CIRJE-F-669

Modelling and Forecasting Noisy Realized Volatility

Manabu Asai
Soka University
Michael McAleer
Erasmus University Rotterdam
and Tinbergen Institute
and CIRJE, Faculty of Economics, University of Tokyo
Marcelo C. Medeiros
Pontifical Catholic University of Rio de Janeiro

September 2009

Modelling and Forecasting Noisy Realized Volatility

Manabu Asai
Faculty of Economics
Soka University, Japan

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

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

Revised: September 2009

* The authors are grateful to Marcel Scharth for efficient research assistance, and to a coeditor and two anonymous referees for useful comments and suggestions. For financial support, the first author acknowledges the Japan Society for the Promotion of Science and the Australian Academy of Science, the second author acknowledges the Australian Research Council and the National Science Council, Taiwan, and the third author wishes to acknowledge CNPq, Brazil.

Abstract

Several methods have recently been proposed in the ultra high frequency financial literature to remove the effects of microstructure noise and to obtain consistent estimates of the integrated volatility (IV) as a measure of ex-post daily volatility. Even bias-corrected and consistent (modified) realized volatility (RV) estimates of the integrated volatility can contain residual microstructure noise and other measurement errors. Such noise is called "realized volatility error". As such measurement errors ignored, we need to take account of them in estimating and forecasting IV. This paper investigates through Monte Carlo simulations the effects of RV errors on estimating and forecasting IV with RV data. It is found that: (i) neglecting RV errors can lead to serious bias in estimators due to model misspecification; (ii) the effects of RV errors on one-step ahead forecasts are minor when consistent estimators are used and when the number of intraday observations is large; and (iii) even the partially corrected R^2 recently proposed in the literature should be fully corrected for evaluating forecasts. This paper proposes a full correction of R^2, which can be applied to linear and nonlinear, short and long memory models. An empirical example for S&P 500 data is used to demonstrate that neglecting RV errors can lead to serious bias in estimating the model of integrated volatility, and that the new method proposed here can eliminate the effects of the RV noise. The empirical results also show that the full correction for R^2 is necessary for an accurate description of goodness-of-fit.

Keywords: realized volatility; diffusion; financial econometrics; measurement errors; forecasting; model evaluation; goodness-of-fit.

1 Introduction

Given the rapid growth in financial markets and the continual development of new and more complex financial instruments, there is an ever-growing need for theoretical and empirical knowledge of volatility in financial time series.

There is, however, an inherent problem in using models where the volatility measure plays a central role. The conditional variance is latent, and hence is not directly observable. It can be estimated, among other approaches, by the (Generalized) Autoregressive Conditional Heteroskedasticity, or (G)ARCH, family of models proposed by Engle (1982) and Bollerslev (1986), stochastic volatility (SV) models (see, for example, Taylor (1986)), or exponentially weighted moving averages (EWMA), as advocated by the Riskmetrics methodology (Morgan, 1996) (see McAleer (2005) for a recent exposition of a wide range of univariate and multivariate, conditional and stochastic, models of volatility, and Asai, McAleer and Yu (2006) for a review of the growing literature on multivariate stochastic volatility models). However, as observed by Bollerslev (1987), Malmsten and Teräsvirta (2004), and Carnero, Peña, and Ruiz (2004), among others, most of the latent volatility models fail to describe satisfactorily several stylized facts that have been observed in financial time series.

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

Ignoring the remaining measurement error, which can be problematic, the ex post volatility essentially becomes "observable", and hence it can be modelled directly, rather than being treated as a latent variable. Based on the theoretical results of Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev, Diebold and Labys (2003) and Meddahi (2002), several recent studies have documented the properties of realized volatility constructed from high frequency data. However, it is well known that neglecting microstructure noise in calculating realized volatility can lead to biased and inconsistent estimates of the integrated volatility as a true measure of daily volatility.

Several methods have recently been proposed in the ultra high frequency financial literature to remove the effects of microstructure noise and to obtain consistent estimates of the integrated volatility (see Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), Christensen, Oomen and Podolskij (2008)**,** Hansen, Large and Lunde (2008), and Zhang, Mykland and Aït-Sahalia (2005)). For an extensive review of the realized volatility literature, see McAleer and Medeiros (2008) and Bandi and Russell (2007). Nevertheless, even bias-corrected and consistent realized volatility estimates of the integrated volatility can contain residual microstructure noise and other measurement errors that should not be ignored. Furthermore, the consistency of the above mentioned estimators is derived under some (strong) assumptions about the microstructure noise. Whenever some of these assumptions are not met in practice, the estimators turn to be inconsistent. Finally, if the number of intraday observations is small (due to illiquidity effects or data availability), the remaining measurement error may not be negligible. Barndorff-Nielsen and Shephard (2002) refer to such remaining noise as the "realized volatility (RV) errors". They suggested a method to estimate the continuous-time SV model, in which volatility follows a non-Gaussian Ornstein-Uhlenbeck (OU) process (see also Corradi and Distaso (2006) for a discussion of measurement errors and realized volatility).

The contribution of this paper is two-fold. First, we extend Barndorff-Nielsen´s and Shephard (2002) approach and estimate three different models of integrated volatility. The common features between Barndorff-Nielsen and Shephard (2002) and the current paper is the use of state space representation to remove such realized volatility errors. This paper deals with discrete-time SV models, in which the logarithm of integrated volatility follows a *K*-component model, a long memory model (ARFIMA), or a heterogeneous autoregressive (HAR) model. Our *K*-component model corresponds to the continuous-time SV model of Chernov et al. (2003). Monte Carlo simulation experiments are presented to investigate the effects of the RV errors on the estimators and forecasts of these three models. Second, we show that, in the presence of RV errors, the R^2 correction proposed by Andersen, Bollerslev and Meddahi (2005) is only a partial correction. We provide a fully corrected R^2 measure in Mincer-Zarnowitz regressions when the dependent variable is a noisy RV measure.

An empirical example is used to show that neglecting the RV error can lead to serious bias in estimating integrated volatility, and that the new method can eliminate the effects of the RV error. Finally, the fully corrected R^2 proposed in this paper is needed in most cases.

The plan of the remainder of the paper is as follows. Section 2 discusses the effects of RV error on estimating and forecasting integrated volatility. Section 3 presents the results of Monte Carlo simulation experiments regarding the effects of RV error, using the *K*-component, long memory and HAR models. Section 4 proposes a new method to fully correct R^2 with RV error. The results of an empirical example are analyzed in Section 5. Some concluding remarks are given in Section 6.

2 Realized Volatility and the Significance of Measurement Errors

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

$$dp(t+\tau) = \mu(t+\tau)d\tau + \sigma(t+\tau)dW(t+\tau), \quad 0 \le \tau \le 1, t = 1, 2, \dots,$$

where $p(t+\tau)$ is the logarithmic price at time $t+\tau$, $\mu(t+\tau)$ is the drift component, $\sigma(t+\tau)$ is the instantaneous volatility (or standard deviation), and $W(t+\tau)$ is a standard Brownian motion. In addition, suppose that $\sigma(t+\tau)$ is orthogonal to $W(t+\tau)$, such that there is no leverage effect. This assumption is standard in the realized volatility literature.

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

$$r_t \Big| \Im_t \sim \mathrm{N}\left(\int_0^1 \mu(t+\tau-1)d\tau, \int_0^1 \sigma^2(t+\tau-1)d\tau, \right).$$

The term $IV_t^2 = \int_0^1 \sigma^2(t+\tau-1)d\tau$ is known as the *integrated variance*, which is a measure of the day-*t* ex post volatility. The integrated variance is typically the object of interest as a measure of the true daily volatility.

In general, $\sigma(t+\tau)$, or a function of $\sigma(t+\tau)$ such as $\sigma^2(t+\tau)$ or $\ln\sigma^2(t+\tau)$, is assumed to follow a continuous time diffusion process (see Ghysels, Harvey and Renault (1996) for example). As integrating on τ makes the Brownian motion of the diffusion process a Gaussian variable, such that the integrated variance is a random variable. In this sense, IV_t^2 plays the same role as the stochastic variance in the class of "Stochastic Volatility (SV)" models. From this viewpoint, the connections among the integrated variance, stochastic variance, and conditional variance are clear. As shown by Nelson (1990), conditional variance models are approximations to continuous-time SV models. In the conditional variance model, the current variance is determined by past information sets, indicating that the approximation can be improved. Usually, continuous-time SV models are approximated by the Euler-Maruyama method, and the resulting models are called "discrete time" SV models. For example, the EGARCH model and the asymmetric SV model of Harvey and Shephard (1996) have the same diffusion limit, in which $\ln\sigma^2(t+\tau)$ follows the OU process with the negative correlation between the Brownian motions of $p(t+\tau)$ and $\ln\sigma^2(t+\tau)$.

Let RV_t be a suitable estimator of the integrated volatility, $IV_t = \sqrt{IV_t^2}$, as suggested by Zhang, Mykland and Aït-Sahalia (2005), hereafter ZMA (2005), or Barndorff-Nielsen, Hansen, Lunde and Shephard (2008) (BHLS (2008)). Then RV_t is consistent, and its order of convergence is n_t^{α}, where n_t is the total number of observations at day *t* and α $\left(1/6 \le \alpha \le 1/2\right)$ depends on the assumptions made about the noise. For simplicity, we may write RV_t as

$$RV_t = IV_t + \frac{1}{n_t^{\alpha}} u_t, \qquad (1)$$

where u_t is an independent process with mean 0 and variance σ_u^2, respectively[1]. Hereafter, $u_t \sim ID(0, \sigma_u^2)$. We call the second term in (1), u_t / n_t^{α}, the "realized volatility error".

The approach proposed in this paper is based on equation (1), which shows that the last term plays a key role as a measurement error. It is known that measurement errors can lead to serious bias in estimating econometric models (see textbooks such as Hayashi (2000)). As the logarithm of RV_t is modelled in the literature, it is useful to consider the measurement error of $\ln RV_t$ when it is based on equation (1). By using a Taylor series expansion of $\ln\left(IV_t + n^{-\alpha} u_t\right)$ around $u_t = 0$, we have

$$\ln RV_t = \ln IV_t + w_t, \qquad (2)$$

where

---

[1] We only assume this for the purpose of describing the idea. We may relax this assumption, as in Barndorff-Nielsen and Shephard (2002), who assume that $E\left(u_t \middle| IV_t\right) = 0$, allowing the variance of u_t to depend on *t*, and u_t to be correlated with IV_t.

$$w_t = \frac{1}{2}\sum_{i=1}^{\infty}\frac{(-1)^{i-1}}{i}\left(\frac{u_t}{n^{\alpha} IV_t}\right)^i .$$

Here, w_t is correlated with RV_t, and is $o_p\left(n^{-\alpha}\right)$.

Consider a general time series model for $\ln IV_t$, such as

$$(1-L)^d \ln IV_{t+1} = v_{t+1},$$

$$v_{t+1} = g\left(v_t, v_{t-1}, \ldots\right) + \eta_t,$$

where L is the lag operator, $g\left(v_t, v_{t-1}, \ldots\right)$ can be a linear or nonlinear function, and η_t is the innovation term. This model includes ARMA and ARFIMA models, by the AR (∞) representation**,** assuming that the invertibility conditions are satisfied. Obviously, it also contains the non-linear AR models. Then we have the model of RV_t as

$$(1-L)^d \ln RV_{t+1} = v_{t+1}^*,$$

$$v_{t+1}^* = g\left(v_t^*, v_{t-1}^*, \ldots\right) + \xi_t,$$

where $v_t^* = v_t + (1-L)^d w_t$ and $\xi_t = \eta_t + (1-L)^d w_{t+1} + g^*\left(X_t^*, W_t\right)$ with the function

$$\begin{aligned} g^*\left(X_t^*, W_t\right) = g\Big((1-L)^d \left\{\ln RV_t - w_t\right\}, (1-L)^d \left\{\ln RV_{t-1} - w_{t-1}\right\}, \ldots\Big) \\ - g\Big((1-L)^d \ln RV_t, (1-L)^d \ln RV_{t-1}, \ldots\Big), \end{aligned}$$

and $X_t^* = \left\{\ln RV_t, \ln RV_{t-1}, \ldots\right\}$ and $W_t = \left\{w_t, w_{t-1}, \ldots\right\}$. This leads a measurement error problem in nonlinear regression models. Estimation neglecting measurement errors produces bias in the estimators, which may affect the bias in the forecasts. Such bias depends both on the model and the size of the RV error.

Consider two examples. If the true $\ln\sigma(t)$ follows an OU process, then $\ln IV_t$ follows an AR(1) process, namely

$$\ln IV_{t+1} = (1-\phi)\mu + \phi \ln IV_t + \eta_t .$$

Then we have a model of RV_t as

$$\ln RV_{t+1} = (1-\phi)\mu + \phi \ln RV_t + \xi_t ,$$

where $\xi_t = \eta_t + w_{t+1} - \phi w_t$. Note that ξ_t is correlated with RV_t, by the structure of the model. Hence, neglecting $w_{t+1} - \phi w_t$ causes a familiar problem of measurement errors in regression models. The OLS estimator for ϕ is biased due to the RV error in RV_t and the correlation between $\ln RV_t$ and the disturbance. On the other hand, taking account of $w_{t+1} - \phi w_t$ leads to an ARMA(1,1) specification of $\ln RV_t$, as an AR(1) plus noise follows an ARMA(1,1) model, in general (see Granger and Morris (1976)). Regarding the forecasts of $\ln IV_t$, it is the same as the forecasts of $\ln RV_t$ as the expectation of w_t is zero. In the case of taking account of w_t, the forecast of $\ln RV_t$ is made of all the past information, $\ln RV_{t-1}, \ln RV_{t-2}, \dots$, due to the AR (∞) representation of a stationary ARMA process. Regarding the case of neglected measurement errors, the forecast of $\ln RV_t$ depends on $\ln RV_{t-1}$ from the AR(1) specification. Hence, forecasts that neglect the measurement errors lead to two kinds of bias, one caused by the bias in the estimate of ϕ, and the other from the lack of information.

Another example is the Heterogeneous Autoregressive (HAR) model of Corsi (2009). Consider the HAR model of $\ln IV_t$ as

$$\ln IV_{t+1} = \beta_0 + \beta_1 \ln IV_t + \beta_2 \frac{1}{5}\sum_{i=1}^{5} \ln IV_{t+1-i} + \beta_3 \frac{1}{22}\sum_{i=1}^{22} \ln IV_{t+1-i} + \eta_t , \qquad (3)$$

which yields the RV_t model as

$$\ln RV_{t+1} = \beta_0 + \beta_1 \ln RV_t + \beta_2 \frac{1}{5}\sum_{i=1}^{5} \ln RV_{t+1-i} + \beta_3 \frac{1}{22}\sum_{i=1}^{22} \ln RV_{t+1-i} + \xi_t ,$$

where

$$\xi_t = \eta_t + w_{t+1} - \beta_1 w_t - \beta_2 \frac{1}{5}\sum_{i=1}^{5} w_{t+1-i} - \beta_3 \frac{1}{22}\sum_{i=1}^{22} w_{t+1-i} .$$

We may apply the discussion which is similar to the above. In this case, $\ln RV_t$ follows an ARMA(22,22) model, in general. Hence, neglecting the effects of w_t and using OLS lead to bias in the estimates of β_0, β_1, β_2 and β_3. Furthermore, forecasts obtained by neglecting measurement errors will be biased due to the bias in the estimates and the lack of information. Overall, the moving average term caused by measurement error plays an important role in estimating and forecasting integrated volatility.

3 Effects of RV Errors

In the following section, we will investigate the effects of the RV error on estimating and forecasting volatility models. We consider three kinds of models, namely the *K*-component, long memory and HAR models, which are familiar in empirical analysis. Then we will conduct Monte Carlo simulations using two quasi-maximum likelihood

(QML) estimators, one taking account of measurement errors caused by RV error, and another which neglects measurement errors. The purpose of the simulations is to (i) compare the finite sample properties of two estimators, (ii) investigate differences in forecasts based on these estimators, and (iii) check the effects on the corrected R^2 values.

3.1 *K*-component Model

With regard to the integrated volatility, IV_t, consider the following *K*-component model:

$$IV_t = \sigma \exp\left(\sum_{i=1}^{K} \alpha_{it}\right),$$
$$\alpha_{i,t+1} = \phi_i \alpha_{it} + \sigma_i \eta_{it}, \quad i = 1, \ldots, K, \tag{4}$$

where η_{it} follows the independent standard normal distribution. In the literature of stochastic volatility based on observed return series, Chernov et al. (2003) and Asai (2008), among others, consider such a *K*-component model in a more general framework. Here, we will consider estimation of the model via a proxy for the latent integrated volatility, namely realized volatility.

Based on equations (1)-(4), we have

$$\ln RV_t = \ln IV_t + w_t = \ln \sigma + \sum_{i=1}^{K} \alpha_{it} + w_t. \tag{5}$$

Thus, we can construct the state space model with the measurement equation (5) and the state equation of α_{it}, which enables an application of the QML method via the Kalman filter. Note that the distribution of the measurement error, w_t, is unknown.

We may have filtered (or smoothed) the estimate of the logarithm of integrated volatility via the Kalman filter (or smoother). For purposes of forecasting out of sample, the one-step ahead predicted value, $\ln\hat{\sigma}+\sum_{i=1}^{K}\hat{\alpha}_{i,T+1}$, is also available.

The method here includes estimation of the *K*-component model in the absence of RV errors. Let σ_w be the standard deviation of w_t. By setting $\sigma_w=0$, the approach can deal with the case of no measurement errors.

3.2 Long Memory Model for Integrated Volatility

In this section we consider a long memory model for the logarithm of integrated volatility. For convenience, we assume that $IV_t=\sigma\exp\left(x_t\right)$ and that x_t follows an ARFIMA(*p*,*d*,*q*) model. Then we have

$$\begin{aligned}&\ln RV_t=\ln IV_t+w_t=\ln\sigma+x_t+w_t,\\&\left(1-L\right)^d\phi\left(L\right)x_t=\theta\left(L\right)\eta_t,\end{aligned}\tag{6}$$

where $\eta_t\sim N\left(0,\sigma_\eta^2\right)$ and w_t is defined by equation (2). The spectral density of the model is given by

$$f\left(\lambda\right)=\frac{\sigma_\eta^2\left|\theta\left(e^{-i\lambda}\right)\right|^2}{2\pi\left|1-e^{-i\lambda}\right|^{2d}\left|\phi\left(e^{-i\lambda}\right)\right|^2}+\frac{\sigma_w^2}{2\pi},\quad -\pi\le\lambda\le\pi .$$

Thus, we may apply the method of Breidt, Crato and de Lima (1998) in order to estimate the above model[2]. With an adaptation of the algorithm given in Harvey (1998), we can obtain the estimates and forecasts of $\ln IV_t$. In order to estimate the model without RV errors, we need only set σ_w to zero.

3.3 HAR Model for Integrated Volatility

We consider the HAR model for integrated volatility as

$$
\begin{aligned}
&\ln RV_t = \ln IV_t + w_t = \ln \sigma + x_t + w_t, \\
&\ln x_{t+1} = \beta_1 \ln x_t + \beta_2 \frac{1}{5}\sum_{i=1}^{5} \ln x_{t+1-i} + \beta_3 \frac{1}{22}\sum_{i=1}^{22} \ln x_{t+1-i} + \eta_t,
\end{aligned}
\tag{7}
$$

where $\eta_t \sim N\left(0, \sigma_\eta^2\right)$ and w_t is defined by equation (2). Note that setting $\sigma = \exp\left[\beta_0 / \left(1 - \beta_1 - \beta_2 - \beta_3\right)\right]$ and $x_t = \ln IV_t - \ln \sigma$ leads to equation (3). As the model is an AR(22) plus noise, we can use the QML method via the Kalman filter. For purposes of forecasting out of sample, the one-step ahead predicted value is also available from the Kalman filter.

For the case of neglecting measurement error, we may handle the case by setting $\sigma_w = 0$. In this case, the QML estimator is equivalent to the OLS estimator.

3.4 Framework of Experiments

We start from equation (1) for specifying the magnitude of the RV error. The variance of

[2] An alternative method is to work with the filtering algorithm proposed by So (1999), but we abandoned it because of its computational burden.

the RV error is given by $\sigma_u^2 / n_t^{2\alpha}$. We consider the case of ZMA (2005), which indicates that $\alpha = 1/6$. Let the variance of IV_t be σ_{iv}^2. Then we define the variance ratio of the RV error to volatility as

$$ev = \frac{\sigma_u^2}{\sigma_{iv}^2 \sqrt{n_t}}. \tag{8}$$

In the following, we set $ev = 0.03$ in order to consider a minor RV error compared with volatility. It should be noted that, if the RV error is large, it will lead to bias in estimating and forecasting the models of IV. Hence, we exclude the obvious case in order to concentrate on the case that the estimator of RV is consistent and well-behaved.

In the following Monte Carlo simulations, we generate data of IV_t with sample size T+1. The parameter setting are as follows; $(\phi_1, \sigma_1, \phi_2, \sigma_2, \sigma) = (0.98, 0.1, 0.4, 0.4, 1)$ for the two component model (equation (4) with K=2), $(d, \sigma_\eta, \phi, \sigma) = (0.4, 0.4, -0.1, 1)$ for the ARFIMA(1,d,0) model (6), and $(\beta_1, \beta_2, \beta_3, \sigma_\eta, \sigma) = (0.8, 0.1, 0.05, 0.25, 1)$ for the HAR model (7). Next, we generate the noise process in the following way. For the parameter values above, we can calculate the variance of $\ln IV_t$ as the variance of the ARMA and ARFIMA models are available. Then, by using the property of the log-normal distribution, we can obtain the value of σ_{iv}^2. With $n_t = 250$ and ev=0.03, we obtain the values of σ_u as $\sigma_u = 0.642$ for the two component model and $\sigma_u = 0.466$ for the ARFIMA model. We generated $U_t \sim N(0, \sigma_u^2)$ in order to caucluate RV_t via (1). The first T observations are used for estimation of the models, while the last observation is used for forecasting evaluation. The number of replications, at 500, is relatively small, but there are no changes to the conclusions below where the number of replications is 400.

For each replication, we estimate the models with and without measurement errors in order to investigate the finite sample properties of the QML estimators, and to compare the performances of the one-step-ahead predictions.

Let $\hat{h}_{T+1|T}^{(i)}$ $(i = 1, 2, \ldots, 500)$ be the one-step-ahead forecast of $\ln IV_{T+1}$ in the i-th replication. We calculate the mean absolute error (MAE) and root mean squared errors (RMSE) based on the true values. In addition to these values, we use two kinds of Mincer-Zarnowitz regression:

$$MZ^{IV}: \quad \ln IV_t = \gamma + \delta \hat{h}_{t|t-1} + \text{error},$$

$$MZ^{RV}: \quad \ln RV_t = \alpha + \beta \hat{h}_{t|t-1} + \text{error},$$

for purposes of investigating the effects of using the noisy RV as the regressand.

3.5 Monte Carlo Results

This subsection reports the results of the Monte Carlo simulations described above. Table 1(a) shows the true parameters and the mean, standard deviations and RMSE of two kinds of QML estimators for 500 replications with T=1000. As it is not easy to obtain the true value of σ_w analytically, we use the simulated value from 500 replications as a proxy. The QML estimator taking account of RV error has a small bias, while the QML estimator neglecting RV error has a relatively large bias, especially for (ϕ_2, σ_2). On the other hand, introducing σ_w makes the standard deviations for (ϕ_2, σ_2) larger, compared with those for the QML neglecting RV error. Overall, the RMSE values for the QML with measurement errors are always smaller than those for the QML neglecting measurement errors.

In the above simulations, we also obtained the predicted values, $\hat{\alpha}_{1,T+1}$, $\hat{\alpha}_{2,T+1}$ and $\ln\hat{\sigma} + \sum_{i=1}^{2} \hat{\alpha}_{i,T+1} \left(= \hat{h}_{T+1|T}\right)$. Table 1(b) presents the MAE and RMSE values for the predictions of $\alpha_{1,T+1}$, $\alpha_{2,T+1}$ and $\ln IV_{T+1}$. The QML estimator taking account of RV error always has smaller MAE and RMSE values than the corresponding QML estimator neglecting the RV error. Table 1(c) shows R^2 and the F test for the MZ^{IV} and MZ^{RV} regressions. The F test is for the null hypothesis $H_0: \alpha = 0$ and $\beta = 1$ ($\gamma = 0$ and $\delta = 1$) in the MZ equation for $\ln RV_t$ ($\ln IV_t$). The p-values in Table 1(c) indicate that the model is correctly specified, as expected.

In both the cases of including and neglecting RV error, MZ^{IV} has larger R^2 values than does MZ^{RV}. This result is reasonable, as the denominator of R^2 is the sum of squared deviations of the regressand, for which $\ln IV_t$ has smaller values than does $\ln RV_t$. An interesting result is that, for MZ^{RV}, the QML estimator taking account of RV error has smaller R^2 than the corresponding QML estimator neglecting RV error. Thus, R^2 for MZ^{RV} yields a misleading result. Therefore, we must be careful in comparing R^2 values based on MZ^{RV}. We will discuss this point further in the next section.

Now we turn to the results for the ARFIMA model, which is given in Table 2. Table 2(a) presents the true parameters and the mean, standard deviations and RMSE of two kinds of QML estimators for 500 replications with *T*=1000. We set σ_w as the simulated value from 500 replications as given previously. In the results for QML neglecting measurement errors, the estimator for *d* has a downward bias, while the estimator for σ_η has an upward bias. The estimator for ϕ is unbiased. In the results for QML taking account of RV error, the bias is minor, except for ϕ, but the standard deviations are relatively large. This may be explained by three reasons: (i) The sample size is relatively small for the analysis of a long memory process; (ii) the measurement error in the current parameter setting is too small to detect; (iii) As in Table 1(a), introducing σ_w for accommodating measurement errors make larger the standard deviations of some parameters which are strongly affected by neglecting the noise.

In order to investigate the effects of sample size, Table 3 reports the results for *T*=2000. The bias in ϕ for the QML accommodating the RV errors becomes smaller. In all cases, the standard deviations and RMSE are smaller than those in Table 2.

For purposes of forecasting integrated volatility, Table 2(b) shows that the QML estimator taking account of RV error always has smaller MAE and RMSE values than the corresponding QML estimator neglecting the RV error. This is the same as in Table 1(b). Table 2(c) presents R^2 and the *F* test for the MZ^{IV} and MZ^{RV} regressions. The *p*-values of the *F* test are for the null hypothesis $H_0 : \alpha = 0$ and $\beta = 1$

($\gamma = 0$ and $\delta = 1$) in the *MZ* equation for $\ln RV_t$ ($\ln IV_t$), and indicate that the model is correctly specified. As before, MZ^{IV} has larger R^2 values than does MZ^{RV}. Furthermore, the QML estimator taking account of the RV error always has larger R^2 values than the corresponding QML estimator neglecting the RV error. We obtain the same conclusions from Tables 3 (b) and 3(c) for *T*=2000.

Third, we discuss the simulation results of the HAR model, which are given in Table 4. Table 4(a) presents the true parameters and the mean, standard deviations and RMSE of two kinds of estimators for 500 replications with *T*=1000. We set σ_w to be the simulated value from 500 replications, as above. In the results for OLS neglecting measurement errors, the estimators of β_1 and β_2 have large bias, while the estimator for σ_η has an upward bias. In the results for QML accounting for RV error, the bias is negligible. As noted previously, introducing σ_w for accommodating RV errors makes larger the standard deviations of some parameters which are strongly affected by neglecting the noise.

For purposes of forecasting integrated volatility, Table 4(b) shows that the QML estimator taking account of RV error always has smaller MAE and RMSE values than the corresponding OLS estimator neglecting RV error. Table 4(c) presents R^2 and the *F* test for the MZ^{IV} and MZ^{RV} regressions. The *p*-values of the *F* test indicate that the model for ln*RV* is correctly specified. As before, MZ^{IV} has larger R^2 values than does MZ^{RV}.

In the Monte Carlo simulations for the effects of a relatively small noise, it is found that: (i) the estimator neglecting the RV error has bias; (ii) the estimator taking account of RV error produces better forecasts than the estimator neglecting the RV error, but the differences are minor; and (iii) the R^2 values based on MZ^{RV} are misleading, and needs to be corrected.

4 Correcting $\mathbf{R^2}$

As shown in the previous section, we need to correct R^2 based on MZ^{RV}. A natural framework is to use the correction suggested by Andersen, Bollerslev and Meddahi (2005). In the following, we will examine the results of Monte Carlo experiments in detail, showing that their partial correction is insufficient. Then we will propose a fully corrected R^2 measure.

4.1 Implications from the Monte Carlo results.

The essence of Andersen, Bollerslev and Meddahi (2005) is to multiply by

$$\frac{V\left(\ln RV_t\right)}{V\left(\ln IV_t\right)}$$

the R^2 values based on MZ^{RV} in the previous section. This is reasonable as the denominator of R^2 is the squared sum of deviations of $\ln RV_t$, but R^2 based on MZ^{IV} uses $\ln IV_t$. For reasons that will become clear below, we will refer to this type of R^2 as the 'partially corrected R^2'.

Regarding the previous Monte Carlo experiments, Table 5(a) shows the R^2 based on MZ^{RV} and MZ^{IV}, and the partially corrected R^2 by using the sample value of

$V(\ln RV_t)/V(\ln IV_t)$. Clearly, the partially corrected R^2 overestimates R^2 for MZ^{IV}, indicating that we have failed to use some important information.

The Appendix shows the algebraic relationship between R^2 for MZ^{RV} and MZ^{IV}, indicating that we need to multiply by not only the sample value of $V(\ln RV_t)/V(\ln IV_t)$ but also by $\left(\hat{\delta}/\hat{\beta}\right)^2$. We will refer to this type of corrected R^2 as the 'fully corrected R^2'. Table 5 also presents the value of $\left(\hat{\delta}/\hat{\beta}\right)^2$ and the fully corrected R^2. This time the resulting R^2 coincides with R^2 for MZ^{IV}. Therefore, the full correction is needed for real data.

4.2 Proposed Approach

From the results above, we need two kinds of correction, the adjustment for the denominator by $V(\ln RV_t)/V(\ln IV_t)$, and also for the numerator by $\left(\hat{\delta}/\hat{\beta}\right)^2$. In real data analysis, $V(\ln IV_t)$ and $\hat{\delta}$ are unavailable, and they have to be estimated. We can estimate $V(\ln RV_t)/V(\ln IV_t)$ by the approach of Andersen, Bollerslev and Meddahi (2005), so that we only need to estimate $\hat{\delta}$, which is given by

$$\hat{\delta} = \hat{\beta} - \frac{\sum\left(\hat{h}_{t|t-1} - \bar{h}\right) w_t}{\sum\left(\hat{h}_{t|t-1} - \bar{h}\right)^2}. \tag{9}$$

Appendix shows how to derive the connection between $\hat{\beta}$ and $\hat{\delta}$. Equation (9) indicates that we also need to estimate w_t.

For this paper, we propose a simple method as follows. First of all, using the whole sample, including those for forecasting, we estimate the model taking account of the measurement errors. Second, for the estimated parameter value, we conduct filtering techniques below in order to obtain the filtered estimate of w_t for the forecasting period. Third, we obtain an estimate of $\hat{\delta}$ by substituting the estimates of w_t for the true value of w_t in (9). Note that this estimate of w_t may be used not only for the model with measurement errors but also for the model neglecting measurement errors.

With respect to the filtering technique, we suggest the following approach. For the short-memory models including the ARMA and *K*-components models, we can use the Kalman filter. For the case of the long-memory ARFIMA model, we may use the filtering algorithm proposed by So (1999). Regarding nonlinear time series models, we can work with particle filters, such as in Kitagawa (1987). Note that another candidate for w_t is the smoothed estimates.

In general, we may assume that $E\left(u_t \middle| IV_t\right) = 0$ for u_t, as in Barndorff-Nielsen and Shephard (2002), such that the variance of u_t depends on *t* and u_t can be correlated with IV_t. The correction proposed here is still valid. For evaluating the forecasts of IV_t and IV_t^2, a similar correction is required, in addition to the partial correction of Andersen, Bollerslev and Meddahi (2005) . The additional correction requires the estimation of w_t in

$$RV_t = IV_t + w_t$$

for volatility, and

$$RV_t^2 = IV_t^2 + w_t$$

for volatility squared. If the models for log-volatility are considered, as in the current paper, we may use the particle filters for obtaining filtered estimates of w_t, in general.

Estimates of w_t and $\ln IV_t$ are available simultaneously. It is an open question whether this estimate is used for empirical analyses. We leave this problem for future research.

4.3 Simulation Results

In order to check the performance of the proposed fully corrected approach, we conduct another Monte Carlo simulation. The previous Monte Carlo experiments considered the series of $\hat{h}^{(i)}_{T+1|T}$ $(i = 1, 2, \ldots, 500)$. In other words, each $\hat{h}^{(i)}_{T+1|T}$ was calculated for the *i*-th replication. Now we generate IV_t and RV_t with the same parameters and with the sample size of *T*+500 once only. Then we forecast the model to obtain $\hat{h}_{T+j|T+j-1}$ $(j = 1, 2, \ldots, 500)$, fixing the window size as *T*. For each forecast, the model is re-estimated. After forecasting, we estimate the model with the whole sample (*T*+500) in order to have the filtered estimate of w_t, $\tilde{w}_t$. On evaluating the forecasts using MZ^{RV}, we can correct R^2 by multiplying $\left(\tilde{\delta}/\hat{\beta}\right)^2$ for the corrected R^2 by the sample value of $V\left(\ln RV_t\right)/V\left(\ln IV_t\right)$, where $\tilde{\delta}$ is the estimate of $\hat{\delta}$ based on (9) and $\tilde{w}_t$. It should be noted that the smoothed estimate of w_t is another proxy for w_t.

Table 5(b) shows the simulation results for R^2 based on MZ^{RV} and MZ^{IV}, the partially corrected R^2 by using the sample value of $V\left(\ln RV_t\right)/V\left(\ln IV_t\right)$, the values of $\left(\hat{\delta}/\hat{\beta}\right)^2$ and $\left(\tilde{\delta}/\hat{\beta}\right)^2$, and the fully corrected R^2 based on $\left(\tilde{\delta}/\hat{\beta}\right)^2$. We first analyse the results for the two component model. Roughly speaking, the difference between R^2 for MZ^{IV} and the partially corrected R^2 is 0.025, while the difference between the true R^2 and fully corrected R^2 is 0.01, such that the fully corrected R^2 yields a better estimate of the true value. For the ARFIMA model, we have a similar result. The difference between R^2 for MZ^{IV} and the partially corrected R^2 is 0.03, while the difference between R^2 for MZ^{IV} and fully corrected R^2 is 0.02. With respect to the HAR model, the difference between R^2 for MZ^{IV} and the partially corrected R^2 is 0.025, while the difference between true R^2 and fully corrected R^2 is 0.02. In short, the fully corrected R^2 can be far more accurate than its partially corrected counterpart in some cases, but it is never worse.

Before concluding the section, we discuss the effects of model misspecification. If the model is misspecified, which is typically the case for most models used in empirical research, the model misspecification error can be confused as a measurement error in finite samples. Hence, we need to separate the effect of measurement error from model misspecification error. For this purpose, we suggest using the most general model considered for the analysis in order to obtain the estimate of w_t. Note that the most general model need not produce the best out-of-sample forecasts, but it is expected to have best in-sample forecasts, that is, fitted values, if the sample size is large enough. In applied work, the true model may encompass the most general model considered for the analysis, yielding model misspecification error. The fully corrected R^2 is not worse in any case than its partially corrected counterpart, by construction. For the case of our simulation experiments, ARFIMA is the most general model, as the remaining two models have only ARMA representations. We obtained the filtered estimate of w_t, employing the ARFIMA model, for the case that each of the other two models is correct. Then we found that the results for the full corrected R^2 remain unchanged.

5 Empirical Example

This section examines the estimates and forecasts using the RV of Standard and Poor's 500 Composite Index (S&P 500). In order to calculate the daily realized volatility, we use the estimation method given in ZMA (2005). The sample period is Jan/3/1996 to March/29/2007, giving T=2796 observations of RV.

An anonymous referee suggested comparing the models given above with models including an MA(1) term, namely, the two component model (AR(1)+ARMA(1,1)), the ARFIMA(1,*d*,1) model, and the ARMA(22,1) with restrictions on the AR coefficients. The additional parameter is the coefficient of the MA(1) term, which is the same as the models with measurement errors. Intuitively, including the MA(1) term is more comprehensive than accommodating the measurement errors. We will use the three models for estimating and forecasting implied volatility.

Before estimating the models, it is useful to test for the existence of measurement errors. Tanaka (2002) proposed the LM statistic to test the presence of measurement errors based on three kinds of processes, namely the AR(p), unit root and long memory models. The test statistics have the standard normal distribution under the null hypothesis of no measurement error, and is one-sided on the right tail. Table 6 shows the descriptive and test statistics for the logarithm of RV, with descriptive statistics of returns and RV itself. When an AR(1) model is assumed to be the true process of the logarithm of integrated volatility, the calculated statistic rejects the null hypothesis of no measurement error. When an ARFIMA(1,d,0) process is assumed to be true, the calculated statistic also rejects the null hypothesis. The empirical results indicate that there are measurement errors which are not negligible, even after ostensibly removing the microstructure noise.

Table 7 shows the QML estimates for the two-component model, accounting for and neglecting measurement errors. For the former, the estimated value of ϕ_1 is close to 0.99, while the estimate of σ_1 is 0.07, which are typical for SV models. For the second component, the estimate of ϕ_2 is 0.80, while that of σ_2 is 0.18. The estimate of σ_w is 0.38, and is significant at the five percent level, indicating that measurement errors are not negligible. For the case of neglecting measurement errors, the estimates of ϕ_1 and ϕ_2 decrease, while those of σ_1 and σ_2 increase. As expected from the Monte Carlo simulations, the differences for the second factor are large and not negligible, showing the large bias that are arise from neglecting the measurement errors. Table 7 also presents the QML estimates for the two component model comprising AR(1) and ARMA(1,1). The estimate of the MA(1) term is negative and significant. All other estimates, apart from σ_2, are close to those of QML with measurement errors.

Table 8 presents two kinds of QML estimates for the ARFIMA(1,*d*,0) model, one based on accommodating measurement error and another neglecting measurement error. For the former, which takes account of measurement error, the estimate of d is 0.49, indicating that $\ln IV_t$ has long range persistence and is stationary. The estimate of ϕ is positive and significant, while the estimate of σ_w is close to that in Table 7. For the latter, the estimate of d is 0.48, for which the downward bias is expected from the Monte Carlo simulations. The estimate of ϕ is negative and significant. Table 8 also gives the estimates of ARFIMA(1,*d*,1) as a counterpart to the ARFIMA(1,*d*,0) model accommodating measurement error. The estimate of the MA term is positive and significant.

Table 9 presents estimates for the HAR model. For the QML estimates accounting measurement errors, the estimate of $\beta_1 + \beta_2 + \beta_3$ is 0.96, which implies high persistence in volatility. The estimate of σ_w is 0.33, which is smaller than those in Tables 7 and 8. Regarding the case of OLS neglecting measurement error, the estimate of $\beta_1 + \beta_2 + \beta_3$ is 0.93, but the estimates of each parameter is different from those of the former. As the HAR model is the AR(22) model with parametric restrictions, we also estimated the ARMA(22,1) model with restrictions on the AR coefficients. The estimate of the MA term is negative and significant. The estimate of $\beta_1 + \beta_2 + \beta_3$ is 0.96, which is close to the QML estimate accounting for measurement errors.

Next, we compare the out-of-sample forecasts based on these three approaches, which are the two-component plus noise model, ARFIMA plus noise model, and HAR plus noise model. The period of forecast is the last 796 observations. For each set of forecasts, the parameters are re-estimated to calculate the one-step-ahead forecasts, fixing the sample size to 2000.

The forecasts are evaluated by estimating the Mincer-Zarnowitz regression, MZ^{RV}. We compute the partially corrected R^2 values, as proposed in Andersen, Bollerslev and Meddahi (2005), as a measure of the ability of the model to track the variance over time. We also calculate the fully corrected R^2 values, as proposed in the previous section, but do not calculate the mean absolute errors or root mean squared errors as they neglect the measurement error in realized volatility. We also computed the heteroskedasticity-robust F statistic of the joint hypothesis $\alpha = 0$ and $\beta = 1$. If the model is correctly specified, the joint null hypothesis is not rejected.

Table 10 reports the partially and fully corrected R^2 values and the *p*-values of the F statistics for the two-component, ARFIMA(1,*d*,0) and HAR models. For estimating w_t with the full correction, we use the ARFIMA model, as discussed in the previous section. The differences between the partially and fully corrected R^2 values are not negligible, implying the importance of taking account of the measurement error fully in correcting R^2. The partially corrected R^2 selects the two component model accommodating the measurement error, while the fully corrected R^2 chooses the ARFIMA models with/without measurement error. This can happen as in the Monte Carlo simulations, which suggest that the fully corrected R^2 provides a more accurate estimate of the true R^2 than does the partially corrected R^2.

Furthermore, Table 10 indicates that the fully corrected R^2 values show that the

ARFIMA models with/without measurement errors have the highest value, while the two-component model with the MA(1) term has the lowest. We will examine the differences among the models with and without measurement errors. As stated previously, the ARFIMA models have the highest value of the fully corrected R^2, while the HAR model with measurement error has the lowest. The second best model is the two-component model with measurement error, while the remaining two models have similar values. As some of the differences in the corrected R^2 values among the six models are very small, we need to assess the results by an alternative approach.

For the complementary analysis, we conduct the tests for forecast encompassing suggested by Harvey, Leybourne and Newbold (1998). Consider a combination of two forecasts, f_{1t} and f_{2t}, as $f_{ct} = (1-\lambda) f_{1t} + \lambda f_{2t}$ $(0 \le \lambda \le 1)$, in order to produce forecasts that are superior to the two individual forecasts. The null hypothesis is $\lambda = 0$, and the alternative hypothesis $(\lambda > 0)$ has an interpretation that f_{2t} contains useful information that is not present in f_{1t}. For the case $\lambda = 0$, f_{1t} is said to "encompass" f_{2t}.

Table 11 gives the *p*-values of the test of Harvey, Leybourne and Newbold (1998) with respect to the models with/without measurement errors. The interpretation of the results is as follows. The forecasts of both ARFIMA models encompass all the other forecasts. The forecast of the two factor model with measurement error encompasses the remaining three models. The forecast of the HAR model with measurement error encompasses no forecasts. The implication obviously supports the fully corrected R^2 values in Table 10. Note that the test of forecast encompassing is not suitable for model evaluation for the following two reasons: (i) it neglects the measurement errors, and (ii) the test may potentially find two forecasts which are unable to encompass each other.

Returning to Table 10, The F statistics are significant at the five percent level, suggesting that the model can be improved, with one possibility being to accommodate leverage effects, which is supported by the skewness of ln*RV*, as shown in Table 6.

Table 12 gives the partially and fully corrected R^2 values for the h step-ahead forecasts (h=5, 10, 20), regarding the two component, ARFIMA and HAR models with/without measurement errors and with the MA(1) term. As noted previously, the differences between the partially and fully corrected R^2 values are not negligible. In all cases, the fully corrected R^2 chooses the HAR model with measurement errors. For the cases h=5 and 10, the two component (AR(1)+ARMA(1,1)) model has the lowest values of the fully corrected R^2, whereas for the case h=20, the ARFIMA model neglecting measurement errors is chosen. Table 12 has omitted the results for the F test of the null hypothesis that $H_0 : \alpha = 0$ and $\beta = 1$ in the *MZ* equation. The tests all reject the null hypothesis, thereby indicating that the models can be improved.

6 Conclusion

Neglecting microstructure noise in calculating realized volatility (RV) can lead to biased and inconsistent estimates of the integrated volatility as a true measure of daily volatility. Consequently, several methods have recently been proposed in the ultra high frequency financial literature to remove the effects of microstructure noise and to obtain consistent estimates of the integrated volatility. However, even bias-corrected and consistent RV estimates of the integrated volatility contain RV errors that should not be ignored.

This paper investigated the effects of RV errors on estimating and forecasting models of integrated volatility. For minor RV errors, the Monte Carlo results showed that: (i) the estimates neglecting measurement error have serious biases; (ii) forecasts accounting for the measurement error outperform those neglecting them, but the differences can be small; and (iii) R^2 for evaluating the forecasts should be corrected appropriately.

This paper also proposed a new method to correct R^2 of the Mincer-Zarnowitz regression, which is based on measurement errors in the estimated model. Theoretical and Monte Carlo results showed that the new fully corrected method is preferred to the partially corrected approach of Andersen, Bollerslev and Meddahi (2005).

The empirical example of S&P 500 showed that neglecting microstructure noise can cause serious bias in estimating integrated volatility. Such bias in forecasting was found to be small. The proposed fully corrected R^2 showed the clear difference with the partially corrected R^2 of Andersen, Bollerslev and Meddahi (2005), implying that the appropriate correction can be empirically useful.

Data Appendix: Construction of Daily Realized Volatility Measures

The empirical analysis focuses on the realized volatility of the S&P 500 index. We start by removing non-standard quotes (that is, discarding quotes where the bid or offer price is missing and selecting observations where the "mode" field in the TAQ file is 3, 5, 10, 12 or 29; see the description below), computing prices through the mean of the bid and ask quotes, filtering possible errors (namely, ruling out implausible returns in relation to the last quotes), and obtaining one second returns for the 9:30 am to 4:00 p.m. period (which are the regular trading hours on the NYSE).

Observing the consistency considerations in Hansen and Lunde (2006), the previous tick method for determining prices at precise second marks is implemented. Based on the results of Hasbrouck (1995), who reports a median 92.7% information share at the NYSE for Dow stocks, and Blume and Gold (1997), who conclude that NYSE quotes match or determine the best displayed quote most of the time, we privilege NYSE quotes if there is more than one quote in a given second.

In order to calculate the daily realized volatility, we use the estimation method given in BHLS (2008).

Appendix: Fully Corrected R^2

Consider the following structure of noise

$$y_t = x_t + w_t\,, \tag{A.1}$$

where x_t follows a univariate dependent process, and w_t is correlated with x_t. Although x_t is assumed to be latent, we can observe y_t. We denote the forecast of x_t as $\hat{h}_t$, where y_t and x_t in (A.1) correspond to $\ln RV_t$ and $\ln IV_t$, respectively, in the text.

Regarding the two Mincer-Zarnowitz regressions:

$$\begin{aligned} x_t &= \gamma + \delta \hat{h}_t + \text{error}, \\ y_t &= \alpha + \beta \hat{h}_t + \text{error}. \end{aligned} \tag{A.2}$$

R^2 is given as

$$R_x^2 = \frac{\hat{\delta}^2 \sum\left(\hat{h}_t - \bar{\hat{h}}\right)^2}{\sum\left(x_t - \bar{x}\right)^2}\,, \tag{A.3}$$

and

$$R_y^2 = \frac{\hat{\beta}^2 \sum\left(\hat{h}_t - \bar{\hat{h}}\right)^2}{\sum\left(y_t - \bar{y}\right)^2}\,, \tag{A.4}$$

respectively, where $\hat{\delta}$ and $\hat{\beta}$ are OLS estimates, and $\bar{\hat{h}}$, $\bar{x}$ and $\bar{y}$ denote the means of $\hat{h}_t$, x_t and y_t, respectively. In (A.2), only the second R^2 can be calculated empirically.

In order to obtain the latent R_x^2 from the observed R_y^2, we need to multiply R_y^2 not only by $\sum(y_t-\overline{y})^2 \big/ \sum(x_t-\overline{x})^2$ but also by $\left(\hat{\delta}\big/\hat{\beta}\right)^2$. It should be noted that

$$\hat{\delta}=\hat{\beta}-\frac{\sum\left(\hat{h}_t-\overline{\hat{h}}\right)w_t}{\sum\left(\hat{h}_t-\overline{\hat{h}}\right)^2}, \tag{A.5}$$

in which the second term in (A.5) does not approach zero as $T\to\infty$ because of the correlation between $\hat{h}_t$ and w_t. Multiplication of R_y^2 by $\sum(y_t-\overline{y})^2 \big/ \sum(x_t-\overline{x})^2$ gives the partially corrected R^2 of Andersen, Bollerslev and Meddahi (2005), while the use of the additional information through $\left(\hat{\delta}\big/\hat{\beta}\right)^2$ gives the fully corrected R^2 developed in this paper.

References

Aït-Sahalia, Y., P.A. Mykland and L. Zhang (2006), "Ultra high frequency volatility estimation with dependent microstructure noise", Working Paper, w11380, NBER.

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

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

Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003), "Modeling and forecasting realized volatility", *Econometrica*, **71**, 529 – 626.

Andersen, T., T. Bollerslev and N. Meddahi (2005), "Correcting the errors: A note on volatility forecast evaluation based on high-frequency data and realized volatilities," *Econometrica*, **73**, 279 – 296.

Asai, M. (2008), "Autoregressive stochastic volatility models with heavy-tailed distribution: A comparison with multifactor volatility models", *Journal of Empirical Finance*, **15**, 332-341.

Asai, M., M. McAleer and J. Yu (2006), "Multivariate stochastic volatility: A review", *Econometric Reviews*, **25**, 145 – 175.

Bandi, F.M. and J.R. Russell (2007), "Volatility estimation", In J. Birge and V. Linetsky eds., *Handbook in Operations Research and Management Science: Financial Engineering*, North Holland, Elsevier, 183－222.

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

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

Blume, M. and M. Goldstein (1997), “Quotes, order flow, and price discovery”, *Journal of Finance*, **52**, 221 – 244.

Bollerslev, T. (1986), “Generalized autoregressive conditional heteroskedasticity”, *Journal of Econometrics*, **21**, 307 – 328.

Bollerslev, T. (1987), “A conditionally heteroskedastic time series model for speculative prices and rates of return”, *Review of Economics and Statistics*, **69**, 542 – 547.

Breidt, F.J., N. Crato and P.J.F. de Lima (1998), “The detection and estimation of long-memory in stochastic volatility”, *Journal of Econometrics*, **83**, 325 – 348.

Carnero, M.A., D. Peña and E. Ruiz (2004), “Persistence and kurtosis in GARCH and stochastic volatility models”, *Journal of Financial Econometrics*, **2**, 319 – 342.

Chernov, M., A.R. Gallant, E. Ghysels and G. Tauchen (2003), “Alternative models for stock price dynamics”, *Journal of Econometrics*, **116**, 225 – 257.

Christoffersen, P. (1998), “Evaluating interval forecasts”, *International Economic Review*, **39**, 841 – 862.

Corradi, V. and W. Distaso (2006), “Semiparametric comparison of stochastic volatility models using realized measures”, *Review of Economic Studies*, **73**, 635－677.

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

Engle, R.F. (1982), “Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation”, *Econometrica*, **50**, 987－1007.

Ghysels, E., A.C. Harvey and E. Renault (1996), “Stochastic volatility”, In C. R.Rao and G. S. Maddala eds., *Statistical Methods in Finance*, pp.119－191. Amsterdam: North-Holland.

Granger, C.W.J. and M. Morris (1976), “Time series modeling and interpretation”, *Journal of the Royal Statistical Society*, Series A, **139**, 246－257.

Hansen, P.R., J. Large and A. Lunde (2008), "Moving average-based estimators of integrated variance", *Econometric Reviews*, **27**, 79－111.

Hansen, P.R. and A. Lunde (2006), "Realized variance and market microstructure noise" (with discussion), *Journal of Business and Economic Statistics*, **24**, 127 – 218.

Harvey, A.C. (1998), "Long memory in stochastic volatility", in J. Knight and E. Satchell (eds.), *Forecasting Volatility in Financial Markets*, Butterworth–Haineman, London, pp. 307–320.

Harvey, A. C. and N. Shephard (1996), "Estimation of an asymmetric stochastic volatility model for asset returns", *Journal of Business and Economic Statistics*, **14**, 429-34.

Hasbrouck, J. (1995), "One security, many markets: Determining the contributions to price discovery", *Journal of Finance*, **50**, 1175–1198.

Hayashi, F. (2000), *Econometrics*, Princeton: Princeton University Press.

Hillebrand, E. and M.C. Medeiros (2008), "Asymmetries, breaks, and long-range dependence: an estimation framework for time series of daily realized volatility", Unpublished paper, Department of Economics, Pontifical Catholic University of Rio de Janeiro.

Kitagawa, G. (1987), "Non-Gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, **82**, 1032－1063 (with discussion).

Malmsten, H. and T. Teräsvirta (2004), "Stylized facts of financial time series and three popular models of volatility", Working Paper Series in Economics and Finance 563, Stockholm School of Economics.

McAleer, M. (2005), "Automated inference and learning in modeling financial volatility", *Econometric Theory*, **21**, 232 – 261.

McAleer, M. and M. Medeiros (2008a), "Realized volatility: A review", *Econometric Reviews*, **27**, 10－45.

McAleer, M. and M. Medeiros (2008b), "A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries", *Journal of Econometrics*, **147**, 104-119.

Meddahi, N. (2002), "A theoretical comparison between integrated and realized volatility", *Journal of Applied Econometrics*, **17**, 479 – 508.

Merton, R.C. (1980), "On estimating the expected return on the market: An exploratory investigation", *Journal of Financial Economics*, **8**, 323 – 361.

Morgan, J.P. (1996), *J.P. Morgan/Reuters Riskmetrics – Technical Document*. J.P. Morgan, New York.

Nelson, D. B. (1990), "ARCH models as diffusion approximations", *Journal of Econometrics*, **45**, 7–38.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, **59**, 347－70.

So, M.K.P. (1999), "Time series with additive noise", *Biometrika*, **86**, 474－482.

Tanaka, K. (2002), "A unified approach to the measurement error problem in time series models", *Econometric Theory*, **18**, 278 – 296.

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

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**, 1394 – 1411.

Table 1: Monte Carlo Results of the QML Estimators for the Two-Component Model for T=1000

(a) Finite Sample Properties

Parameter	True Value	With Measurement Errors			Neglecting Measurement Errors		
		Mean	St.Dev.	RMSE	Mean	St.Dev.	RMSE
ϕ_1	0.98	0.9708	0.0135	0.0163	0.9664	0.0157	0.0208
σ_1	0.10	0.1029	0.0207	0.0209	0.1130	0.0226	0.0260
ϕ_2	0.40	0.4065	0.1109	0.1111	0.2920	0.0535	0.1206
σ_2	0.40	0.4135	0.0601	0.0616	0.4752	0.0283	0.0803
σ	1.00	1.1487	0.1538	0.2139	1.1491	0.1539	0.2143
σ_w	0.2351	01814	0.1304	0.1410			

Note: The true value of σ_w is obtained by simulation.

(b) One-Step-Ahead Predictions

Statistic	With Measurement Errors			Neglecting Measurement Errors		
	$\alpha_{1,T+1}$	$\alpha_{2,T+1}$	$\ln IV_{T+1}$	$\alpha_{1,T+1}$	$\alpha_{2,T+1}$	$\ln IV_{T+1}$
MAE	0.2449	0.3361	0.3621	0.2453	0.3375	0.3625
RMSE	0.3043	0.4199	0.4505	0.3048	0.4221	0.4512

(c) Mincer-Zarnowitz Regression

Method	With Measurement Errors		Neglecting Measurement Errors	
	R^2	F test	R^2	F test
MZ^{IV}	0.5393	0.6992	0.5378	0.7335
MZ^{RV}	0.4313	0.1972	0.4329	0.1823

Note: 'F test' denotes the p-value of the (robust) χ^2 test for the null hypothesis of $\alpha=0$ and $\beta=1$ ($\gamma=0$ and $\delta=1$) in the MZ equation for $\ln RV_t$ ($\ln IV_t$).

Table 2: Monte Carlo Results of the QML Estimators for the ARFIMA Model for *T*=1000

(a) Finite Sample Properties

Parameter	True Value	With Measurement Errors			Neglecting Measurement Errors		
		Mean	St.Dev.	RMSE	Mean	St.Dev.	RMSE
d	0.40	0.4093	0.0595	0.0602	0.3764	0.0424	0.0486
σ_η	0.40	0.3802	0.0727	0.0754	0.4421	0.0244	0.0486
ϕ	-0.10	-0.0493	0.1529	0.1615	-0.1045	0.0476	0.0478
σ_w	0.1537	01503	0.1320	0.1321			

Note: The true value of σ_w is obtained by simulation.

(b) One-Step-Ahead Predictions of $\ln IV_{T+1}$

Statistic	With Measurement Errors	Neglecting Measurement Errors
MAE	0.2953	02959
RMSE	0.3724	0.3728

(c) Mincer-Zarnowitz Regression

Method	With Measurement Errors		Neglecting Measurement Errors	
	R^2	*F* test	R^2	*F* test
MZ^{IV}	0.5012	0.4992	0.5003	0.4875
MZ^{RV}	0.4630	0.4154	0.4620	0.3964

Note: '*F* test' denotes *P*-value of the (robust) χ^2 test for the null hypothesis of $\alpha = 0$ and $\beta = 1$ ($\gamma = 0$ and $\delta = 1$) in the *MZ* equation for $\ln RV_t$ ($\ln IV_t$).

Table 3: Monte Carlo Results of the QML Estimators for the ARFIMA Model for T=2000

(a) Finite Sample Properties

Parameter	True Value	With Measurement Errors			Neglecting Measurement Errors		
		Mean	St.Dev.	RMSE	Mean	St.Dev.	RMSE
d	0.40	0.4112	0.0447	0.0461	0.3773	0.0278	0.0359
σ_η	0.40	0.3830	0.0652	0.0673	0.4386	0.0186	0.0428
ϕ	-0.10	-0.0631	0.1079	0.1141	-0.1044	0.0332	0.0335
σ_w	0.1537	01410	0.1277	0.1283			

Note: The true value of σ_w is obtained by simulation.

(b) One-Step-Ahead Predictions of $\ln IV_{T+1}$

Statistic	With Measurement Errors	Neglecting Measurement Errors
MAE	0.3161	0.3165
RMSE	0.3944	0.3953

(c) Mincer-Zarnowitz Regression

Method	With Measurement Errors		Neglecting Measurement Errors	
	R^2	F test	R^2	F test
MZ^{IV}	0.4291	0.5242	0.4263	0.5511
MZ^{RV}	0.3912	0.8674	0.3878	0.8681

Note: 'F test' denotes P-value of the (robust) χ^2 test for the null hypothesis of $\alpha = 0$ and $\beta = 1$ ($\gamma = 0$ and $\delta = 1$) in the MZ equation for $\ln RV_t$ ($\ln IV_t$).

Table 4: Monte Carlo Results of the Estimators for the HAR Model for T=1000

(a) Finite Sample Properties

Parameter	True Value	With Measurement Errors			Neglecting Measurement Errors		
		Mean	St.Dev.	RMSE	Mean	St.Dev.	RMSE
β_1	0.80	0.7994	0.1384	0.1384	0.4411	0.0844	0.3687
β_2	0.10	0.0912	0.1286	0.1289	0.3904	0.0866	0.3030
β_3	0.05	0.0481	0.0308	0.0308	0.0725	0.0462	0.0514
σ_η	0.25	0.2607	0.0524	0.0535	0.40324	0.0458	0.1599
σ	1.00	0.9841	0.1496	0.1504			
β_0	0.00				-0.0041	0.0158	0.0163
σ_w	0.2638	0.2472	0.0680	0.0699			

Note: The true value of σ_w is obtained by simulation. $\beta_0 = (1-\beta_1-\beta_2-\beta_3)\ln\sigma$.

(b) One-Step-Ahead Predictions of $\ln IV_{T+1}$

Statistic	With Measurement Errors	Neglecting Measurement Errors
MAE	0.2216	0.2216
RMSE	0.2794	0.2794

(c) Mincer-Zarnowitz Regression

Method	With Measurement Errors		Neglecting Measurement Errors	
	R^2	F test	R^2	F test
MZ^{IV}	0.8001	0.0001*	0.8022	0.0005*
MZ^{RV}	0.7236	0.056	0.7235	0.150

Note: 'F test' denotes P-value of the (robust) χ^2 test for the null hypothesis of $\alpha=0$ and $\beta=1$ ($\gamma=0$ and $\delta=1$) in MZ equation for $\ln RV_t$ ($\ln IV_t$).

Table 5: Partially and Fully Corrected R^2

(a) Forecasts $\hat{h}^{(i)}_{T+1|T}$ $(i=1,2,\ldots,500)$

Descriptive Statistics	Two Component		ARFIMA		HAR	
	With ME	Neglecting ME	With ME	Neglecting ME	With ME	Neglecting ME
R^2 for MZ^{IV}	0.5393	0.5378	0.4291	0.4263	0.8001	0.8022
R^2 for MZ^{RV}	0.4313	0.4329	0.3912	0.3878	0.7236	0.7232
$\hat{V}(\ln RV_t)/\hat{V}(\ln IV_t)$	1.6086	1.6086	1.2537	1.2537	1.1247	1.1247
$(\hat{\delta}/\hat{\beta})^2$	0.7773	0.7723	0.8748	0.8770	0.9832	0.9856
Partially Corrected R^2	0.6938	0.6964	0.4905	0.4861	0.8138	0.8137
Fully Corrected R^2	0.5393	0.5378	0.4291	0.4263	0.8001	0.8023

Note: ME denotes 'Measurement Errors'.

(b) Forecasts $\hat{h}_{T+j|T+j-1}$ $(j=1,2,\ldots,500)$

Descriptive Statistics	Two Components		ARFIMA		HAR	
	With ME	Neglecting ME	With ME	Neglecting ME	With ME	Neglecting ME
R^2 for MZ^{IV}	0.5179	0.5165	0.4551	0.4550	0.8385	0.8355
R^2 for MZ^{RV}	0.4889	0.4872	0.4166	0.4165	0.7603	0.7584
$\hat{V}(\ln RV_t)/\hat{V}(\ln IV_t)$	1.1103	1.1103	1.1641	1.1641	1.1358	1.1358
$(\hat{\delta}/\hat{\beta})^2$	0.9541	0.9548	0.9385	0.9385	0.9709	0.9700
$(\tilde{\delta}/\hat{\beta})^2$	0.9364	0.9404	0.8976	0.8976	0.9486	0.9465
Partially Corrected R^2	0.5428	0.5410	0.4849	0.4849	0.8636	0.8614
Fully Corrected R^2	0.5083	0.5087	0.4353	0.4353	0.8192	0.8153

Table 6: Descriptive Statistics for S&P 500

Stock	Mean	St.Dev.	Skewness	Kurtosis	LM test	
					AR(1)	ARFIMA(1,*d*,0)
Return	0.0296	1.1227	-0.1781	6.3870		
RV	1.0434	1.2010	4.4862	33.737		
ln*RV*	-0.3074	0.7774	0.5873	3.2424	17.513*	15.053*

Note: The LM test of Tanaka (2002) is a test of measurement errors. The test statistic has the standard normal distribution under the null hypothesis of no measurement error, and rejects the null hypothesis if the calculated value exceeds the right side critical value. '*' indicates significance at 5%.

Table 7: QML Estimates of the Two-Component Model

QML	ϕ_1	σ_1	ϕ_2	σ_2	θ_2	σ	σ_w	Q Log-like
With ME	0.9922	0.0743	0.8044	0.1839		0.7150	0.3839	-1907.60
	(0.0046)	(0.0226)	(0.1038)	(0.0226)		(0.1248)	(0.0131)	
Neglecting ME	0.9824	0.1202	0.1391	0.4254		0.7270		-1910.79
	(0.0043)	(0.0097)	(0.0288)	(0.0080)		(0.0925)		
With MA(1)	0.9918	0.0769	0.7898	0.4562	-0.5549	0.7136		-1907.69
	(0.0045)	(0.0215)	(0.1069)	(0.0131)	(0.0834)	(0.1226)		

Note: Standard errors are in parentheses. ME denotes 'Measurement Errors'.

Table 8: QML Estimates of the ARFIMA(1,*d*,0) Model

QML	d	σ_η	ϕ	θ	σ_w
With ME	0.4934	0.1651	0.5761		0.3931
	(0.0990)	(0.0596)	(0.2901)		(0.0195)
Neglecting ME	0.4766	0.4791	-0.1358		
	(0.0192)	(0.0064)	(0.0255)		
With MA(1)	0.4995	0.4783	0.1715	0.3385	
	(0.0319)	(0.0064)	(0.0991)	(0.1126)	

Note: Standard errors are in parentheses. ME denotes 'Measurement Errors'.

Table 9: Estimates of the HAR Model

Method	β_1	β_2	β_3	θ	σ_η	σ	β_0	σ_w	Q Log-like
With ME	0.5435	0.3002	0.1170		0.3031	0.7241		0.3339	-1907.10
	(0.1281)	(0.1055)	(0.0352)		(0.0497)	(0.1248)		(0.0340)	
Neglecting	0.2424	0.4585	0.2343		0.4793		-0.0184		-1916.04
ME	(0.0227)	(0.0384)	(0.0330)		(0.0064)		(0.0101)		
With	0.5950	0.2124	0.1487	-0.3115	0.4785	0.7524			-1911.49
MA(1)	(0.1200)	(0.0876)	(0.0371)	(0.1091)	(0.0064)	(0.1071)			

Note: ME denotes 'Measurement errors'. Standard errors are in parentheses. $\beta_0 = (1-\beta_1-\beta_2-\beta_3)\ln\sigma$.

Table 10: Results for One-Step-Ahead Forecasts

Method	Two-Component Models			ARFIMA(1,*d*,0) models			HAR models		
	Partially Corrected R^2	Fully Corrected R^2	*F* test	Partially Corrected R^2	Fully Corrected R^2	*F* test	Partially Corrected R^2	Fully Corrected R^2	*F* test
With ME	0.308	0.384	0.000*	0.306	0.386	0.032*	0.276	0.347	0.000*
Neglecting ME	0.305	0.380	0.000*	0.307	0.386	0.016*	0.301	0.377	0.000*
With MA(1)	0.261	0.330	0.000*	0.263	0.343	0.000*	0.305	0.383	0.000*

Note: ME denotes 'Measurement errors'. The table reports the out-of-sample forecasting results for daily realized volatility. The partially and fully corrected R^2 values are corrected by the methods of Andersen, Bollerslev and Meddahi (2005) and the current paper, respectively. '*F* test' denotes the *p*-value of the (robust) χ^2 test for the null hypothesis of $H_0: \alpha = 0$ and $\beta = 1$ in the *MZ* equation.

Table 11: Forecast Encompassing Tests for One-Step-Ahead Prediction

$$f_{ct} = (1-\lambda) f_{1t} + \lambda f_{2t} \quad (0 \le \lambda \le 1)$$

f_{1t} \ f_{2t}	Two Component		ARFIMA		HAR	
	$i=1$	$i=2$	$i=1$	$i=2$	$i=1$	$i=2$
Two Component						
$i=1$	---	0.098	0.003	0.013	0.585	0.667
$i=2$	0.004	---	0.000	0.001	0.829	0.506
ARFIMA						
$i=1$	0.389	0.428	---	0.606	0.234	0.503
$i=2$	0.786	0.782	0.264	---	0.244	0.930
HAR						
$i=1$	0.000	0.000	0.000	0.000	---	0.000
$i=2$	0.004	0.147	0.000	0.000	0.665	---

Note: The entries are p-values for the null hypothesis $\lambda = 0$. We denote i as the model with measurement error $(i=1)$, and the model neglecting measurement error $(i=2)$.

Table 12: Results for Multi-Step-Ahead Forecasts

Method	Two-Component Models		ARFIMA(1,*d*,0) models		HAR models	
	Partially Corrected R^2	Fully Corrected R^2	Partially Corrected R^2	Fully Corrected R^2	Partially Corrected R^2	Fully Corrected R^2
5 step-ahead						
With ME	0.296	0.366	0.287	0.366	0.299	0.378
Neglecting ME	0.302	0.370	0.289	0.369	0.281	0.358
With MA(1)	0.277	0.351	0.287	0.372	0.290	0.370
10 step-ahead						
With ME	0.289	0.357	0.261	0.338	0.296	0.373
Neglecting ME	0.301	0.366	0.263	0.339	0.267	0.339
With MA(1)	0.260	0.332	0.268	0.351	0.270	0.343
20 step-ahead						
With ME	0.280	0.343	0.226	0.297	0.294	0.368
Neglecting ME	0.291	0.352	0.225	0.295	0.272	0.343
With MA(1)	0.248	0.312	0.235	0.312	0.270	0.342

Note: ME denotes 'Measurement errors'. The table reports the out-of-sample forecasting results for daily realized volatility. The partially and fully corrected R^2 values are corrected by the methods of Andersen, Bollerslev and Meddahi (2005) and the current paper, respectively.