Forecasting Volatility using High Frequency Data

PETER REINHARD HANSEN   ASGER LUNDE
Stanford University, USA   *Aarhus University & CREATES, Denmark*

Contents

1   Introduction

Over the last decade the easy access to high frequency financial data has spurred much activity in financial economics. In particular, our knowledge of financial volatility and its dynamic properties is an area that has gained tremendously from the availability of high frequency data. The introduction of empirical estimators of the quadratic variation to measure the ex post variation of asset prices is a prime example, and such estimators have had an immense influence on financial econometrics. There is now a range of volatility estimators that are computed from high frequency data, and we shall refer to such as realized measures.

For the problem of forecasting volatility, high frequency data have been valuable in a number of ways. The main reason that high-frequency data can greatly improve the forecast accuracy is simply that volatility is highly persistent, so that a more accurate measure of current volatility, which high frequency data provide, is valuable for forecasting future volatility. Next, we list six ways that high frequency data have improved volatility forecasting.

1. High frequency data improve our understanding of the dynamic properties of volatility which is key for forecasting.

2. Realized measures are valuable predictors of future volatility in reduced form models.

3. Realized measures have enabled the development of new volatility models that provide more accurate forecasts.

4. High frequency data have improved the evaluation of volatility forecasts in important ways.

5. Realized measures can facilitate and improve the estimation of complex volatility models, such as continuous time volatility models. The resulting reduction in the parameter uncertainty will improve predictions based on such models.

6. High frequency data have improved our understanding of the driving forces of volatility and their relative importance. For instance, high frequency data have enabled a detailed analysis of news announcements and their effect on the financial markets.

The current interest in high frequency data was largely spurred by Andersen and Bollerslev (1998*b*) who used the realized variance to show that standard volatility models deliver accurate forecasts. The realized variance is the most commonly used realized measure, and the realized variance is simply constructed by adding up squared intraday returns. There exists an extensive statistical theory for this object that is

derived in papers by Barndorff-Nielsen and Shephard (2002*a*), Meddahi (2002), Andersen, Bollerslev, Diebold and Labys (2003) and Mykland and Zhang (2009).

1.1 High Frequency Data and the Dynamic Properties of Volatility

Several studies have estimated GARCH models using intraday returns, such as 5 minute returns. These studies have documented a strong diurnal pattern in intraday volatility, see e.g. Andersen and Bollerslev (1997*a*,*b*, 1998*a*). Such features are important for intraday volatility forecasting. In this chapter we focus on volatility forecasting over longer horizons, so we shall not attempt to review this literature. Nor shall we attempt to review the related literature on duration based models. The interested reader may consult Ghysels and Jasiak (1997), Engle and Russell (1998), Engle (2000), Engle and Sun (2007), Racicot, Théoret and Coön (2008) and Tay, Ting, Tse and Warachka (2009).

Many of the papers that we will review in this chapter aim at uncovering the dynamic properties of daily volatility by aid of high frequency data. Here an AR(FI)MA framework is commonly used because the ARMA structure arises naturally in many models. For example, when the spot variance is modeled as a linear combination of p independent OU processes, Barndorff-Nielsen and Shephard (2002*a*) showed that integrated and realized variances both are ARMA(p,p) processes. Meddahi (2003) derive the ARMA representation of both integrated and realized variances when the spot variance depends linearly on two autoregressive factors. These papers highlight the fact that it is important to discriminate between the realized measure of volatility and the underlying population quantity. Hansen and Lunde (2010) demonstrate that measurement errors of the magnitude found in realized measures of daily volatility cause the autocorrelation function of the observed time series to look distinctively different from that of latent volatility. We shall review part of this literature in Section 3.

1.2 High Frequency Data and Volatility Forecast Evaluation

The paper by Andersen and Bollerslev (1998*b*) was a response to a critique of GARCH models. Several papers had noted that GARCH models were unable to explain much of the variability in squared returns when evaluated out-of-sample, despite the fact that GARCH models had good in-sample fit, see e.g. Jorion (1995) and Figlewski (1997). The false conclusion that was drawn from this observation was that GARCH models were of little practical value. By using the realized variance, which is a more accurate measure of volatility than squared returns, Andersen and Bollerslev (1998*b*) showed that standard volatility models perform rather well. So the apparent 'poor' performance can be attributed to the fact that the squared return is a very noisy proxy of the conditional variance.

The comparison of volatility forecasts is a problem for which accurate measures of volatility is of critical importance. Hansen and Lunde (2005*b*) showed that the use of a noisy proxy, in place of the true latent volatility, can severely distort a comparison of volatility models, because the empirical ranking of models may be inconsistent for the true (population) ranking. So an evaluation that is based on squared returns may select an inferior model as the 'best' with a probability that converges to one as the sample size increases. Hansen and Lunde (2005*b*) show that some evaluation criteria are immune to this inconsistency problem, and Patton (2008) characterizes this class of criteria when the proxy can be assumed to be unbiased for the underlying population quantity.

1.3 Volatility Forecasting with Realized Measures

Volatility forecasting using high frequency data can largely be divided into two main approaches that we refer to as *reduced form* volatility forecasting and *model based* forecasting, respectively.

The reduced-form approach covers cases where the realized measures are modeled with a time series model (e.g. ARIMA) and the estimated model is used to produce volatility forecasts. An influential paper in this area is Andersen et al. (2003) who construct and analyze long-memory Gaussian vector autoregressive (VAR) models for a vector of realized variances. This approach has subsequently been used in numerous other studies. A related type of reduced-form volatility forecasts are the regression based forecasts, such as those based on the HAR structure by Corsi (2009) and MIDAS by Ghysels, Santa-Clara and Valkanov (2006). We review this literature in Section 5.

The model-based approach to volatility forecasting is constructed from a model for returns, such as a GARCH type model that specifies the entire distribution of returns. The name of the game for this type of forecasting is to estimate the parametric model and use the estimated model to predict volatility. High-frequency data have been used in two different ways in this context. Either to enhance an existing volatility model by including a realized measure into the model, or to utilize high-frequency based statistics to improve or simplify the estimation of the statistical model. Conventional GARCH models are based on daily returns and go back to the seminal paper by Engle (1982). Blair, Poon and Taylor (2004), Martens (2001, 2002) and Engle (2002) were some of the first to extend the conventional GARCH model to include the realized variance in the equation that defines the dynamic properties of the conditional variance. This idea has recently been extended in a number of interesting ways and we review this literature in Section 6.

1.4 High Frequency Data to Facilitate Estimation of Volatility Models

Continuous time stochastic volatility (SV) models were introduced by Taylor (1982), and high frequency data has been used to facilitate and improve the estimation of SV models. In particular, Barndorff-Nielsen and Shephard (2002*a*), Bollerslev and Zhou (2002), Meddahi (2003) and Corradi and Distaso (2006) have shown how realized measures can be used to estimate the parameters in these models. This idea has been extended and utilized for the purpose of volatility forecasting by Barndorff-Nielsen, Nielsen, Shephard and Ysusi (2004), Andersen, Bollerslev and Meddahi (2004, 2009) and Aït-Sahalia and Mancini (2008). We will not attempt to review this part of the literature in this chapter. However, we want to emphasize that this use of high frequency data is potentially very useful for forecasting, because the reduction in the parameter uncertainty, which is one of the benefits of using high frequency data.

1.5 Outline of this Chapter

The structure of the chapter is as follows. In Section 2, we provide a formal framework in which we can discuss and classify the various forecasting models and methods that we review. We also relate this framework to population quantities, such as the quadratic variation and the realized variance. Moreover, we review the most popular approaches to ex-post measurement of volatility. This is followed by Section 3, which is concerned with the statistical properties of volatility. We emphasize the importance of discriminating between the realized measure of volatility and the underlying population quantity. This has distinctive implications for inferring the dynamic properties of the volatility process, which, in turn, has implications for forecasting. Section 4 concerns evaluation of volatility forecasts, which is an area where realized measures have proven extremely valuable for delivering precise ex post measurement of the target of interest. The next two sections review the two leading approaches to volatility forecasting. In Section 5 we discuss reduced form forecasts that are constructed from time series of realized measures, and in Section 6 we consider model based forecasts, that are constructed from a parametric model for the return distribution. Finally, Section 7 provides some concluding remarks.

2 High-Frequency Data and Realized Measures of Volatility

We start by introducing the notational framework. Let $Y(t)$ denote the logarithmic price of some asset, where $Y(t)$ is a continuous time stochastic process. In the leading case the time scale is a day, so that

$$y_t = Y(t) - Y(t-1),$$

is a daily return. Given a partition, $t - 1 = \tau_0 < \tau_1 < \cdots < \tau_{N_t} = t$, the intraday returns are defined by

$$y_{t,i} = Y(\tau_i) - Y(\tau_{i-1}), \qquad \text{for } i = 1, \ldots, N_t.$$

Note that even though we have an underlying continuous time model in mind we will explicitly formulate the forecasting problem to be concerned with the volatility of discretely sampled observations. We will rely on the existence of the conditional second moment of y_t, and denote the conditional mean and variance of y_t by $\mu_t = \mathrm{E}\,[y_t|\mathcal{F}_{t-1}]$, and $\sigma_t^2 = \mathrm{var}\,[y_t|\mathcal{F}_{t-1}]$, respectively, where $\mathcal{F}_t$ is the information set available at time t. The conditional variance, σ_t^2, is often the object of interest. The richness of the filtration, $\mathcal{F}_t$, is key for the definition of σ_t^2, and the forecasting problem. In the *classical* approach to volatility forecasting $\mathcal{F}_t$ is typically generated by sparse daily information, such as opening or closing prices. In this chapter we specifically review and discuss research where $\mathcal{F}_t$ also comprises high frequency information. So $\mathcal{F}_t$ contains observations such as intraday transaction prices, bid/ask quotes, trading volume and depth. One strategy for volatility forecasting is to parameterize μ_t and σ_t^2 or the full conditional distribution of y_t.

Another approach is based on so-called realized measures. To motivate this approach it is helpful to make the connection to the underlying continuous time model. So let the return process be written as standard Ito process, that is

$$dY(t) = \mu(t)dt + \sigma(t)\mathrm{d}W(t), \tag{1}$$

where $\mu(t)$ is the drift, $\sigma(t)$ denotes the *spot* volatility and $W(t)$ is a standard Brownian motion. This model is also known as a stochastic volatility (SV) model. It has been analyzed extensively and the results that we present here are from Barndorff-Nielsen and Shephard (2002*a*,*b*). An SV model is a special type of semimartingale that has some very convenient properties. In particular, when $\mu(t)$ and $\sigma(t)$ are jointly independent from $W(t)$, we have

$$y_t|\mu_t, IV_t \sim N(\mu_t, IV_t),$$

where

$$\mu_t = \int_{t-1}^{t} \mu(s)\mathrm{d}s \qquad \text{and} \qquad IV_t \equiv \int_{t-1}^{t} \sigma^2(s)\mathrm{d}s.$$

So in this framework the integrated variance, IV_t, is the population measure of actual return variance.

The quadratic variation of a stochastic process over the interval $[t - 1, t]$ is given by

$$QV_t \equiv \plim_{N \to \infty} \sum_{j=1}^{N} \{Y(\tau_j) - Y(\tau_{j-1})\}^2,$$

where $t - 1 = \tau_0 < \tau_1 < \ldots < \tau_N = t$ is a mesh that satisfies $\max_{1 \leq j \leq N} |\tau_j - \tau_{j-1}| \to 0$ as $N \to \infty$. The empirical counter part of QV_t is called the realized variance which is simply the sum of the squared observed interperiod returns,

$$RV_t \equiv \sum_{j=1}^{N_t} y_{t,j}^2 \qquad \text{where } y_{t,j} = Y(\tau_j) - Y(\tau_{j-1}), \quad j = 1, \ldots, N_t.$$

It is a well know result that RV_t consistently estimates the corresponding QV_t for all semimartingales, and for SV models it holds that $QV_t = IV_t$. So for SV models the population measure, IV_t, is also consistently estimated by RV_t.

The RV_t can also be related to the conditional variance, σ_t^2, but will only be a consistent estimator of σ_t^2 in special cases. For instance, if $\{\mu(s), \sigma(s)\}_{s=t-1}^t$ is $\mathcal{F}_{t-1}$-measurable we have $IV_t = \sigma_t^2$, but more generally we can view the realized variance as a noisy estimate of σ_t^2 where the noise has two sources. One source is the sampling error, $RV_t - IV_t$, and the other source arises from the random nature of volatility $IV_t - \sigma_t^2$.

The literature has established relationships between realized measures and population measures to a higher level of generality than we have spelled out here. It is such relationships that have motivated the use of realized measures in volatility modeling, forecasting, and forecast evaluation.

Sometimes we use volatility quantities that span more or less than one day, e.g. the volatility over the interval $[t, t + h]$. For such quantities we use the subscript "t, h" for example, $IV_{t,h}$ is defined to be $\int_t^{t+h} \sigma^2(s)\mathrm{d}s$.

2.1 Measuring Volatility

The fact that the population measures of volatility – such as the conditional variance and the IV_t – are latent variables, complicates forecasting and makes it difficult to evaluate the performance of volatility models. A common resolution to this problem is to substitute a proxy for the true volatility and evaluate the models by comparing their predicted volatility to the proxy. Given the results by Andersen and Bollerslev (1998*b*) it is important to use an accurate proxy when evaluating the merits of volatility models. Thus if we let ε_t be defined to be the difference between the realized measure, RM_t, and the population measure of interest, then we would want ε_t to have a distribution that is tightly concentrated about zero. We will discuss the importance of this in a forecasting exercise. It enables the estimation of a predictive equation that precisely captures the dynamics of $\mathrm{var}(y_t|\mathcal{F}_{t-1})$, and it is needed to make consistent model evaluation. We discuss these topics briefly in the following subsections.

2.2 Realized Measures of Volatility

In a widely cited paper Merton (1980) noted that in a noiseless environment the variance of returns can be estimated much more precisely from realized returns than the expected return. This is essentially the insight that Barndorff-Nielsen and Shephard (2002*a*) extend to the SV model in (1). They show that RV_t is consistent for IV_t (which equals $\mathrm{var}(y_t|\mathcal{F}_t)$) and that

$$\sqrt{N}(RV_t - IV_t) \overset{law}{\to} N(0, 2IQ_t), \qquad \text{as } N \to \infty,$$

where $QV_t \equiv \int_{t-1}^{t} \sigma^4(s)\mathrm{d}s$. The asymptotic variance is here stated for the case with equidistant sampling, $\tau_i - \tau_{i-1} = \frac{1}{N}$. The full theoretical foundation for this result is nicely reviewed in Barndorff-Nielsen and Shephard (2007, sec. 3.1).

Taken literally, this theory suggests that we should sample prices as often as possible. This would amount to estimate IV_t by RV_t from tick-by-tick data. However, as was noted in Merton (1980) "in practice, the choice of an even-shorter observation interval introduces another type of error which will 'swamp' the benefit [...] long before the continuous limit is reached". The modern terminology for this phenomenon is known as *market microstructure effects* that cause the observed market price to deviate from the efficient price, so that only the latter has the semi-martingale property. This motivated the idea of viewing observed prices, $X(t)$, as noisy measures of the latent true price process, $Y(t)$. So for estimating IV_t we will have observations

$$X(\tau_0), ..., X(\tau_N), \quad t - 1 = \tau_0 < \tau_1 < ... < \tau_N = t,$$

where $X(\tau_j)$ is a noisy version of $Y(\tau_j)$, that is

$$X(\tau_j) = Y(\tau_j) + U(\tau_j).$$

We think of U as noise that can be due to, for example, liquidity effects, bid/ask bounce and misrecordings. A classical example of how market frictions distort efficient prices is Roll (1984), that shows how the presence of a bid–ask spread leads to a negative first-order correlation in observed price changes. Specific models for U have been suggested in this context by, for example, Zhou (1996), Hansen and Lunde (2006), Li and Mykland (2007) and Diebold and Strasser (2008). See Hansen and Lunde (2006) for a comprehensive analysis of the statistical properties of U.

Following Zhou (1996) several methods have been developed for estimating the integrated variance and the quadratic variation in the presence of noise. Leading references include Andersen, Bollerslev, Diebold and Labys (2000), Bandi and Russell (2008), Zhang, Mykland and Aït-Sahalia (2005), Zhang (2006),

Jacod, Li, Mykland, Podolskij and Vetter (2009), Barndorff-Nielsen, Hansen, Lunde and Shephard (2008) and Hansen and Horel (2009). For instance the realized kernel estimator takes the form

$$RK_t = \sum_{h=-H}^{H} k\left(\tfrac{h}{H}\right) \gamma_h, \quad \gamma_h = \sum_{j=|h|+1}^{n} x_j x_{j-|h|}, \quad x_j = X(\tau_j) - X(\tau_{j-1}) \tag{2}$$

where $k(x)$ is a kernel weight function, such as the Parzen kernel. Details regrading the statistical properties of this estimator are presented in Barndorff-Nielsen, Hansen, Lunde and Shephard (2010) that also applies to multivariate price processes. For additional details about the implementation see Barndorff-Nielsen, Hansen, Lunde and Shephard (2009).

3 Statistical Properties of Volatility

Careful data cleaning is one of the most important aspects of volatility estimation from high-frequency data. This task has been given special attention in e.g. (Dacorogna, Gencay, Müller, Olsen and Pictet, 2001, chapter 4), Falkenberry (2001), Hansen and Lunde (2006), Brownless and Gallo (2006) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2009). Interestingly, Hansen and Lunde (2006) show that tossing out a large number of observations can improve the accuracy of volatility estimators. This result may seem counter intuitive at first, but the reasoning is fairly simple. An estimator that makes optimal use of all data will typically put high weight on accurate data and be less influenced by the least accurate observations. The generalized least squares (GLS) estimator in the classical regression model is a good analogy. On the other hand, the precision of the standard least squares estimator can deteriorate when relatively noisy observations are included in the estimation. So the inclusion of poor quality observations can cause more harm than good to the least squares estimator and this is the relevant comparison to the present situation.

The most commonly used data source in academic literature is the TAQ data base, NYSE (2007). Several other data sources are surveyed in Wood (2000). Regardless of the data source, a good understanding of the microstructure of the financial market from which the data originates is essential for constructing efficient data filters. Some fairly comprehensive descriptions are provided in Comerton-Forde and Rydge (2004) and (Hasbrouck, 2007, chapter 2 and appendix about U.S. equity markets). Also of interest are Andersen (2000) and Ghysels (2000) who provide valuable discussions of various aspects of analyzing high frequency data.

3.1 Assessing the Persistence of Volatility

The reduced form approach that we will discuss in the next section relies on daily time series of realized measures for forecasting the future latent volatility over some time span. Each element of these time

series can be viewed as a noisy estimate of the latent volatility. Despite the progress that has been made in estimating financial volatility from high-frequency data, it is still important to discriminate between the realized measure of volatility and the underlying population quantity. Even with the most accurate estimators of daily volatility that utilize thousands of high-frequency prices, the standard error for a single estimate is rarely less than 10% of the point estimate, see e.g. Barndorff-Nielsen et al. (2008). Measurement errors of this magnitude cause the autocorrelation function of the observed time series to look distinctively different from that of the underlying time series. Hansen and Lunde (2010) propose a novel estimator of the autocorrelation function (ACF) for the latent time series. This estimator is informative about the ACF of the latent process, in particular when the latter is a persistent process.

Let $y_t = \log(\mathrm{IV}_t)$ and $x_t = \log(\mathrm{RM}_t)$, where RM_t is a realized measure. If y_t were observed then we could estimate the autocorrelations, $\rho_y(h)$, by the empirical autocorrelations, $\mathrm{ACF}_y(h)$, and we would have $\rho_y(h) = \mathrm{plim}_{n\to\infty} \mathrm{ACF}_y(h)$. The corresponding autocorrelations for the observed time series $\rho_x(h)$ are inconsistent for $\rho_y(h)$ due to the errors-in-variables problem. Hansen and Lunde (2010) suggest estimating $\rho_y(h)$ using an IV estimator

$$\mathrm{ACF}_x^*(h) = \frac{\sum_{t=1}^{n} z_t x_{t+h}}{\sum_{t=1}^{n} z_t x_t},$$

where the instrument z_t could be a lagged value of $x_{t-j} - \bar{x}_j$ or a linear combination of these such as $z_t = Z_t'\alpha$ with $Z_t = (x_{t-j} - \bar{x}_j, \ldots, x_{t-J} - \bar{x}_J)'$. $\mathrm{ACF}_x^*(h)$, $h = 1, 2, \ldots$, is called an approximate autocorrelation function, because it is not consistent for $\rho_y(h)$ in general. However if the process, y_t, is highly persistent, then $\mathrm{ACF}_x^*(h)$ will closely approximate $\rho_y(h)$. In particular if y_t is an AR(1) process then $\mathrm{ACF}_x^*(h)$ is consistent for $\rho_y(h)$.

In the following we present the empirical results of applying the $\mathrm{ACF}_x^*(h)$ to realized measures of volatility for a stock in the Dow-Jones Industrial Average (DJIA). The sample period runs from January 3, 2002 to July 31, 2009 with a total of 1,907 trading days for most of the series. The high frequency data used to compute the realized measures of volatility is the collection of trades and quotes recorded on the NYSE. These high frequency data were extracted from the TAQ database through the Wharton Research Data Services (WRDS) system. Both the realized kernel and the realized variance are computed with transaction prices that have been cleaned by the step-by-step cleaning procedure proposed by Barndorff-Nielsen, Hansen, Lunde and Shephard (2009).

The first time series is based on realized variance, computed with 30 minute returns, the other is computed with the realized kernel estimator implemented as detailed in Barndorff-Nielsen, Hansen, Lunde and Shephard (2009). We can view both series as noisy proxies of the same population quantity, the underlying quadratic variation. We compute the approximate autocorrelations ACF_x^* using the two-stage least squares

estimator presented above, where seven lags, $x_{t-4}, \ldots, x_{t-10}$, are used as instrumental variables.

Figure 1 displays the estimated autocorrelation functions for the realized measures of volatility computed with high frequency returns on INTC (Intel). The two upper curves are the ACF_x^* based on the realized kernel and the realized variance. The approximate ACF is designed to reflect the the persistence of the latent time series, which – in this case – is the same underlying volatility process for both time series. As should be expected the ACF^*-estimates for the two time series are quite similar. The two lower curves in Figure 1 are computed with the traditional empirical autocorrelation function, ACF_x, and the realized kernel and the realized variance result in distinctly different ACF_x-estimates. The difference between the two simply reflects that the measurement errors in the two series are different.
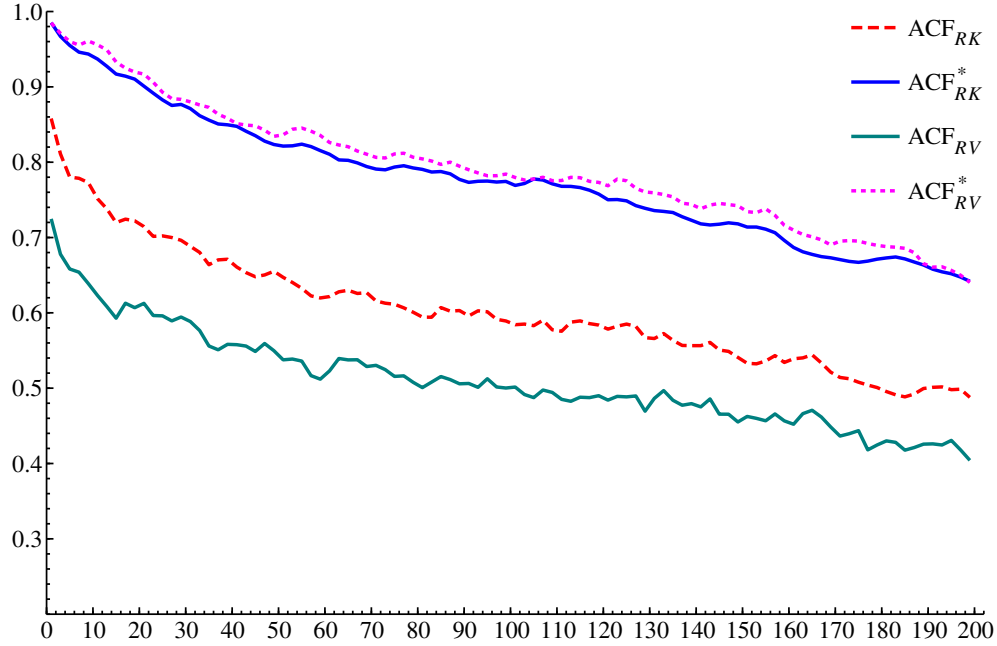

Figure 1: The empirical and approximate autocorrelation functions computed with realized measures of volatility for INTC. The approximate ACF* better reflects the autocorrelation function of latent volatility. While the conventional ACFs are quite different the two ACF*-estimates are in agreement.

The traditional ACFs suggest that the realized kernel is somewhat more persistent than the realized variance. A key point is here that the discrepancy between the two ACFs is induced by the realized variance being a less accurate estimator of the latent volatility, and that neither of the conventional autocorrelation functions properly reflects the persistence of the population measure of volatility. The persistence is better assessed with the approximate estimation of the autocorrelation function, ACF^*, that produces very

similar estimates for the realized kernel and the realized variance.

This analysis is tangible evidence of the measurement error in the series of realized measure as proxies for latent volatility. Let us illustrate implications of this for modelling and forecasting purposes. Consider a simple AR(p) specification for the latent time series, y_t, and treat the observed volatility, x_t, as a noisy and possibly biased estimate of y_t. Specifically, if

$$x_t = y_t + \xi + \eta_t \quad \text{and} \quad \varphi(L)(y_t - \delta) = \varepsilon_t,$$

then the important implication for the observed time series is that

$$\varphi(L)\left(x_t - \delta - \xi\right) = \varepsilon_t + \varphi(L)\eta_t.$$

This shows that x_t is an ARMA process with the exact same autoregressive polynomial as y_t. It also stresses the importance of introducing moving average terms in reduced form volatility equations. In the context of time series of volatility this result was pointed out Barndorff-Nielsen and Shephard (2002*b*) and Meddahi (2003).

4   The use of Realized Measures in Forecast Evaluations and Comparisons

A critique of conventional GARCH models was building momentum in the mid nineties. A series of papers had argued that GARCH models were of little practical value. This (false) conclusion was based on the empirical observation that GARCH models were unable to explain much of the variability in squared returns when evaluated out-of-sample, despite the fact that GARCH models had good in-sample fit, see e.g. Jorion (1995) and Figlewski (1997).

This critique was forcefully rebutted by Andersen and Bollerslev (1998*b*). By using the realized variance, which is a more accurate measure of volatility than squared returns, Andersen and Bollerslev (1998*b*) showed that standard volatility models perform rather well. So the apparent 'poor' performance can be attributed to the fact that the squared return is a very noisy proxy of the conditional variance. The paper by Andersen and Bollerslev (1998*b*) sparked the interest for using realized measures in the context of volatility forecasting.

Another important use of realized measures in forecast evaluations and comparisons was presented in Hansen and Lunde (2005*b*). They showed that the evaluation based on a proxy for the latent volatility can cause the empirical ranking to be inconsistent for the true population ranking. So the empirical ranking of volatility models can mislead practitioners to conclude that an inferior model is the 'best'. Conditions that ensure the equivalence of the rankings that are induced by the true volatility and the proxy is derived,

and some of the popular criteria for evaluating and comparing volatility models are shown not to satisfy the necessary conditions. Some simulation experiments and an empirical comparison of volatility models based on IBM stock returns provide an additional argument for using the realized variance (*RV*) as a proxy for the true volatility. The reason is simple, as *RV* is a more precise measure of volatility than is the squared return, it is less likely to induce an inconsistent ranking of volatility models.

Specifically, let σ_t^2 denote the population measure of volatility and let $\tilde{\sigma}_t^2$ denote the corresponding proxy for σ_t^2, $t = 1, \ldots, n$. Further, let h_t be generic for a model-based measure of volatility and suppose that the 'precision' of h_t is assessed in terms of the expected loss, $\mathrm{E}[L(\sigma_t^2, h_t)]$, where $L(\cdot, \cdot)$ is some known loss function. The substitution of $\tilde{\sigma}_t^2$ for σ_t^2 leads to an expected loss, $\mathrm{E}[L(\tilde{\sigma}_t^2, h_t)]$, that need not be equal to $\mathrm{E}[L(\sigma_t^2, h_t)]$. So a criterion that is based on the proxy need not be useful for a *quantitative* assessment of volatility models, which is the point made by Andersen and Bollerslev (1998*b*).

Another question is whether the proxy distorts the relative ranking of volatility forecasts. The problem is that the population measure and the proxy induce preordering of an arbitrary set of volatility models, that need not be identical. This consistency problem was analyzed in Hansen and Lunde (2005*b*), who derived conditions that ensure that

$$\mathrm{E}[L(\sigma_t^2, h_{i,t})] < \mathrm{E}[L(\sigma_t^2, h_{j,t})] \text{ if and only if } \mathrm{E}[L(\tilde{\sigma}_t^2, h_{i,t})] < \mathrm{E}[L(\tilde{\sigma}_t^2, h_{j,t})],$$

with i and j referring to two competing volatility models.

The preordering induced by σ_t^2 and the proxy, $\tilde{\sigma}_t^2$, are defined from $\mathrm{E}[L(\sigma_t^2, \cdot)]$ and $\mathrm{E}[L(\tilde{\sigma}_t^2, \cdot)]$. Similarly, the sample average, $n^{-1} \sum_{t=1}^{n} L(\tilde{\sigma}_t^2, \cdot)$, induces a third preordering so we have a *true preordering*, an *approximate preordering*, and an *empirical preordering*, respectively. It is shown that some (but not all) of the commonly used criteria for comparing volatility models satisfy the conditions that ensure that the true preordering and the approximate preordering are equivalent. One important exception is the R^2-criterion, when the R^2 is calculated from the Mincer-Zarnowitz regression using logarithmically transformed variables. So this criterion may identify an inferior volatility model as the 'best', and the inferior model may spuriously be found to be 'significantly' better than all other models, with a probability that converges to one as the sample size increases. In particular, Hansen and Lunde (2005*b*) showed that a sufficient condition for a loss function to yield a consistent ranking is that $\partial^2 L(\sigma_t^2, h_t)/\partial(\sigma_t^2)^2$ does not depend on h_t. Patton (2008) generalized this result by providing necessary and sufficient conditions to ensure that the true preordering and the approximate preordering for a univariate loss function are equivalent. Further discussion and some additional examples are provided in Hansen and Lunde (2005*a*), Hansen, Lunde and Nason (2003), Patton and Sheppard (2009*b*) and Laurent, Rombouts and Violante (2009).

5 Reduced Form Volatility Forecasts

There are several approaches to reduced form forecasting using realized measures of volatility. Most of these methods are quite similar in terms of the forecasts they produce. In fact the various methods can be viewed as flexible extensions of the simple exponential smoothing forecast. This observation is not surprising because volatility is known to be highly persistent (close to unit root) and a realized measure is simply a noisy measurement of the underlying population volatility. Since, the MSE optimal forecast within the local level model (random walk with noise) is given by exponential smoothing, we should not be surprised to see the reduced form forecasts yield point forecasts that closely resembles that of exponential smoothing.

In this section we first consider reduced form models with a simple distributed lag structure. Then we turn to the forecasts based on ARFIMA models and the so-called HAR model. Finally we discuss some results from investigations of nonlinearity in volatility.

5.1 Distributed Lag Models

A simple and elegant approach to block sampling filters is presented in Barndorff-Nielsen et al. (2004). Within the context of stochastic volatility models, their analysis takes advantage of the joint convergence,

$$\sqrt{N}(RV_{s:p} - IV_{s:p}) \overset{law}{\to} N(0, 2\text{diag}(IQ_{s:p})),$$

where $RV_{s:p} \equiv (RV_s, RV_{s+1}, \ldots, RV_p)'$, $IV_{s:p} \equiv (IV_s, IV_{s+1}, \ldots, IV_p)'$ and $IQ_{s:p} \equiv (IQ_s, IQ_{s+1}, \ldots, IQ_p)'$. Based on this they show that (if the realized variances constitute a covariance stationary process) the best linear forecast of IV_{p+1} is given by

$$\widehat{IV}_{p+1} = c\text{E}[IV_{p+1}] + aRV_{s:p},$$

where a is the $1 \times (p - s)$ matrix given by $\text{cov}\left[RV_{p+1}, RV_{s:p}\right](\text{var}\left[RV_{s:p}\right])^{-1}$ and $c = (1 - a\iota')$. This gives a particular simple forecast that has the following generic form

$$\widehat{IV}_{p+1} = \hat{c}\frac{1}{p}\sum_{i=0}^{p} RV_j + \sum_{i=0}^{p-s} \hat{a}_i RV_{p-i}.$$

This forecast should be viewed as a simple intercept constrained predecessor of the more flexible ARFIMA approach that we discuss later in this section.

Andreou and Ghysels (2002) extend the continuous record analysis of Foster and Nelson (1996) to estimation of the integrated variance process, $IV_{t-h,t}$, by treating $IV_{t-h,t}$ as a continuous time process in t for

fixed h. The generic form of their estimator is

$$\widehat{IV}_{t-h,t(k,m_1,m_2)} = \sum_{i=0}^{km_1} w_i \, RV^{(m_2)}_{t-i/m_1-h,t-i/m_1} \tag{3}$$

where w_i is a weighting scheme and $RV^{(m_2)}_{t-i/m_1-h,t-i/m_1}$ is the realized variance estimated from squared returns sampled m_2 times each day. So $\widehat{IV}_{t-h,t(k,m_1,m_2)}$ is a weighted average of realized variances rolling along the intervals $[t-i/m_1-h:t-i/m_1]$ for $i=0,\ldots,km_1$. In a simulation study and an empirical study they compare flat or exponentially decaying weights with $m_1 = m_2$ corresponding to 5 minute sampling. They set the window length to cover 1, 2 and 3 days/months. They find that the optimal window length extends well beyond h, and that exponentially decaying weights are better than flat.

Fleming, Kirby and Ostdiek (2003) were the first to apply such methods to study the economic value of high-frequency data in the context of investment decisions. In a trivariate setting they estimate the daily conditional covariance matrix of the returns on the S&P 500 futures, gold futures and Treasury bond futures. They use a variant of (3) with non-overlapping intervals for the RVs that can be expressed as $\widehat{IV}^1_{t,t+1(t,1,5\,\text{min})}$ where the w_i decays exponentially. They find that the utilization of high-frequency returns as opposed to daily returns deliver gains that amount to $50-200$ basis point per year. Similar performance evaluations have been conducted in many of the papers that we discuss below. While the analysis in Fleming et al. (2003) is solely based on realized variances estimated from 5 minute returns, more recent studies are based on more sophisticated estimators, that better utilize the entire data base of high frequency data.

Bandi, Russell and Zhu (2008) use the methodology of Fleming et al. (2003) to assess the profitability of methods that optimally sample (in an MSE sense) intraday returns in the presence of market microstructure noise. They compare the value of their optimally sampled (see Bandi and Russell (2008)) daily estimates to the value of conventional estimates based on 5- and 15-minute intervals. In contrast to Fleming et al. (2003) they construct out-of-sample forecasts using univariate ARFIMA models. For the sample period from January 1993 to December 2003 using quote data for three representative stocks from the S&P 100 they find mildly significant gains for the optimal sampling strategy. The gains range from 3 to 80 basis points.

de Pooter, Martens and van Dijk (2008) use the forecasting and evaluation approach used in Fleming et al. (2003), but expand the empirical analysis to include all assets in the S&P 100 index as well as the S&P 500 index futures. Their sample spans the period from April 16, 1997 to May 27, 2004. They set of realized measures include the standard realized covariance matrix, a two time-scales estimator inspired by Zhang et al. (2005) and a one-lead one-lag corrected estimator. They compare the performance of these

estimators when based on various sampling frequencies for the high-frequency data. These range from 1 minute to 130 minutes. In contrast to Bandi et al. (2008) they find quite surprisingly that the optimal (in terms of profit) sampling frequency is in the neighborhood of one hour rather than 5 to 30 minutes. However, the way that the forecasts are set up is quite different. While Bandi et al. (2008) apply ARFIMA methods to the individual realized variance/covariances, de Pooter et al. (2008) apply outer product based estimators to the 100-dimensional vector of intraday return. The latter method is vulnerable to the so-called eps-effect. This effect has to do with the fact that all assets do not trade at the same time. The trades have asynchronous time stamps, so when constructing high dimensional return vectors there will be several stale prices that result in zero returns. The result is that the realized covariances are biased towards zero. The problem and remedies are presented in Barndorff-Nielsen et al. (2010) that indicate that the bias corrected estimators used by de Pooter et al. (2008) do not fully resolve the problems arising from eps-effects. So it is quite likely that an analysis based on the multivariate realized kernel would result in a higher optimal sampling frequency for the high frequency data.

The approach of Andreou and Ghysels (2002) is generalized in Ghysels et al. (2006) that consider volatility (MIDAS) regressions such as

$$RV_{t,h} = \mu + \phi \sum_{k=0}^{K} b(k,\theta) \tilde{X}_{t-k} + \varepsilon_t, \tag{4}$$

where $RV_{t,h} = RV_t + \cdots + RV_{t+h}$. Their idea is that while the target $RV_{t,h}$ measures the integrated variance, $IV_{t,h}$, over a h-interval, the regressors, $\tilde{X}_{t-k}$, can be sampled at any frequency that seems relevant for the purpose of predicting $IV_{t,h}$. This is obviously an extension of the distributed lag models for which hypothesis testing and estimation of alternative form for $b(k,\theta)$ have been discussed extensively in the literature, see e.g. Griliches (1967) or Nerlove (1972). In Ghysels et al. (2006) the weights, $b(k,\theta)$, are specified according to the Beta function, that allows flat, gradually declining and hump-shaped patterns. Several different choices for $\tilde{X}_{t-k}$ are investigated, including daily realized variance, $RV_{t,1}$, 5 minute or daily squared/absolute returns, the daily range and the daily sum of intraday absolute returns. A logarithmic version of (4) is also included in the analysis. The general conclusion is that predictors based on absolute returns are more successful than those based on squared returns at explaining the variation of future volatility. This feature arises quite naturally in the eigenfunction approach to volatility model (see Meddahi (2001)) because the absolute return is more closely related to the first eigenfunction than is squared returns in certain models. Moreover, this is also a natural consequence of some empirical results in Hansen and Lunde (2010), that we reviewed in the previous section. Specifically the fact that the absolute return is a less noisy proxy of latent volatility than is the squared returns, can explain that the absolute return is a better predictor of future volatility than is squared return. In the next section we

discuss forecasting with ARFIMA type models.

5.2 ARFIMA Models

The idea of fitting ARFIMA models to realized measures was put forth in 1999 and later published as Andersen et al. (2003). These authors estimate a trivariate long-memory Gaussian VAR for the logarithmic realized volatilities,

$$\Phi(L)(1 - L)^d(y_t - \mu) = \varepsilon_t, \tag{5}$$

where ε_t is a vector white noise process. The dependent variable, y_t, is a vector with $\log(RV_t)/2$ components for DM/\$, ¥/\$ and ¥/DM exchange rates. They found $\hat{d} = 0.401$ using the methods developed in Robinson (1995). A general finding of empirical studies with ARFIMA models in this context, is that this framework produces volatility forecasts that dominate those of conventional GARCH models that are based on daily returns. In Andersen, Bollerslev, Diebold and Ebens (2001) similar models are estimated with realized measures computed with stock returns, but this paper does not report out-of-sample forecasting results. As mentioned above Ghysels et al. (2006) benchmark their mixed frequency distributed lag approach to this ARFIMA approach. So while (5) is only allowed to use lags of $\log(RV_t)/2$ the models based on (4) can use more or less information. When (4) is based on interday information such as r_t^2 or $|r_t|$ then forecasts delivered by (5) give a lower out-of-sample MSE. However, when (4) is based on intraday measures it was found to outperform the ARFIMA model. This also holds when $\tilde{X}_{t-k}$ consists of 5 minute intraday squared/absolute returns which deliver the overall lowest MSE ratios for the competing models. One may note that these conclusions are only indicative as no measure of precision is attached to the MSE ratios presented in Ghysels et al. (2006).

Following Andersen et al. (2003) there have been many applications and extensions of the ARFIMA models to realized variance and other realized measures of the integrated variance. For example Deo, Hurvich and Lu (2005) that estimate a long-memory stochastic volatility (LMSV) model from 30 minute intraday returns, and find its performance to be quite similar to that of Andersen et al. (2003). Oomen (2001, 2004) provide a detailed study of the ARFIMA approach with realized variances computed from minute by minute data of the FTSE-100 index. In Oomen (2001) a simulation study demonstrates that the forecasting performance of the ARFIMA model outperforms that of conventional GARCH models. Interestingly, Oomen (2004) augments the ARFIMA model with an EGARCH style function of lagged returns, lagged daily trading volume and interest rates (1 month UK Interbank rate). All regressors are found to have predictive power assessed by the in-sample fit, but the paper does not report out-of-sample results.

Li (2002) compares the ARFIMA forecasting power to that of implied volatility from options on three currencies. Martens and Zein (2002) question the statistical validity of Li (2002) because overlapping data are used in the forecast evaluation. For the S&P 500, YEN/USD, and Light, Sweet Crude Oil Martens and Zein (2002) find that long memory forecasts dramatically improve upon daily GARCH forecasts, confirming the results of Andersen et al. (2003). Moreover, as Li (2002) they also find that high-frequency time-series forecasts have incremental information over that contained in implied volatilities. In contrast, Pong, Shackleton, Taylor and Xu (2004) find that ARFIMA forecasts have incremental information over implied volatilities for short forecast horizons such as one day and one week. But implied volatilities are found to incorporate most of the relevant information when the horizon is either one month or three months. However, the out-of-sample implied volatility forecasts that Pong et al. (2004) employ are "enhanced" by regressing ex-post realized volatilities on implied volatilities, and using the regression coefficients to better match the target. In contrast, Martens and Zein (2002) use unrefined implied volatilities directly as extracted from option prices.

Koopman, Jungbacker and Hol (2005) also compare the forecasting ability of realized volatility to that of option-implied volatility and plain squared interday returns. They estimate long memory models in the style of (5) and an unobserved components (UC-RV) model taken from Barndorff-Nielsen et al. (2004). The UC-RV model consists of independent ARMA components that are estimated by Kalman filtering methods. The predictive ability of the alternative models is compared using formal testing via the test for superior predictive ability (SPA) by Hansen (2005). The most accurate forecasts are found to be those delivered by the ARFIMA specification, while the UC-RV model comes in as a close second.

Finally, Chiriac and Voev (2009) address the problem of modeling the dynamic properties of realized covariance matrices. They decompose the realized covariance matrices into Cholesky factors and model the elements of the decomposition with a multivariate vector fractionally integrated ARMA (VARFIMA) model. In their forecast evaluation they report that their model significantly outperforms the relevant benchmarks at various prediction horizons. The results hold both in terms of minimizing the RMSE of the forecast, and with respect to improving the performance of a mean-variance efficient portfolios.

5.3 HAR Models

Inspired by the HARCH model of Müller, Dacorogna, Davé, Olsen, Pictet and von Weizsacker (1997) and Dacorogna, Müller, Olsen and Pictet (1998) and by the asymmetric propagation of volatility between long and short time horizons, Corsi (2009) proposes an additive cascade model of different volatility components designed to mimic the actions of different types of market participants. The model is called

the heterogeneous autoregressive model of realized variance (HAR-RV), and it is a predictive model for the daily integrated volatility, $\sqrt{IV_{t+1}}$. The model has the following simple time series representation

$$\sqrt{RV_{t+1}} = c + \beta^{(d)}\sqrt{RV_t} + \beta^{(w)}\sqrt{RV_{t-5,5}} + \beta^{(m)}\sqrt{RV_{t-22,22}} + \varepsilon_{t+1}. \tag{6}$$

So the model predicts future volatility using a daily, a weekly and a monthly component. Corsi (2009) also discusses adding other components and alternative targets such as IV_{t+1} or its logarithmic transformation. The HAR model is found to be very successful in practice, which is impressive given its relatively simple structure. Corsi (2009) presents out-of-sample forecasting results for three realized volatility series of USD/CHF, S&P500, and T-Bond, where the model in (6) consistently outperforms short-memory models at all the time horizons considered (one day, one week, and two weeks) and is comparable to the ARFIMA models.

The HAR model is sometimes associates with long-memory models. An alternative interpretation is that the prediction produced by HAR model is a kind of exponential smoothing of lagged values of $\sqrt{RV_t}$. Although the weights are forced to be piecewise constant, the framework offers a flexible framework with 3 parameters, unlike the traditional exponential smoothing where weights are all defined from a single parameter.

Following the initial version of Corsi (2009) many extensions have been explored. Andersen, Bollerslev and Diebold (2007) separate the three volatility measures in jump and continuous volatility components. In Corsi, Mittnik, Pigorsch and Pigorsch (2008) the authors give a careful study of the model fit resulting from applying (6) to S&P 500 index futures. They show that allowing for time-varying volatility of realized volatility and logarithmic realized variance substantially improves the fit as well as predictive performance. A similar approach is taken in Bollerslev, Kretschmer, Pigorsch and Tauchen (2009) who build a multivariate discrete-time volatility model for the returns and the realized continuous sample path and jump variation measures. Among their reduced form equations they have (6) applied to the logarithmic bipower variation, and as Corsi et al. (2008) they allow for time-varying volatility in a GARCH style fashion. They present evidence of a very accurate in-sample fit to their sample of S&P500 Index futures contracts traded on the Chicago Mercantile Exchange, in the period covering January 1, 1985 to December 31, 2004. However, the authors do not report out-of-sample forecasting results.

5.4 Separating the Continuous and Jump Components of Volatility

The price process can be expresses as the sum of a continuous process plus a pure jump process, and the quadratic variation can similarly be decomposed into the variability of the continuous sample path plus the (squared) jump component. In terms of predictability, the jump component appears to be much

less predicable than the continuous component, and the existing empirical evidence suggests that the continuous component is a better predictor than a realized measure that contains both components.

Andersen et al. (2007) where the first to make this observation. These authors estimated the jump component as the difference between the realized variance and the the bipower variation, see Barndorff-Nielsen and Shephard (2004, 2006). In their empirical analysis of exchange rates, equity index returns, and bond yields Andersen et al. (2007) find that the volatility jump component is distinctly less persistent than the continuous component, yet an important predictor of future volatility. Moreover, the separation of the jumps from the smooth continuous part results in significant out-of-sample volatility forecast improvements. Similar results have been reported in Anderson and Vahid (2007), Lanne (2007), and Andersen, Bollerslev and Huang (2010)

5.5   Nonlinearity

The ARFIMA modeling framework for realized variance has also be modified in various non-linear directions. Maheu and McCurdy (2002) develop a Markov switching model for the integrated variance which they apply to ten years of realized variances for DM/\$ exchange rates. These realized variances are based on the sum of squared 5-minute MA(4) filtered return. They find that their non-linear model provides a marginal (R^2 increase from 0.446 to 0.459) improvement of the out-of-sample forecasting performance relative to a plain ARMA model augmented with daily squared returns. However, this empirical result is based on a single time series so it is unclear whether this result will hold more broadly.

Asai, McAleer and Medeiros (2009*a,b*) consider extensions to the ARFIMA framework that allow asymmetric effects of lagged positive or negative shocks, and Hillebrand and Medeiros (2009) investigate nonlinear variations that nest the Andersen et al. (2003) framework.

Hillebrand and Medeiros (2008) use a bagging strategy to expand the set of regressors in a log-linear version of (6). They also explore a nonlinear version of the HAR model based on neural networks. They find that bagging reduces the prediction MSE for both model types, and that bagging eliminates the advantage of the nonlinear specification. McAleer and Medeiros (2008) proposed an extension of the linear HAR model by incorporating smooth transitions. In a comparison to (6) few rejections of equal predictive ability are found for the sixteen DJ stocks they consider. A review of this research is given in McAleer and Medeiros (2009). Our overall conclusion is that significant gains from highly non-linear models have still not surfaced in out-of-sample volatility forecasting.

6 Model-Based Volatility Forecasts

Discrete time GARCH models are widely used for predicting volatility in practice, and realized measures have recently made their inroad into these models. In this section we discuss the GARCH-type models that utilize realized measures of volatility. This area of research is in its infancy, but it has convincingly shown that realized measures can greatly improve the empirical fit of these models.

For volatility prediction a key variable of interest is the conditional variance

$$\sigma_t^2 = \text{var}(r_t | \mathcal{F}_{t-1}),$$

where the filtration, $\mathcal{F}_t$, is the σ-field generated by all variables that are available for prediction at times t. Conventional GARCH models provide a specification for a conditional variance that is defined with a simpler filtration, which is generated exclusively by past returns, $\mathcal{F}_t^r = \sigma(r_t, r_{t-1}, \ldots)$. Realized measures are useful additions to GARCH models because

$$\text{var}(r_t | \mathcal{F}_{t-1}^r) \neq \text{var}(r_t | \mathcal{F}_{t-1}^{r,RM}),$$

where $\mathcal{F}_t^{r,RM} = \sigma(RM_t, r_t, RM_{t-1}, r_{t-1}, \ldots)$. In fact, the difference between the two conditional variances tends to be more pronounced after a sudden change in the conditional variance, as we illustrate with a simple example below.

To distinguish between the true conditional variance, $\sigma_t^2 = \text{var}(r_t | \mathcal{F}_{t-1})$, and the model-based equivalent, we denote the latter by its standard notation, h_t. Moreover, to simplify the exposition we consider the case where $\text{E}(r_t | \mathcal{F}_{t-1}) \stackrel{a.s.}{=} 0$.

The simplest GARCH(1,1) model is given by

$$r_t = \sqrt{h_t^G} z_t, \tag{7}$$
$$h_t^G = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}^G, \tag{8}$$

where $z_t \sim \text{iid}(0, 1)$. In this model the conditional variance is driven up and down by the squared inter-day returns.

A simple way to enhance the conventional GARCH model is to include a realized measure in the equation for conditional variance,

$$h_t^{GX} = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}^{GX} + \gamma RM_{t-1}. \tag{9}$$

This is an example of a GARCHX model, because the realized measure, RM_{t-1}, is treated as an exogenous variable. GARCHX models of this type have been estimated by Blair et al. (2004), Martens (2001, 2002),

Engle (2002), Barndorff-Nielsen and Shephard (2007), Barndorff-Nielsen et al. (2010), amongst others. For instance, Engle (2002) found that α (the squared return) is insignificant once the realized variance was included, and Barndorff-Nielsen and Shephard (2007) arrived at the same conclusion if the bipower variation is included. Barndorff-Nielsen et al. (2010) also find this in a multivariate framework when estimating a scalar BEKK version of (7) and (9). For all these models, including a realized measure often result in a substantial boost of the log-likelihood function.

Note that both models, GARCH and GARCHX, could, in principle, be correctly specified in the sense that

$$h_t^G = \text{var}(r_t|\mathcal{F}_{t-1}^r), \qquad \text{and} \qquad h_t^{GX} = \text{var}(r_t|\mathcal{F}_{t-1}^{r,RM}),$$

even if $\gamma \neq 0$. Naturally, for the purpose of forecasting it is fair game to include any variable that adds predictive power. The empirical evidence that $\gamma \neq 0$ is overwhelming, and the implication is that the realized measure is a useful predictor of volatility. For instance, in the present context it follows that

$$E[\{r_t^2 - \text{var}(r_t|\mathcal{F}_{t-1}^r)\}^2] > E[\{r_t^2 - \text{var}(r_t|\mathcal{F}_{t-1}^{r,RM})\}^2],$$

when $\gamma \neq 0$.

Before we review the growing literature on this topic, we shall discuss the intuition behind the value of introducing realized measures of volatility into GARCH models. This intuition is nicely laid out in Shephard and Sheppard (2009). First we note that if the GARCH(1,1) model was the true data generating process with $h_t^G = \sigma_t^2$, then there would be absolutely no population gains from including a realized measure in the GARCH equation (9). Since the realized measures are found to be highly beneficial in practice this goes to show that the model-implied conditional variance, h_t^G, from a conventional GARCH model should merely be viewed as an approximation to the conditional variance defined by σ_t^2. It is therefore relevant to discuss how well h_t (h_t^G or h_t^{GX}) approximates the conditional variance, σ_t^2, in various circumstances. So consider the following simple example that we have borrowed from a presentation of the paper by Hansen, Huang and Shek (2009). Suppose that, σ_t^2, is such that the volatility is $\sigma_t = 20\%$ for $t < T$ and then jumps to $\sigma_t = 40\%$ for $t \geq T$, and suppose that the squared return and the realized measure are both unbiased estimates of σ_t^2. The parameter values $\omega = 0$, $\beta = 0.95$, and $\alpha = 0.05$ are in line with typical estimates of a GARCH(1,1) model when estimated with daily observations. In this thought experiment the implication is that $E(h_{T+k}^G) = \alpha \sum_{j=0}^{\infty} \beta^j E(r_{T+k-j}^2)$ so for $k \geq 0$

$$E(h_{T+k}^G) = \alpha \sum_{j=0}^{k-1} \beta^j (40\%)^2 + \alpha \sum_{j=k}^{\infty} \beta^j (20\%)^2 = (1 - \beta^k)(40\%)^2 + \beta^k (20\%)^2,$$

where we have used that $\alpha = 1 - \beta$. Similarly, the parameter values, $\omega = 0$, $\alpha = 0$, $\beta = 0.50$, and $\gamma = 0.50$, are in line with typical estimates of the GARCHX model, (9), and we can compute $E(h_{T+h}^{GX})$ using the same expression with γ in place of α. Figure 2 displays the square-root of the expected "conditional variance" that is implied by each of the two models, along with the true volatility. The figure illustrates how long it takes a conventional GARCH model to "catch up" with the new level of volatility. The 60 periods it takes for h_t^G to get slightly above 39% (on average) can be interpreted as 3 months, because the parameter values are chosen to match typical estimates with daily observations. In comparison, it takes just four days for the GARCHX model to bring h_t^{GX} above 39% (on average).
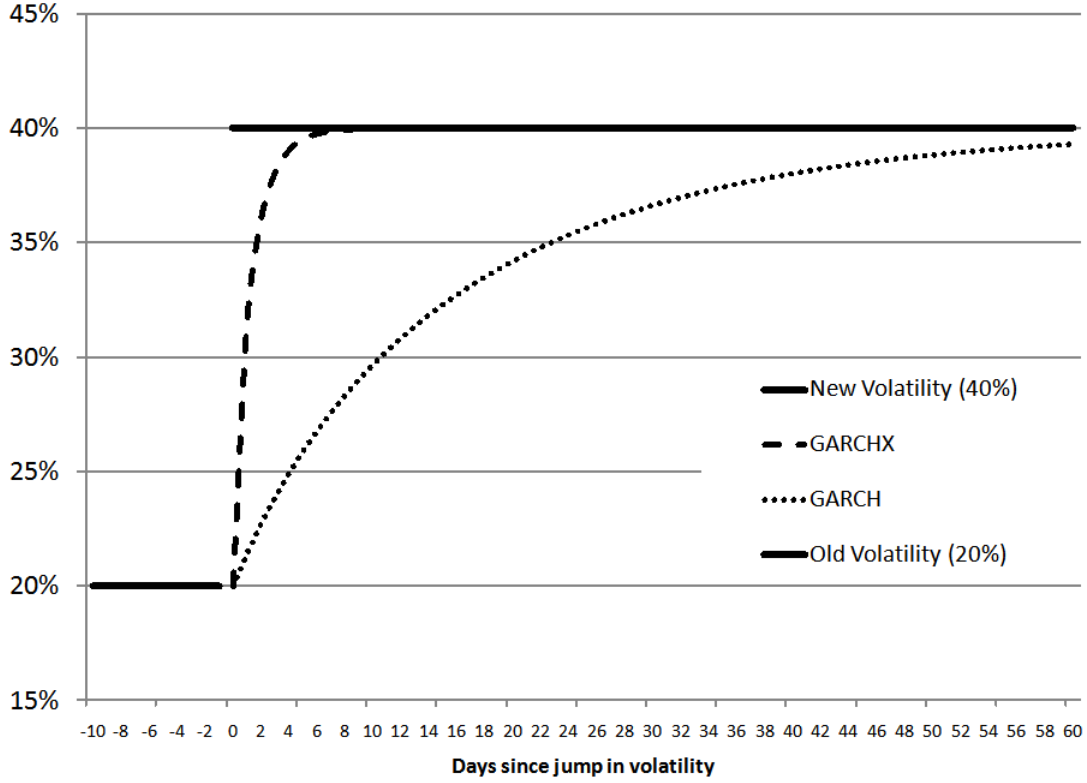

Figure 2: The figure reveals the speed by which a conventional GARCH model can adjust its model-implied volatility after a sudden jump in volatility. In this example it takes h_t^G about 3 months (60 days) to approach the new level, whereas the GARCHX model achieves most of the adjustment within a few days.

6.1 Joint Models for Returns and Realized Measures

The GARCHX model defined (7) and (9) is an incomplete model in the sense that it makes no attempt to model the realized measure, RM_t. For the problem of predicting volatility one-period ahead there is

not necessary to specify a model for RM_t. However, for predicting the volatility a longer horizons a specification for RM_t is needed to "complete" the model.

Two different ways to complete the model have been proposed in the literature. First, Engle and Gallo (2006) proposed to model each of the realized measures with additional GARCH-type models. We refer to this as a parallel GARCH structure. The second approach is due to Hansen et al. (2009) who introduced a model structure in which the realized measure is tied to the conditional variance of returns. A key difference between these two approaches is the number of latent volatility variables. The Realized GARCH model by Hansen et al. (2009) has a single latent volatility factor – the conditional variance of returns, h_t, – as is the case in conventional GARCH models. Contrary, the parallel GARCH structure that is used in Engle and Gallo (2006) and Shephard and Sheppard (2009), adds an additional latent volatility variable for every realized measures that is used in the model

6.2   Parallel GARCH Structure

The parallel GARCH structure can be classified as a multiplicative error model (MEM), which was first introduced by Engle (2002). In the present context with realized measures the MEM structure was first used Engle and Gallo (2006), who followed Engle (2002) and labeled their model a MEM. This parallel GARCH structure is also used in Shephard and Sheppard (2009), who labeled their model the HEAVY model.

The conventional GARCH model, (7), implies that

$$r_t^2 = h_t z_t^2.$$

The approach by Engle and Gallo (2006) is to impose a similar structure on the realized measure,

$$RM_t = \tilde{h}_t \tilde{z}_t^2,$$

where $\tilde{z}_t \sim \text{iid}(0, 1)$. The formulation in Engle and Gallo (2006) is that $\eta_t = \tilde{z}_t^2$ is iid with $\text{E}(\eta_t) = 1$, which is (observationally) equivalent to the formulation we use here with $\tilde{z}_t$. An implication is that $\tilde{h}_t$ is the conditional expectation of RM_t, i.e. $\tilde{h}_t = \text{E}(RM_t | \mathcal{F}_{t-1})$. This introduces an additional latent variable that is being modelled in a manner that is analogous to h_t, such as

$$\tilde{h}_t = \tilde{\omega} + \tilde{\alpha} RM_{t-1} + \tilde{\beta} \tilde{h}_{t-1},$$

So RM_t is effectively being modeled analogously to the way r_t^2 is modeled in the GARCHX model. This makes it easy to formulate the joint likelihood function for returns and realized measures, and the

framework is easy to extend to multiple realized measures, by introducing additional GARCH models for each of the realized measures. Engle and Gallo (2006) focusses on the case with two realized measures, whereas Shephard and Sheppard (2009) used a simplified parallel GARCH structure with one realized measure.

6.3 Realized GARCH

The Realized GARCH model by Hansen et al. (2009) takes a different approach to modelling the realized measure. A key component of the Realized GARCH model is a measurement equation that ties the realized measure to the conditional variance of returns. For instance, the measurement equation could take the form

$$RM_t = \xi + \varphi h_t + \tau(z_t) + u_t,$$

where $u_t \sim iid(0, \sigma_u^2)$ and independent of z_t. The leverage function, $\tau(z_t)$, is used to capture the dependence between shocks to returns and shocks to volatility, which is known to be empirically important. Hansen et al. (2009) proposed to construct τ from Hermite polynomial. The first two Hermite polynomials yield the quadratic specification, $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$, which can accommodate some degree of asymmetry. This choice is convenient because it ensures that $E[\tau(z_t)] = 0$ whenever z_t has mean zero and unit variance.

An implication of the measurement equation is that

$$E(RM_t|\mathcal{F}_{t-1}) = \xi + \varphi h_t,$$

so the second volatility factor, $\tilde{h}_t$, that is introduced with a parallel GARCH structure is a simple transformation of the conditional h_t.

The measurement equation facilitates estimation of the so-called *new impact curve*, and since φ may be less than one, this framework can accommodate the situation where the realized measure is computed over a period that is shorter than the period that the return, r_t, spans. This is important in practice because for the modelling of daily, close-to-close returns, the availability of high-frequency data is often limited to a much short (open-to-close) period, e.g. 6.5 hours. Hansen et al. (2009) actually recommend a logarithmic specification because it provides a better specification for the equity data they analyze.

6.3.1 Realized GARCH Forecasting

The Realized GARCH model implies an VARMA structure for (h_t, RM_t), that is very useful for forecasting. In this section we focus on forecasting with the simplest Realized GARCH model. Forecasting with more complex Realized GARCH models, including the log-linear model that implies a VARMA structure for , is straight forward and we refer to Hansen et al. (2009) for details. The linear RealGARCH(1,1) model is defined by the three equations,

$$r_t = h_t^{1/2} z_t,$$

$$h_t = \omega + \beta h_{t-1} + \gamma RM_{t-1},$$

$$RM_t = \xi + \varphi h_t + w_t,$$

with $w_t = \tau(z_t) + u_t$ and $\tau(z) = \tau_1 z + \tau_2(z^2 - 1)$. It is simple to see that this structure implies

$$\begin{bmatrix} h_t \\ RM_t \end{bmatrix} = \begin{bmatrix} \omega \\ \xi + \varphi \omega \end{bmatrix} + \begin{bmatrix} \beta & \gamma \\ \varphi \beta & \varphi \gamma \end{bmatrix} \begin{bmatrix} h_{t-1} \\ RM_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ w_t \end{bmatrix}.$$

So H-periods ahead prediction can be done with

$$\begin{bmatrix} h_{T+H} \\ RM_{T+H} \end{bmatrix} = A^H \begin{bmatrix} h_T \\ RM_T \end{bmatrix} + \sum_{h=0}^{H-1} A^h \left(\begin{bmatrix} \omega \\ \xi + \varphi \omega \end{bmatrix} + \begin{bmatrix} 0 \\ w_{T+H-h} \end{bmatrix} \right),$$

where the matrix

$$A = \begin{bmatrix} \beta & \gamma \\ \varphi \beta & \varphi \gamma \end{bmatrix},$$

has a particular simple structure so that $A^h = (\beta + \varphi \gamma)^{h-1} A$, for $h \geq 1$.

Moreover, if the prime objective is to predict the future path of the conditional variance, h_t, the auxiliary predictions of x_t are not required in this context, because the Realized GARCH model implies a reduced form model for h_t. Specifically we have

$$h_t = (\omega + \gamma \xi) + (\beta + \varphi \gamma) h_{t-1} + \gamma w_{t-1},$$

so that

$$h_{T+H} = (\beta + \varphi \gamma)^H h_T + \sum_{h=0}^{H-1} (\beta + \varphi \gamma)^h [(\omega + \gamma \xi) + \gamma w_{T+H-1-h}].$$

This expression is derived within the simplest specification with one lag of h_t and one lag of RM_t in the GARCH equation ($p = q = 1$). The expressions that arise from the more general specification with additional lags, $p, q > 1$ are similar.

27

6.4   Other Related Approaches

Chen, Ghysels and Wang (2010) propose a framework called HYBRID GARCH which is a bit difficult to classify in this context, because it consists of a range of different estimation approaches, and operates with a volatility process that need not be defined to be the conditional variance of returns. A central theme in Chen et al. (2010) is the use of sophisticated realized measures that, in part, are inspired by the MIDAS structure by Ghysels et al. (2006). Given the empirical success of MIDAS models the HYBRID GARCH structure may offer a promising alternative to volatility forecasting.

7   Concluding Remarks and Some Directions for Future research

In this chapter we have focused on some aspects of high-frequency data and their use in volatility forecasting. High-frequency data can be used to construct volatility forecasts. We have reviewed two leading approaches to this. One approach is the reduced form forecast, where the forecast is constructed from a time series model for realized measures, or a simple regression based approach such as the HAR model. The other approach is based on more traditional discrete time volatility models that include a modeling of returns. Such models can be generalized to utilize information provided by realized measures. We have also discusses how volatility forecasts, produced by complex volatility models, can benefit from high frequency data in an indirect manner, through the used of realized measures to facilitate and improve the estimation of complex models.

The time series properties of volatility are fundamentally linked to the problem of forecasting volatility. A key insight is that high-frequency based realized volatility measures facilitate a better understanding of latent volatility and its dynamic properties. Despite the progress that has been made in estimating financial volatility from high-frequency data, it is still important to discriminate between the realized measure of volatility and the underlying population quantity. The modeling implication of this is that measurement errors must be accounted for in the analysis. This observation is strongly supported by many empirical studies that have found the moving average term(s) to be highly significant in the reduced form AR(FI)MA models, when these models are estimated with realized measures. Finally, high-frequency data have played a central role in the evaluation and comparison of volatility forecasts. For this problem the introduction of the realized measures have proven extremely valuable. In fact, without the realized measures of volatility it would be difficult to distinguish good volatility models from bad models.

Let us conclude by giving some directions for further research. There are still many open issues in this field. One interesting topic is the extend to which microstructure effects are relevant for the forecasting problem. For instance, is it important to use realized measures that are robust to microstructure effects, or

are less robust measures just as good for the purpose of prediction. Some work towards an understanding of this is Brownless and Gallo (2010), but there is still much more to investigate. Also, the presence of jumps in the price process may have information for forecasting volatility. Andersen et al. (2007) give a detailed study that confirms this in their HAR modelling frame. However, more research is needed before any stylized facts can be established.

Another, promising line of research is the use of realized semivariance in volatility forecasting. In Barndorff-Nielsen, Kinnebrock and Shephard (2009) and Patton and Sheppard (2009*a*) the authors demonstrate that is the high frequency variation that corresponds to negative returns is very informative for future variance.

Finally, there are other related areas where high frequency data can be helpful. One important problem is the estimation and forecasting of systematic risk that is a key object in financial economics. In Bollerslev and Zhang (2003) and Andersen, Bollerslev, Diebold and Wu (2005, 2006) systematic risk is measured by the so-called realized beta that is derived from the realized covariance matrix. There are many open questions in this area that need attention. In particular, the issue of how to deal with the measurement error left in realized beta is an important problem.

References

Aït-Sahalia, Y. and Mancini, L. (2008), 'Out of sample forecasts of quadratic variation', *Journal of Econometrics* **147**, 17–33.

Andersen, T. G. (2000), 'Some Reflections on Analysis of High-Frequency Data', *Journal of Business & Economic Statistics* **18**(2), 146–153.

Andersen, T. G. and Bollerslev, T. (1997*a*), 'Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns', *Journal of Finance* **52**(3), 975–1005.

Andersen, T. G. and Bollerslev, T. (1997*b*), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**, 115–158.

Andersen, T. G. and Bollerslev, T. (1998*a*), 'Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies', *Journal of Finance* **53**(1), 219–265.

Andersen, T. G. and Bollerslev, T. (1998*b*), '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 & Statistics* **89**, 707–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000), 'Great realizations', *Risk* **13**(3), 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**(2), 579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Wu, J. (2005), 'Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?', *American Economic Review* **95**, 398–404.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Wu, J. (2006), Realized Beta: Persistence and Predictability, *in* T. B. Fomby and D. Terrell, eds, 'Advances in Econometrics: Econometric Analysis of Financial and Economic Time Series/Part B', Vol. 20, New York: JAI Press, pp. 1–40.

Andersen, T. G., Bollerslev, T. and Huang, X. (2010), 'A reduced form framework for modeling volatility of speculative prices based on realized variation measures', *Journal of Econometrics, Forthcoming* .

Andersen, T. G., Bollerslev, T. and Meddahi, N. (2004), 'Analytic Evaluation of Volatility Forecasts', *International Economic Review* **45**, 1079–1110.

Andersen, T. G., Bollerslev, T. and Meddahi, N. (2009), 'Realized volatility forecasting and market microstructure noise', *Journal of Econometrics, forthcoming* .

Anderson, H. M. and Vahid, F. (2007), 'Forecasting the Volatility of Australian Stock Returns: Do Common Factors Help?', *Journal of Business & Economic Statistics* **25**, 76–90.

Andreou, E. and Ghysels, E. (2002), 'Rolling-sample volatility estimators: Some new theoretical, simulation, and empirical results', *Journal of Business & Economic Statistics* **20**(3), 363–376.

Asai, M., McAleer, M. and Medeiros, M. C. (2009*a*), Modelling and Forecasting Noisy Realized Volatility. http://ssrn.com/abstract=1476044.

Asai, M., McAleer, M. and Medeiros, M. C. (2009*b*), Asymmetry and Leverage in Realized Volatility. http://ssrn.com/abstract=1464350.

Bandi, F. M. and Russell, J. R. (2008), 'Microstructure Noise, Realized Variance, and Optimal Sampling', *Review of Economic Studies* **75**, 339–369.

Bandi, F. M., Russell, J. R. and Zhu, Y. (2008), 'Using High-Frequency Data in Dynamic Portfolio Choice', *Econometric Reviews* **27**, 163–198.

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

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2010), Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Unpublished paper: Oxford-Man Institute, University of Oxford.

Barndorff-Nielsen, O. E., Kinnebrock, S. and Shephard, N. (2009), Measuring downside risk - realised semivariance, *in* M. W. Watson, T. Bollerslev and J. Russell, eds, 'Volatility and Time Series Econometrics: Essays in Honor of Robert Engle', Oxford University Press.

Barndorff-Nielsen, O. E., Nielsen, B., Shephard, N. and Ysusi, C. (2004), Measuring and forecasting financial variability using realised variance with and without a model, *in* A. C. Harvey, S. J. Koopman and N. Shephard, eds, 'State Space and Unobserved Components Models: Theory and Applications', Cambridge University Press, Cambridge, pp. 205–235.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2002*b*), 'Estimating quadratic variation using realised volatility', *Journal of Applied Econometrics* **17**, 457–477.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2006), 'Econometrics of testing for jumps in financial economics using bipower variation', *Journal of Financial Econometrics* **4**, 1–30.

Barndorff-Nielsen, O. E. and Shephard, N. (2007), Variation, jumps and high frequency data in financial econometrics, *in* R. Blundell, T. Persson and W. K. Newey, eds, 'Advances in Economics and

Econometrics. Theory and Applications, Ninth World Congress', Econometric Society Monographs, Cambridge University Press, pp. 328–372.

Blair, B. J., Poon, S.-H. and Taylor, S. J. (2004), 'Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns', *Journal of Econometrics* **105**(1), 5–26.

Bollerslev, T., Kretschmer, U., Pigorsch, C. and Tauchen, G. (2009), 'A Discrete-Time Model for Daily S&P 500 Returns and Realized Variations: Jumps and Leverage Effects', *Journal of Econometrics* **150**, 151–166.

Bollerslev, T. and Zhang, B. Y. B. (2003), 'Measuring and modeling systematic risk in factor pricing models using high-frequency data', *Journal of Empirical Finance* **10**, 533–558.

Bollerslev, T. and Zhou, H. (2002), 'Estimating stochastic volatility diffusion using conditional moments of integrated volatility', *Journal of Econometrics* **109**, 33–65.

Brownless, C. T. and Gallo, G. M. (2006), 'Financial econometric analysis at ultra-high frequency: Data handling concerns', *Computational Statistics & Data Analysis* **51**, 2232–2245.

Brownless, C. T. and Gallo, G. M. (2010), 'Comparison of Volatility Measures: a Risk Management Perspective', *Journal of Financial Econometrics* **8**, 29–56.

Chen, X., Ghysels, E. and Wang, F. (2010), 'The HYBRID GARCH Class of Models', *Working Paper, Department of Economics, University of North Carolina at Chapel Hill* .

Chiriac, R. and Voev, V. (2009), 'Modelling and Forecasting Multivariate Realized Volatility', *Journal of Econometrics* . forthcoming.

Comerton-Forde, C. and Rydge, J. (2004), A Review of Stock Market Microstructure: A study of market microstructure in eighteen of the worlds largest and most influential stock markets. Securities Industry Research Centre of Asia-Pacific, Sydney.

Corradi, V. and Distaso, W. (2006), 'Semi-Parametric Comparison of Stochastic Volatility Models using Realized Measures', *Review of Economic Studies* **73**, 635–667.

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

Corsi, F., Mittnik, S., Pigorsch, C. and Pigorsch, U. (2008), 'The Volatility of Realized Volatility', *Econometric Reviews* **27**, 46–78.

Dacorogna, M. M., Gencay, R., Müller, U., Olsen, R. B. and Pictet, O. V. (2001), *An Introduction to High-Frequency Finance*, Academic Press, London.

Dacorogna, M. M., Müller, U. A., Olsen, R. B. and Pictet, O. V. (1998), Modelling short term volatility with GARCH and HARCH, *in* C. Dunis and B. Zhou, eds, 'Nonlinear Modelling of High Frequency Financial Time Series', Wiley, Chichester, pp. 161–176.

de Pooter, M., Martens, M. and van Dijk, D. (2008), 'Predicting the daily covariance matrix for S&P 100 stocks using intraday data but which frequency to use?', *Econometric Reviews* **27**, 199–229.

Deo, R., Hurvich, C. and Lu, Y. (2005), 'Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment', *Journal of Econometrics* **131**, 29–58.

Diebold, F. and Strasser, G. (2008), On the Correlation Structure of Microstructure Noise in Theory and Practice. Unpublished paper: University of Pennsylvania.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation', *Econometrica* **45**, 987–1007.

Engle, R. F. (2000), 'The econometrics of ultra-high frequency data', *Econometrica* **68**(1), 1–22.

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

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

Engle, R. F. and Russell, J. R. (1998), 'Autoregressive conditional duration:A new model for irregularly spaced transaction data', *Econometrica* **66**(5), 1127–1163.

Engle, R. and Sun, Z. (2007), When Is Noise Not Noise - A Microstructure Estimate of Realized Volatility. Unpublished Manuscript, New York University, Stern School of Business.

Falkenberry, T. N. (2001), High frequency data filtering, Technical report, Tick Data.

Figlewski, S. (1997), 'Forecasting volatility', *Financial Markets, Institutions & Instruments* **6**(1), 1–88.

Fleming, J., Kirby, C. and Ostdiek, B. (2003), 'The economic value of volatility timing using realised volatility', *Journal of Financial Economics* **67**, 473–509.

Foster, D. P. and Nelson, D. B. (1996), 'Continuous record asymptotics for rolling sample variance estimators', *Econometrica* **64**(1), 139–174.

Ghysels, E. (2000), 'Some Econometric Recipes for High-Frequency Data Cooking', *Journal of Business & Economic Statistics* **18**(2), 146–153.

Ghysels, E. and Jasiak, J. (1997), 'GARCH for irregularly spaced financial data: The ACD-GARCH model', *Studies in Nonlinear Dynamics and Econometrics* **2**, 133–149.

Ghysels, E., Santa-Clara, P. and Valkanov, R. (2006), 'Predicting volatility: getting the most out of return data sampled at different frequencies', *Journal of Econometrics* **131**, 59–95.

Griliches, Z. (1967), 'Distributed lags: A survey', *Econometrica* **35**, 16–49.

Hansen, P. R. (2005), 'A Test for Superior Predictive Ability', *Journal of Business and Economic Statistics* **23**, 365–380.

Hansen, P. R. and Horel, G. (2009), 'Quadratic Variation by Markov Chains', *working paper, Stanford University* .

Hansen, P. R., Huang, Z. and Shek, H. H. (2009), 'Realized GARCH: A Complete Model of Returns and Realized Measures of Volatility', *working paper, Stanford University* .

Hansen, P. R. and Lunde, A. (2005*a*), 'A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?', *Journal of Applied Econometrics* **20**, 873–889.

Hansen, P. R. and Lunde, A. (2005*b*), 'Consistent Ranking of Volatility Models', *Journal of Econometrics* **131**, 97–121.

Hansen, P. R. and Lunde, A. (2006), 'Realized Variance and Market Microstructure Noise', *Journal of Business and Economic Statistics* **24**, 127–218. The 2005 Invited Address with Comments and Rejoinder.

Hansen, P. R. and Lunde, A. (2010), 'Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error'. CREATES, Aarhus University and Stanford University.

Hansen, P. R., Lunde, A. and Nason, J. M. (2003), 'Choosing the Best Volatility Models: The Model Confidence Set Approach', *Oxford Bulletin of Economics and Statistics* **65**, 839–861.

Hasbrouck, J. (2007), *Empirical Market Microstructure: Economic and Statistical Perspectives on the Dynamics of Trade in Securities Markets*, Oxford University Press.

Hillebrand, E. and Medeiros, M. C. (2008), 'The Benefits of Bagging for Forecast Models of Realized Volatility', *Econometric Reviews, forthcoming* .

Hillebrand, E. and Medeiros, M. C. (2009), Asymmetries, Breaks, and Long-Range Dependence: An Estimation Framework for Time Series of Daily Realized Volatility. Unpublished Manuscript, Department of Economics, Lousiana State University.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. and Vetter, M. (2009), 'Microstructure noise in the continuous case: the pre-averaging approach', *Stochastic Processes and Their Applications* **119**, 2249–2276.

Jorion, P. (1995), 'Predicting Volatility in the Foreign Exchange Market', *Journal of Finance* **50**(2), 507–528.

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**, 445–475.

Lanne, M. (2007), 'Forecasting realized exchange rate volatility by decomposition', *International Journal of Forecasting* **23**, 307–320.

Laurent, S., Rombouts, J. V. K. and Violante, F. (2009), Consistent ranking of multivariate volatility models, CORE Discussion Papers 2009002, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Li, K. (2002), 'Long-memory versus option-implied volatility predictions', *Journal of Derivatives* **9**, 9–25.

Li, Y. and Mykland, P. (2007), 'Are volatility estimators robust to modelling assumptions?', *Bernoulli* **13**, 601–622.

Maheu, J. M. and McCurdy, T. H. (2002), 'Nonlinear features of realized FX volatility', *Review of Economics & Statistics* **84**, 668–681.

Martens, M. (2001), 'Forecasting daily exchange rate volatility using intraday returns', *Journal of International Money and Finance* **20**(1), 1–23.

Martens, M. (2002), 'Measuring and forecasting S&P 500 index futures volatility using high-frequency data', *Journal of Futures Markets* **22**(6), 497–518.

Martens, M. and Zein, J. (2002), 'Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility', *Journal of Futures Markets* **24**(11), 1005–1028.

McAleer, M. and Medeiros, M. C. (2008), 'A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries', *Journal of Econometrics* **147**, 104–119.

McAleer, M. and Medeiros, M. C. (2009), 'Forecasting Realized Volatility with Linear and Nonlinear Models', *Journal of Economic Surveys, forthcoming* .

Meddahi, N. (2001), An eigenfunction approach for volatility modeling, Technical report, Cirano Working Paper 2001s-70.

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

Meddahi, N. (2003), 'ARMA representation of integrated and realized variances', *Econometrics Journal* **6**, 335–356.

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

Müller, U., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsacker, J. E. (1997), 'Volatilities of different time resolutions - Analyzing the dynamics of market components', *Journal of Financial Economics* **4**, 213–239.

Mykland, P. A. and Zhang, L. (2009), 'Inference for continuous semimartingales observed at high frequency: A general approach', *Econometrica* **77**, 1403–1445.

Nerlove, M. (1972), 'Lags in economic behavior', *Econometrica* **40**, 221–251.

NYSE, I. (2007), TAQ 3 Users Guide. New York Stock Exchange, Inc.

Oomen, R. A. C. (2001), 'Using high-frequency data stock market index data calculate, model and forecast realized return variance'. European University Institute, Economics Discussion Paper No. 2001/6.

Oomen, R. A. C. (2004), 'Modelling Realized Variance when Returns are Serially Correlated'. Manuscript, Warwick Business School.

Patton, A. (2008), 'Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies', *Forthcoming in Journal of Econometrics* .

Patton, A. J. and Sheppard, K. (2009*a*), 'Good Volatility, Bad Volatility: Signed Jumps and the Persistence of Volatility', *Working paper, Oxford-Man Institute of Quantitative Finance, Oxford UK* .

Patton, A. and Sheppard, K. (2009*b*), Evaluating Volatility and Correlation Forecasts, *in* T. G. Andersen, R. Davis, J.-P. Kreiss and T. Mikosch, eds, 'Handbook of Financial Time Series', Springer Verlag, pp. 801–838.

Pong, S., Shackleton, M. B., Taylor, S. J. and Xu, X. (2004), 'Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models', *Journal of Banking and Finance* **28**(10), 2541–2563.

Racicot, F., Théoret, R. and Coön, A. (2008), 'Forecasting Irregularly Spaced UHF Financial Data: Realized Volatility vs UHF-GARCH Models', *International Advances in Economic Research* **14**, 112–124.

Robinson, P. M. (1995), 'Non-parametric estimation for time series', *Annals of Statistics* **23**, 1048–1072.

Roll, R. (1984), 'A simple implicit measure of the effective bid-ask spread in an efficient market', *Journal of Finance* **39**, 1127–1139.

Shephard, N. and Sheppard, K. (2009), 'Realising The Future: Forecasting With High-Frequency-Based Volatility (HEAVY) Models', *Journal of Applied Econometrics, forthcoming* .

Tay, A. S., Ting, C., Tse, Y. K. and Warachka, M. (2009), 'The impact of transaction duration, volume and direction on price dynamics and volatility', *Quantitative Finance, forthcoming* .

Taylor, S. J. (1982), Financial returns modelled by the product of two stochastic processes — A study of daily sugar prices 1961-79, *in* O. D. Anderson, ed., 'Time Series Analysis: Theory and Practice, 1', North-Holland, Amsterdam, pp. 203–226.

Wood, R. A. (2000), 'Market Microstructure Research Databases: History and Projections', *Journal of Business & Economic Statistics* **18**(2), 140–145.

Zhang, L. (2006), 'Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach', *Bernoulli* **12**, 1019–1043.

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**, 1394–1411.

Zhou, B. (1996), 'High-frequency data and volatility in foreign exchange rates', *Journal of Business & Economic Statistics* **14**(1), 45–52.