

2008

Realized volatility uncertainty

David E. Allen

Michael McAleer

Marcel Scharth

Follow this and additional works at: <https://ro.ecu.edu.au/ecuworks>



Part of the [Statistics and Probability Commons](#)

Realized Volatility Uncertainty

By

David E. Allen¹, Michael McAleer² and Marcel Scharth³

¹School of Accounting, Finance and Economics, Edith Cowan University

²UWA Business School, University of Western Australia

³Tinbergen Institute, University of Amsterdam

School of Accounting, Finance and Economics & FEMARC Working Paper Series
Edith Cowan University
August 2008
Working Paper 0807

Correspondence author:

David E. Allen
School of Accounting, Finance and Economics
Faculty of Business and Law
Edith Cowan University
Joondalup, WA 6027
Australia
Phone: +618 6304 5471
Fax: +618 6304 5271
Email: d.allen@ecu.edu.au

ABSTRACT

The presence of high and time-varying volatility of volatility and leverage effects bring additional uncertainty in the tails of the distribution of asset returns, even though returns standardized by (ex-post) quadratic variation measures are nearly gaussian. We argue that in this setting modeling shocks to volatility is more relevant for applications than extracting more precise predictions of the variable, as point forecasts differences are swamped by the size of the volatility of volatility and rendered less informative by the nongaussianity in the ex-ante distribution of returns. Using S&P 500 data, we document that this volatility of volatility is subject to strong leverage effects, short-lived explosive regimes and is strongly and positively related to the level of volatility. Starting from a mixing hypothesis for volatility and returns, we propose a Monte Carlo method for translating these features into refined density forecasts for returns from our conditional volatility, skewness and kurtosis framework. We show how this method brings large improvements over traditional approaches for density and risk forecasting.

Keywords: Realized volatility; Volatility of volatility; Value-at-risk; Expected shortfall; Forecasting; Empirical finance

Acknowledgements: The authors are grateful to Marcelo Medeiros and participants of the Modeling and Managing Ultra-High Frequency Data conference for their helpful comments. Allen and McAleer are grateful for funding support from the Australian Research Council (ARC).

1. INTRODUCTION

The advent of ultra high frequency stock market data and the subsequent introduction of realized volatility measures¹ represented a substantial step forward in the accuracy with which econometric models of volatility could be evaluated (Andersen and Bollerslev, 1998) and allowed for the development of new and more precise parametric models of time varying volatility. Several researchers have looked into the properties of *ex post* volatility measures derived from high frequency data and developed time series models that invariably outperform latent variable models of the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) family of models (Andersen et al., 2003) on forecasting future volatility, to the point that the comparison has been dropped altogether in recent papers.

As this new literature emerged, the leading property observed in realized volatility which served as the main foundation for models that have been developed has been that of very high persistence observed in those series, which consistently exhibit empirical autocorrelation functions with hyperbolic decaying patterns; this empirical feature is referred to in a broad sense as long memory or long range dependence. Motivated by early empirical work in macroeconomics and later in the GARCH and Stochastic Volatility literature (for instance Baillie et al., 1996, Comte and Renault, 1996, and Breidt et al., 1998), the first prominent references in the realized volatility literature (e.g., Andersen et al., 2003) turned to the econometrics of fractional integration to establish a framework for modeling and forecasting realized volatility. Fractionally integrated processes ($I(d)$, where $0 < d < 1$) can be seen as a halfway paradigm between short memory ($I(0)$) and infinite memory ($I(1)$) processes and are able to render hyperbolic decaying patterns in autocorrelations and other properties that formally characterize a long memory model (see Baillie, 1996, for an exposition and formal definitions). Although no theoretical foundation has been developed to substantiate the long memory specification or elucidate the high persistence from past shocks, $I(d)$ processes emerged as a consonant description of the data generating process of volatility series, becoming a standard paradigm; modeling the long range dependence in the volatility of stocks and foreign exchange rates is among the greatest empirical successes of fractionally integrated models.

As authors looked into modeling and forecast refinements that could be achieved by changing the long memory specification and by considering additional regressors soon other models followed in the realized volatility literature. As new theoretical results demonstrated how long memory properties are not distinctive of fractionally integrated models but can be originated for example from stochastic regime switching or occasional structural breaks that generate slowly decaying autocorrelations and other properties of $I(d)$ processes (see also Mikosch and Starica, 2004, Hillebrand, 2005), empirical work suggested that both fractional integration and structural changes can describe the volatility of asset returns (Lobato and Savin, 1998, Martens et al., 2004, Beltratti and Morana, 2006, Morana and Beltratti, 2004, Hyung and Franses, 2002), some researchers begin to apply simpler time series models which are consistent with high persistence in relevant horizons, even though they do not rigorously exhibit long memory (hence being labeled as

¹Realized variance is defined as the sum of squared intraday returns sampled at a sufficiently high frequency, consistently approximating the integrated variance over the fixed interval where the observations are summed. Realized volatility is the squared-root of the realized variance. In practice, high frequency measures are contaminated by microstructure noise such as bid-ask bounce, asynchronous trading, infrequent trading, price discreteness, among others; see Biais et al. (2005). Ignoring the remaining measurement error, this *ex post* volatility measure can be modeled as an “observable” variable, in contrast to the latent variable models. See Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) for the theoretical foundations of realized volatility. Several recent papers have proposed corrections to estimation of RV in order to take the microstructure noise into account; see McAleer and Medeiros (2008b) and Gatheral and Oomen (2007) for reviews. In this paper we refer to realized volatility as a consistent estimator of the squared root of the integrated variance.

'quasi-long memory' models). Some examples are the MIDAS (mixed data sample, see for example Ghysels et al., 2007) and HAR (heterogeneous autoregressive, Corsi, 2004) models, both of which explore data sampled at different frequencies, and the unobserved ARMA component (UC) of Koopman et al. (2005).

Other contributions for the realized volatility modeling and forecasting literature are exemplified by Martens et al. (2004), who develop a nonlinear (ARFIMA) model to accommodate level shifts, day-of-the-week effects, leverage effects and volatility level effects, Andersen et al. (2007) and Tauchen and Zhou (2005) who argue that the inclusion of jump components significantly improves forecasting performance. McAleer and Medeiros (2008a) extend the HAR model to account for nonlinearities. Hillebrand and Medeiros (2007) also consider nonlinear models and evaluate the benefits of bootstrap aggregation (bagging) for volatility forecasting. Ghysels et al. (2007) argue that realized absolute values outperform square return-based volatility measures in predicting future increments in quadratic variation. Scharth and Medeiros (2006) introduce multiple regime models linked to asymmetric effects. Liu and Maheu (2007) derive a bayesian averaging approach for forecasting realized volatility.

What are after all the characteristics of realized volatility series that can have significant impact for risk management and other applications? This paper claims that despite not having received a great deal of attention so far, an essential property of the observed realized volatility series is that they exhibit a very large degree of volatility themselves and that this volatility of volatility is itself time varying; periods of stable and more predictable volatility alternate with episodes where the series displays large swings and assume values within a potentially broad range. This has important implications for the tails of the distribution of returns.

We will explore in detail the reasons behind this claim. First, the presence of high and time-varying volatility of volatility means high uncertainty in the tails of the distribution of daily returns. If the volatility in the next day is relatively unpredictable, then a conditional expectation of this variable won't contain much information about what might happen in terms of very negative returns. Suppose a naive practitioner calculates a point forecast of volatility and then obtains a given value-at-risk coverage assuming a normal distribution for returns given volatility based on the empirical evidence that returns standardized by well measured realized volatility are very closely normally distributed. With a large volatility of volatility, the volatility itself may turn out to be much higher than the point forecast and extreme returns that would be virtually impossible with the distribution based on the point forecast may be observed. Even though returns standardized by (ex-post) quadratic variation measures are nearly gaussian, returns standardized by fitted or predicted values of time series volatility models are far from normal. Given the uncertainty in volatility this is expected and should not be seen as evidence against those models; explicitly modeling the higher moment is necessary.

Second, we argue that forecasting improvements brought by the body of work discussed previously are marginal and swamped by the size of the volatility of volatility. In an extreme example, Scharth and Medeiros (2006) calculate that even a simple exponentially weighted moving average (EWMA) of realized volatility delivers predictions which are very close to ARFIMA and HAR specifications; it is crucial to account for the fact that the series is highly persistent, but the way this is done has very little economic relevance. Hence, it is very easy to predict the level of volatility in relation to the history of the series –but there is not much point to it given the uncertainty in the variable. We thus take the view that even though point forecasts have been the main output from which volatility models in general have been evaluated (extensive comparisons of forecasting performance like the one performed by Hansen and

Lunde (2005) are common in the literature) those statistics do not necessarily convey much information about the relative economic significance of the volatility models (see Fleming et al., 2001, 2003, W. Chan and Kalimipalli, 2006, for empirical analysis of the economic value of volatility models). In particular, and perhaps not surprisingly, small and possibly statistically insignificant forecasting performance differences may overshadow important relative modeling qualities.

When high frequency data was not widely available, the volatility of volatility could not be observed and latent volatility models were the only available option, the typical solution of the literature for volatility models that did not generate normally distributed standardized returns was to assume an ad hoc distribution for returns conditional on volatility that would sufficiently inflate the tails. The implications of the volatility of volatility for the tails of returns were understood and this modification could account in part for the mixing properties of volatility and returns. However, this is not entirely satisfactory when the volatility of volatility is time varying and consequently the size of the tails is not constant. Hence, the first building block of our analysis is to analyze the time series properties of the volatility of volatility of the S&P 500 index. To do so, we turn to a concept of realized quarticity analyzed in Barndorff-Nielsen and Shephard (2002, 2004, 2006) and Andersen et al. (2007), which can be seen as estimate for the variance of the return variation, analyzing that the volatility of volatility is characterized by long memory properties, strong leverage effects, short-lived explosive regimes and high correlation with the level of volatility. To the best of our knowledge, the last three of these characteristics haven't been documented in the literature so far.

The next step is to directly model both the volatility and the volatility of volatility. Corsi et al. (2008), in the first paper to consider the volatility of realized volatility, implement a latent variable approach and extend the typical framework for modeling volatility by specifying a GARCH process to allow for clustering in the squared residuals of those realized volatility models and assuming a normal inverse Gaussian (NIG) distribution to accommodate fat tailedness and skewness in the distribution of the residuals (which as we illustrate also cause excessive kurtosis on the ex-ante distribution of returns). This approach is shown by the authors to enhance the properties of estimations of realized volatility models. We suggest, however, that this approach should be extended to make use of the enhanced information available in the context of high-frequency data. In particular, modeling the (possibly non-linear) relation between the volatility of volatility and the level of volatility relevantly improve the realism of the model and the efficiency of the estimations. On the other hand, we show that despite being informative current realized quarticity measures have little practical importance.

With the assumptions for the realized volatility process and following the evidence in assuming that returns conditional volatility are normally distributed we have a mixing hypothesis that will enable us to propose a straightforward two step Monte Carlo method for calculating value-at-risk, expected shortfall and other density related measures for returns. The procedure consists in first simulating realized volatility and then using each of these simulated values to simulate returns. The empirical distribution function of the simulated returns can then be used for obtaining a prediction of value-at-risk and other density related measures of interest.

Our empirical analysis uses tick-by-tick data from the S&P 500 index from 1996 to 2007 to analyze the performance of the extended realized volatility with GARCH innovations model in conjunction with our Monte Carlo method in the calculation of density forecasts, value-at-risk and expected tail loss measures. The results suggest that the method of using only the point forecast of volatility is indeed not sufficient for obtaining adequate coverage and systematically underestimates the VaR intervals, but the introduction of

the Monte Carlo method corrects this failure in the lower tail, albeit the results are significantly better when intraday leverage effects are taken into account. In the case of expected shortfall, our method significantly improves forecasts and the results strongly favor specifications with time varying volatility of volatility and asymmetric effects. Our results are stronger than the ones obtained by Corsi et al. (2008) in that we can conclude that ignoring time varying volatility of volatility and intraday leverage effects render risk measures strongly biased and density forecasts inaccurate.

This paper is structured as follows. The next section presents our empirical setup, describing our data and realized volatility measurement, our general time series model for realized volatility and then finally proceeds to discuss the properties of the ex-ante distribution of the S&P500 returns when high-frequency data information is available, which guides our subsequent analysis. Section three discusses the introduction of time varying volatility of volatility into our general realized volatility time series model and describes how Monte Carlo techniques can be used for translating the features of our conditional volatility, skewness and kurtosis framework into refined density forecasts for returns. In section four we compare the out-of-sample point forecasting, value-at-risk, expected tail loss and density forecasting performances of different specifications. The implications of our analysis for the asset pricing and risk literatures are discussed in the conclusion.

2. EMPIRICAL SETUP

2.1. Data and realized volatility measurement. The empirical analysis focuses on the realized volatility of the S&P 500 index. The raw intraday data was obtained from the TaqTiq/SIRCA (Securities Industry Research Centre of Asia-Pacific) database and is constituted of tick-by-tick quotes originated in the Chicago Mercantile Exchange futures market. The period of analysis starts in January 2, 1996, and ends in March 29, 2007, providing a total of 2798 trading days. We start by removing non-standard quotes, computing mid-quote prices, filtering possible errors, and obtaining one second returns for the 9:30am to 16:00 p.m. period. Following the results of Hansen and Lunde (2006), we adopt the *previous tick* method for determining prices at precise time marks.

The search for unbiased, consistent and efficient methods for measuring realized volatility in the presence of microstructure noise has been one of the most active research topics in financial econometrics over the last years. While early references such as Andersen et al. (2001) suggest the simple selection of an arbitrary lower frequency (typically 5-15 minutes) to balance accuracy and dissipation of microstructure bias, a procedure known as *sparse sampling*, a number of recent articles developed estimators that dominate this procedure. Those contributions falls in several categories: some examples are the selection of optimal sampling frequency in sparse sampling Bandi and Russell (2005), subsampling (Zhang et al., 2005 and Aït-Sahalia et al., 2005), kernel-based estimators (Zhou, 1996, Barndorff-Nielsen et al., 2005) and MA filtering Hansen et al. (2008). See McAleer and Medeiros (2008b) and Gatheral and Oomen (2007) for a review and comparison of methods. In this paper, we turn to the theory developed by Barndorff-Nielsen et al. (2007a,b) and implement a subsampled realized kernel estimator based on one minute returns and the modified Tukey-Hanning kernel, which is consistent in the presence of microstructure noise.

2.2. Modeling Realized Volatility. Our general specification is given by:

$$(1) \quad \phi_p(L)(1-L)^d(RV_t - \mu(W_t)) = X_t\lambda + h_t(h_{t-1}, \varepsilon_{t-1}, Z_t)\varepsilon_t$$

where here d denotes the fractional differencing parameter, L the lag operator, ε_t is *i.i.d* with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) = 1$, $\phi_p(L)$ is a polynomial of order L , W_t is a vector of variables affecting the mean of RV_t , X_t is a vector of explanatory variables and Z_t is a vector of variables that may affect the volatility of volatility. We discuss below the specification for the conditional mean of volatility and the distribution of the errors; we postpone the analysis of heteroskedasticity for subsequent subsections.

2.2.1. Autoregressive Fractionally Integrated specification. When $-0.5 < d < 0.5$, we have the a stationary Autoregressive Fractionally Integrated model for the realized volatility. After running a battery of specification tests centered on the Schwarz Information Criterion, we set $\phi_p(L) = (1 - \phi_1 L)$ (that is, a *ARFIMA*(1, d , 0) model) for our all estimations.

Such models have been extensively estimated for realized volatility, for example in Andersen et al. (2003), Areal and Taylor (2002), Beltratti and Morana (2005), Deo et al. (2006), Martens et al. (2004), Thomakos and Wang (2003), among others. Nevertheless, the estimation of ARFIMA models in the context realized volatility has a few shortcomings. As remarked in the introduction, although $I(d)$ processes are a seemingly reasonable approximation for the data generating process of volatility series, there is no underlying theory to formally support this specification. Instead, the results of Diebold and Inoue (2001) and Granger and Hyung (2004) further challenge fractional integration as the correct long memory specification for realized volatility series. Statistical tests for distinguishing between those alternatives, such as the one proposed Ohanissian et al. (2004), have been hampered by low power.

Additionally, Granger and Ding (1996) and Scharth and Medeiros (2006) discuss how estimates of the fractional differencing parameter are subject to excessive variation over time, occasionally causing very poor in sample performance; this is less of a problem for forecasting, as long as the predictions come from an updated estimation. Finally, another problem that often arises in empirical applications is the estimation of highly non-stationary fractional differencing parameters, which in principle cannot be justified. Scharth and Medeiros (2006) argue that this is due to the presence of multiple regimes, which increase the persistence of the series and bias the d estimator up. Therefore, we keep in mind that the evidence casts doubt on the use of fractionally integrated processes suggesting they are at least incomplete descriptions of the conditional mean of volatility processes.

2.2.2. Heterogenous Autoregressive (HAR) specification. The HAR (Heterogeneous Autoregressive) model proposed by Corsi (2004) is an unfolding of the Heterogeneous ARCH (HARCH) model developed by Müller et al. (1997). It is specified as a multi-component volatility model with an additive hierarchical structure, leading to an additive time series model of the realized volatility which specifies the volatility as a sum of volatility components over different horizons.

Turning to our general specification, let $d = 0$ (underlining the view that the model does not genuinely exhibit long memory) and $\phi_p(L) = 1 - \phi_1 L - \phi_2 \sum_{i=1}^5 L^i - \phi_3 \sum_{i=1}^{22} L^i$. Furthermore, consider the notation $RV_{t,j} = \sum_{i=t-j+1}^t RV_{t,i}/j$, which will be used extensively in this paper. We can then write our HAR model with daily, weekly and monthly components as:

$$(2) \quad RV_t = \mu(W_t) + \phi_1 RV_{t-1} + \phi_2 RV_{t-2} + \phi_3 RV_{t-1,5} + \phi_4 RV_{t-1,22} + X_t \beta + h_t(Z_t) \varepsilon_t$$

We can see that the HAR specification is an *AR*(22) model rendered parsimonious by several parameter restrictions. Simulations reported in Corsi (2004) show that the generous number of autoregressive lags

renders the HAR model capable of reproducing the observed hyperbolic decay of the sample autocorrelations of realized volatility series over not too long horizons. Moreover, the model displays forecasting performance which is similar to that of ARFIMA models, which is generally true for any model that exhibits high persistence (and not necessarily authentic long memory properties).² For its estimation simplicity, the HAR-RV has been commonly favored in the high frequency econometrics literature (e.g, Andersen et al., 2007). Nevertheless, it is difficult to further justify the HAR model. One of its drawbacks is that it tends to estimate parameters that are generally inconsistent with each other when different direct forecasting estimations are implemented.

2.2.3. *Asymmetric Effects and Jumps.* Bollerslev et al. (2006) and Scharth and Medeiros (2006) highlight the impact of leverage effects on the dynamics of realized volatility. The latter argues for the existence of regime switching behavior in volatility, with large falls (rises) in prices being associated with persistent regimes of high (low) variance in stock returns. The authors show that the incorporation of cumulated daily returns as a explanatory variable brings modeling advantages by capturing this effect, which can be quite large; after analyzing certain stocks in the Dow Jones index, the authors document that falls in the horizon of less than two months are associated with volatility levels that are up to 60% higher than the average of periods with stable or rising prices. We estimate models with and without such effects. While Scharth and Medeiros (2006) consider multiple regimes in a nonlinear model, we focus on a simpler linear relationship to account for the large correlation between past cumulated returns and realized volatility. For reasons that will become clear over the next sections, we also consider the relation between returns and volatility on a same day; we shall refer to it as a volatility feedback effect.

Moreover, we consider jump components which have been receiving growing attention in the realized volatility literature. Building on theoretical results for bi-power variation measures, articles such as Andersen et al. (2007), Tauchen and Zhou (2005) and Barndorff-Nielsen and Shephard (2006) established related frameworks for the non-parametric estimation of the jump component in asset return volatility by explicitly considering the presence of less persistent elements in the volatility of stocks in contrast with the smooth and very slowly mean-reverting part associated with long memory properties. Empirically, Andersen et al. (2007) incorporates the distinction between jump and non-jump components into a forecasting model for the DM/USD exchange rate, the S&P500 market index, and the 30-year U.S. Treasury bond yield realized volatility series and find substantial performance improvements in daily weekly, and monthly predictions. In this paper, we follow Ghysels et al. (2007) and take the realized absolute variation (denoted RAV_t , calculated as the sum of intraday absolute returns) as a more robust measure of the persistent component in volatility, thus separating the effect of jumps. We find only the first lag of this variable to be significant, yielding in the least parsimonious case:³

$$\begin{aligned}
X_t \lambda = & \lambda_1 I(r_{t-1} < 0) r_{t-1} + \lambda_2 I(r_{t-1} > 0) r_{t-1} + \lambda_3 I(r_{t-1,5} < 0) r_{t-1,5} \\
(3) \quad & + \lambda_4 I(r_{t-1,5} > 0) r_{t-1,5} + \lambda_5 I(r_{t-1,22} < 0) r_{t-1,22} + \lambda_6 I(r_{t-1,22} > 0) r_{t-1,22} \\
& + \lambda_7 \sqrt{RAV_{t-1}} + I(r_t < 0) (r_t / RV_t) + I(r_t > 0) (r_t / RV_t),
\end{aligned}$$

²In fact, Scharth and Medeiros (2006) document that even a simple moving average of realized volatility achieves the same result.

³In what follows, we will make use of the abbreviations AE, VF and RAV in the acronyms of the models which contain leverage effects, volatility feedback and/or jumps respectively (for example, an HAR/AE model).

where the indicator functions have been included to reinforce the asymmetry between the effect of positive and negative returns and r_t/RV_t is interpreted as an exogenous shock following the standard normal distribution.

2.2.4. *Day of the Week and Holiday effects.* To reduce bias on our estimators and avoid distortions of the error distribution, we control the mean of the dependent variable for day of the week and holiday effects. Martens et al. (2004) and Scharth and Medeiros (2006) show that volatility sometimes tend to be lower on Mondays and Fridays, while substantially less volatility is observed around certain holidays.

2.2.5. *The Distribution of ε_t .* To account for the non-gaussianity of the error terms we follow Corsi et al. (2008) and assume that the (unconditional) *i.i.d.* innovations ε_t are distributed normal inverse Gaussian (NIG), which is flexible enough to allow for excessive kurtosis and skewness and reproduce a number of symmetric and asymmetric distributions (including the normal itself). A more convoluted approach would rely on the generalized hyperbolic distribution, which encompasses the NIG distribution and requires the estimation of an extra parameter. On the other hand, typical distributions with support on the interval $(0, \infty)$, which would be a desirable feature for our case, failed preliminary diagnostic tests. The density of the NIG distribution is given by:

$$(4) \quad f(x; \alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi} \frac{K_1 \left(\alpha \delta \sqrt{1 + \left(\frac{x-\mu}{\delta} \right)^2} \right)}{\sqrt{1 + \left(\frac{x-\mu}{\delta} \right)^2}} \exp \left\{ \delta \left(\sqrt{\alpha^2 - \beta^2} + \beta \left(\frac{x-\mu}{\delta} \right) \right) \right\}$$

where $K_i(x)$ is the modified Bessel function of the second kind with index i ; $\mu \in \mathbb{R}$ denotes the location parameter, $\delta > 0$ the scale, $\alpha > 0$ the shape, and $\beta \in (-\alpha, \alpha)$ the skewness parameter. Mean and variance are given by

$$(5) \quad \begin{aligned} E(x) &= \mu + \frac{\alpha}{\alpha^2 - \beta^2} \\ Var(x) &= \frac{\alpha^2}{\sqrt{\alpha^2 - \beta^2}^3}, \end{aligned}$$

so that the distribution is standardized by setting $\mu = -\frac{(\alpha^2 - \beta^2)}{\alpha^2}$ and $\delta = \frac{(\alpha^2 - \beta^2)^{3/2}}{\alpha^2}$

2.2.6. *Estimation.* The parameters are estimated by maximizing the log-likelihood function:

$$(6) \quad \begin{aligned} \ell(\hat{\phi}, \hat{\lambda}, \hat{\alpha}, \hat{\beta}; RV_{1..T}, X_{1..T}, Z_{1..T}) &= T \log(\hat{\alpha}) - T \log(\pi) + \sum_{t=1}^T \log \left[K_1(\hat{\alpha} \hat{\delta} (1 + \hat{y}_t^2)^{1/2}) \right] \\ &- 0.5 \sum_{t=1}^T \log(1 + \hat{y}_t^2) + T \hat{\delta} (\hat{\alpha}^2 - \hat{\beta}^2)^{1/2} + \hat{\delta} \hat{\beta} \sum_{t=1}^T \hat{y}_t \\ &- 0.5 \sum_{t=1}^T \log(\hat{h}_t^{1/2}) \end{aligned}$$

where $\gamma = (\hat{\alpha}^2 - \hat{\beta}^2)^{1/2}$ and $\hat{y}_t = \frac{\hat{\varepsilon}_t / \hat{h}_t^{1/2} - \hat{\mu}}{\hat{\delta}}$

In the case of the ARFIMA model, we employ a two-step estimation. In the first we apply the widely used log periodogram estimator (GPH) of Geweke and Porter-Hudak (1983) to filter the data. The number of ordinates ℓ used in each regression is selected by the plug-in method of Hurvich and Deo (1999). We then apply the maximum likelihood estimator above for the filtered series.

2.3. The Ex-Ante Distribution of S&P 500 returns. With our basic time series model for the realized volatility defined, we are ready to state the empirical problem at the center of our analysis. Even though it has been long recognized that the distributions of the stock returns scaled by realized standard deviations are approximately gaussian (e.g., Andersen et al., 2001), Figure 1 and Table 1 reveal that this is far from the case when we scale returns by the in-sample predicted values of our best fitting model in terms of in-sample forecasts (the HAR model with leverage effects and the square root of the realized absolute variation as explanatory variables). Despite the sizable forecasting gains made possible by volatility models based on high frequency data, our descriptive results can be directly related to the failure of GARCH volatility models to completely account for the excess kurtosis of returns (see for example Malmsten and Teräsvirta, 2004, Carnero et al., 2004). The researcher or practitioner interested in evaluating the density of returns from the perspective of a time series model still lives in a fat-tailed world and models of the conditional expectation of volatility have little to say about it (see figure 2 for a visual example taken from the S&P 500 series).

In this paper, we do not interpret those facts as evidence against those models, but as a consequence of high day-to-day unpredictability of the shocks that effect the volatility (causing excessive kurtosis) and the intraday leverage effects (negative skewness). In the following section, we will argue that an adequate volatility model for return density forecasting and risk management in this setting should illuminate the dynamics of the higher moments. To pursue this objective, we will turn to the idea of time-varying volatility of realized volatility (Corsi et al., 2008), which will allow for time-varying kurtosis on the general model introduced previously.

A simple motivation for this method can be summarized as follows. Suppose we fit a fat-tailed distribution to past returns standardized by a certain estimate of the conditional mean of volatility and try to use this framework for forecasting return densities in the future. This is the best possible alternative with latent volatility modeling. The problem with this approach, besides inefficiently relying on an unconditional distribution, is that there does not exist a large enough sample of returns to reliably assess the magnitude of the uncertainty in the tails of their distribution. We can say that the modeling qualities of our framework are enhanced not only from what happened in the data, but also from what *could* have happened within the much less strict limitations of our model. Extremely fat tailed days (i.e., days with high conditional volatility of volatility) are far more frequent than days when returns actually realize on them.

TABLE 1. Descriptive Statistics for Returns Standardized by Realized Volatility Predictions.

Statistic	Sample Estimate
Mean	0.012
Maximum	3.630
Minimum	-7.251
Std. deviation	1.195
Skewness	-0.384
Kurtosis	4.334

FIGURE 1. Empirical Density for Returns Standardized by Realized Volatility Predictions.

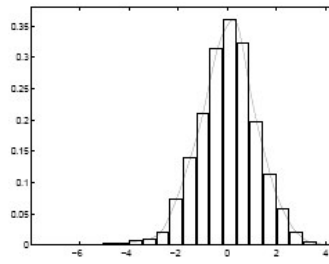
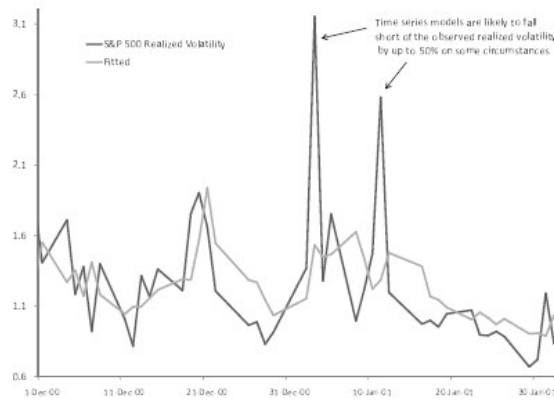


FIGURE 2. Why Shocks to Realized Volatility Matter.



3. VOLATILITY OF VOLATILITY

The theory of quadratic variation and integrated quarticity developed in Barndorff-Nielsen and Shephard (2002, 2004, 2006) and Andersen et al. (2007) points that a realized quarticity measure can be seen as estimate for the variance of the return variation. The initial point of our analysis is therefore to ask if this variable is sufficient for incorporating time varying volatility of volatility to the model described in the last section; if the realized quarticity is a good measure of the variance of realized variance, a more efficient maximum likelihood estimation would use this variable directly instead of relying on the latent variable (GARCH) approach implemented by Corsi et al. (2008). We begin by briefly reviewing the theory of quadratic variation and integrated variance and examining the time series properties of the realized quarticity measure.

3.1. Measuring the Realized Quarticity. Suppose that the logarithmic price of a financial asset, denoted by p_t , evolves according to the following stochastic volatility process:

$$(7) \quad p_t = p_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s)$$

where μ and σ are càdlàg; W is a standard Brownian motion; and σ is assumed to be independent of W . The quadratic variation process is defined as:

$$(8) \quad QV = \text{plim} \sum_{j=0}^{n-1} (p_{t_{j+1}} - p_{t_j})^2$$

Letting $\sup (t_{j+1} - t_j) \rightarrow 0$. The integrated quarticity is given by:

$$(9) \quad IQ = \int_0^t \sigma(s)^4 ds$$

Different estimators of integrated quarticity based on the theory of power variation developed by Barndorff Nielsen and Shephard (2002, 2004, 2006) have been proposed in the literature. So far these estimators deal with microstructure noise difficulties on a less advanced level than the recent realized volatility literature mentioned previously. Setting a day as a unit of measure, as usual, and sampling the continuously compounded intraday returns of day t at frequency M ,

$$r_{t,j} = p_{t-1+j/M} - p_{t-1+(j-1)/M},$$

the realized quarticity over day t is defined by

$$RQ_t = \frac{M}{3} \sum_{j=1}^M r_{t,k}^4 \rightarrow \int_0^t \sigma(s)^4 ds$$

A more robust estimator in the presence of jumps is the realized tri-power quarticity estimator proposed by Andersen et al. (2007):

$$RQ_t = M \frac{\Gamma(1/2)^2}{\Gamma(7/6)^3} \sum_{j=3}^M |r_{t,j}^4| |r_{t,j-1}^4| |r_{t,j-2}^4| \rightarrow \int_0^t \sigma(s)^4 ds$$

Our preferred estimator has been suggested by Barndorff-Nielsen et al. (2007a) and is given by:

$$RQ_t = \left(1 - \frac{2}{M}\right)^{-1} \sum_{j=3}^M (r_{t,j}^2 - 2\hat{\omega}^2)(r_{t,j-2}^2 - 2\hat{\omega}^2) \rightarrow \int_0^t \sigma(s)^4 ds$$

where $\hat{\omega}^2$ is an estimate of the microstructure noise variance.

3.2. Dynamic Properties of the Realized Quarticity: a Full Sample Analysis. The time series and empirical autocorrelation function of the square root of the realized quarticity estimator (the volatility of volatility) for the S&P500 index are depicted in Figures 3 and 4. As with the realized volatility, the series exhibit clustering behavior, alternating periods of stability which can last for years with periods where the series is subject to large swings and extreme spikes. The autocorrelations by their turn suggest that the volatility of volatility may also be characterized by long memory. However, we do not test this possibility formally as it is well known that long memory testing requires unrealistically large samples (Ohanissian et al., 2004, see for example). GPH estimates for the fractional differencing parameter of the realized quarticity and the square root of the realized quarticity are 0.199 and 0.447 respectively.

FIGURE 3. S&P 500 \sqrt{RQ}

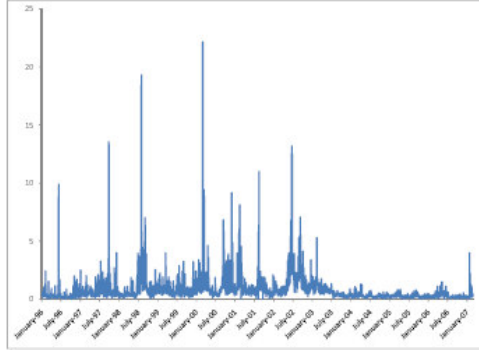
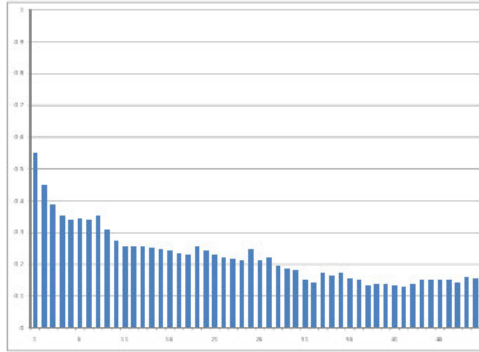


FIGURE 4. S&P500 \sqrt{RQ} : autocorrelation.



The following HAR model is estimated by OLS from the full sample:

$$\begin{aligned}
 \sqrt{RQ}_t &= 0.131 + 0.275\sqrt{RQ}_{t-1} \\
 &\quad \quad \quad (0.029) \quad \quad (0.022) \\
 &\quad + 0.085\sqrt{RQ}_{t-2} + 0.276\sqrt{RQ}_{t-i,5} \\
 &\quad \quad \quad (0.024) \quad \quad \quad (0.047) \\
 &\quad + 0.232\sqrt{RQ}_{t-i,22} + \hat{u}_t \\
 &\quad \quad \quad (0.037) \\
 R^2 &= 0.418, BIC = 2.900
 \end{aligned}
 \tag{10}$$

It is also interesting to evaluate possible relations between the realized quarticity and the other variables. The presence of markedly different patterns across different periods verified in Figure 3 suggest that the series can be described by multiple regimes. Following the the realized volatility specification and the results of Scharth and Medeiros (2006), we consider the possibility that the regime switching behavior is linked to asymmetric effects and can be accounted for by incorporating past cumulated returns as explanatory variables. Simulation results in Scharth and Medeiros (2006) suggest that this effect can account for some of the apparent long memory on the series.

The different estimations for the square root of the realized quarticity are given in Table 2, revealing strong evidence that leverage effects indeed impact the volatility of volatility and can be a strong predictor

for the variable. As in the realized volatility case, the realized absolute variation is also relevant. The large sample correlation between realized volatility and the square root of the realized quarticity (0.899) also renders the variable statistically significant. The last two cases are less interesting, however, as the explanatory power of the realized absolute variation and the realized volatility can be simply explained by the fact that they are less noisy explanatory variables than the lagged values of the realized quarticity.

TABLE 2. Realized Quarticity Estimations.

The table reports OLS estimations of linear models for \sqrt{RQ} consisting of (10) augmented by asymmetric effects, realized volatility and the absolute realized variation (RAV) for the full sample (January 2, 1996 to March 29, 2007, in a total of 2796 trading days).

	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Const.	-0.098	(0.035)	-0.354	(0.067)	-0.458	(0.059)	-0.024	(0.133)
\sqrt{RQ}_{t-1}	0.092	(0.023)	0.062	(0.035)	0.014	(0.032)	-0.293	(0.055)
\sqrt{RQ}_{t-2}	0.050	(0.023)	0.071	(0.024)	0.053	(0.024)	0.032	(0.023)
$\sqrt{RQ}_{t-1,5}$	0.194	(0.045)	0.248	(0.046)	0.242	(0.046)	0.167	(0.044)
$\sqrt{RQ}_{t-1,22}$	0.278	(0.036)	0.121	(0.039)	0.099	(0.038)	0.221	(0.038)
$I(r_{t-1} < 0)r_{t-1}$	-0.366	(0.034)	-	-	-	-	-0.356	(0.034)
$I(r_{t-1} > 0)r_{t-1}$	0.044	(0.032)	-	-	-	-	0.041	(0.031)
$I(r_{t-1,5} < 0)r_{t-1,5}$	-0.163	(0.019)	-	-	-	-	-0.149	(0.019)
$I(r_{t-1,5} > 0)r_{t-1,5}$	0.005	(0.016)	-	-	-	-	0.010	(0.015)
$I(r_{t-1,22} < 0)r_{t-1,22}$	-0.068	(0.009)	-	-	-	-	-0.063	(0.009)
$I(r_{t-1,22} > 0)r_{t-1,22}$	0.036	(0.008)	-	-	-	-	0.035	(0.008)
RAV_{t-1}	-	-	0.986	(0.122)	-	-	-0.041	(0.073)
RV_{t-1}^2	-	-	-	-	0.200	(0.017)	0.562	(0.075)
R^2	0.511		0.431		0.444		0.521	
BIC	2.742		2.879		2.856		2.727	

3.3. Modeling the Volatility of Volatility. In this section, we analyze the problem of incorporating the volatility of volatility on model (1). While the general interpretation of the realized quarticity as the volatility of volatility may loosely suggest that we could simply set h_t as some function of $RQ_t^{1/2}$, this approach suffers from a number of shortcomings. First, there is no theoretical justification for setting this variable as the conditional variance of the errors of the time series model, even though it is a strong indicator of much how the realized volatility is likely to fluctuate. Second, the realized quarticity estimator inherits both the problems of fourth-moment estimators (i.e., large sensibility) and market microstructure noise, with the latter being magnified in this case. In general, whenever realized quarticity is used by itself the optimization of (6) is less stable and maybe subject to convergence problems.

Despite these problems, it may be not necessary to throw the information contained in RQ_t away. Hence, we follow Corsi et al. (2008) in specifying a GARCH(1,1) model for the conditional heteroskedasticity of the realized volatility, but evaluate the effects of adding RQ_t as a explanatory variable and letting the estimation procedure determine the optimal way to weight the information contained in the residuals of the model and the realized quarticity. Even though the the last section suggests that a long memory model of the FIGARCH (Baillie et al., 1996) family maybe be more suitable, we find that this more complex to estimate specification does not bring additional modeling qualities; moreover, the presence of RQ_t should account for the long memory component if there is one. Initially, we consider the models $h_t = \theta_0 + \theta_1 h_{t-1} + \theta_2 \varepsilon_t^2$ and $h_t = \theta_0 + \theta_1 h_{t-1} + \theta_2 \varepsilon_t^2 + \theta_3 \sqrt{RQ_t}$. For conciseness we omit the estimation which relies solely on the realized quarticity for the conditional heteroskedasticity, which proved suboptimal at least with realized

quarticity estimators which are currently available. The estimated models are displayed below:

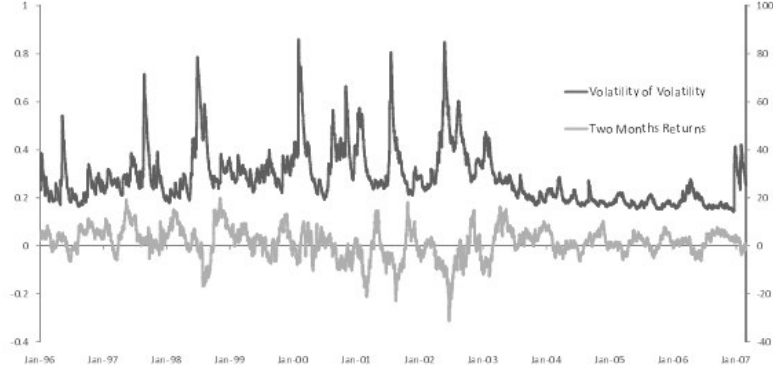
$$\begin{aligned}
\text{Model 1: } RV_t &= \underset{(0.017)}{-0.034} + \underset{0.014}{0.116} RAV_{t-1} - \underset{(0.020)}{0.010} RV_{t-1} \\
&+ \underset{(0.017)}{0.017} RV_{t-2} + \underset{(0.026)}{0.331} RV_{t-1,5} - \underset{(0.020)}{0.313} RV_{t-1,22} \\
&- \underset{(0.010)}{0.062} I(r_{t-1} < 0)r_{t-1} - \underset{(0.000)}{0.014} I(r_{t-1} > 0)r_{t-1} \\
&- \underset{(0.000)}{0.021} I(r_{t-1,5} < 0)r_{t-1,5} - \underset{(0.000)}{0.003} I(r_{t-1,5} > 0)r_{t-1,5} \\
&- \underset{(0.000)}{0.008} I(r_{t-1,22} < 0)r_{t-1,22} + \underset{(0.000)}{0.003} I(r_{t-1,22} > 0)r_{t-1,22} \\
&+ \text{Controls} + \hat{\varepsilon}_t \\
\hat{h}_t &= \underset{(0.000)}{0.000} + \underset{(0.014)}{0.893} h_{t-1} + \underset{(0.010)}{0.082} \hat{\varepsilon}_t^2 \\
\hat{\alpha} &= \underset{(0.136)}{1.808}, \hat{\beta} = \underset{(0.114)}{1.057}, R^2 = 0.693, BIC = -1384.9
\end{aligned}
\tag{11}$$

$$\begin{aligned}
\text{Model 2: } RV_t &= \underset{(0.017)}{-0.062} + \underset{0.014}{0.134} RAV_{t-1} - \underset{(0.022)}{0.029} RV_{t-1} \\
&+ \underset{(0.017)}{0.057} RV_{t-2} + \underset{(0.037)}{0.370} RV_{t-1,5} - \underset{(0.028)}{0.288} RV_{t-1,22} \\
&- \underset{(0.010)}{0.063} I(r_{t-1} < 0)r_{t-1} - \underset{(0.000)}{0.013} I(r_{t-1} > 0)r_{t-1} \\
&- \underset{(0.000)}{0.021} I(r_{t-1,5} < 0)r_{t-1,5} - \underset{(0.000)}{0.003} I(r_{t-1,5} > 0)r_{t-1,5} \\
&- \underset{(0.000)}{0.008} I(r_{t-1,22} < 0)r_{t-1,22} + \underset{(0.000)}{0.003} I(r_{t-1,22} > 0)r_{t-1,22} \\
&+ \text{Controls} + \hat{\varepsilon}_t \\
\hat{h}_t &= \underset{(0.000)}{0.001} + \underset{(0.032)}{0.804} h_{t-1} + \underset{(0.014)}{0.055} \hat{\varepsilon}_t^2 + \underset{(0.000)}{0.007} \sqrt{RQ}_t \\
\hat{\alpha} &= \underset{(0.166)}{2.100}, \hat{\beta} = \underset{(0.140)}{1.299}, R^2 = 0.696, BIC = -1439.6
\end{aligned}
\tag{12}$$

Even though the estimated volatility of volatility series for the pure GARCH(1,1) specification is fairly correlated with the square root of the realized quarticity (0.45, against 0.52 for the specification with the variable itself), the third model improves the efficiency of the estimation. The NIG parameters for the models one and two imply a skewness of 1.48 and 1.43 and a kurtosis of 8 and 7.52 for the unconditional distribution of the standardized residuals respectively. On an important note, the sample contemporaneous correlation between the volatility of volatility as fitted by the last model and the realized volatility is 0.78.

The estimated volatility of volatility series for the second model (that is, $\sqrt{\hat{h}_t}$) is depicted in Figure 5. At this point, we reintroduce the findings of the last section to our analysis and also visualize the relation between returns and the volatility of volatility. It is clear from the figure that as in the case of the conditional mean, the spikes in the volatility of volatility are related to very negative returns over a recent period. The variable then sharply falls back to average levels once positive returns are observed. Since the relation may not be stable and depends on the recent history of a particular period, we let the returns magnify or shrink the effect of past shocks over the volatility of volatility. Below, we estimate $h_t = \theta_0 + \theta_1 h_{t-1} + \theta_2 \varepsilon_t^2 + \theta_3 I(r_{t-1} < 0)r_{t-1} h_{t-1} + \theta_4 I(r_{t-1} > 0)r_{t-1} h_{t-1}$ as an example. Again, the inclusion of the additional explanatory variables in the return equation improves the descriptive power

FIGURE 5. S&P500 Estimated Volatility of Volatility.



of the model, as measured by the Schwarz Criterion. In this case, negative returns below the threshold of around -1% generate an explosive regime in the volatility of volatility. This result is driven by more negative returns and the spikes such as the one observed in Figure 5, which are more pronounced in this estimation.

$$\begin{aligned}
 \text{Model 3: } RV_t &= \underset{(0.017)}{-0.032} + \underset{(0.010)}{0.116} RAV_{t-1} - \underset{(0.022)}{0.010} RV_{t-1} \\
 &+ \underset{(0.017)}{0.071} RV_{t-2} + \underset{(0.021)}{0.298} RV_{t-1,5} - \underset{(0.020)}{0.329} RV_{t-1,22} \\
 &- \underset{(0.010)}{0.064} I(r_{t-1} < 0) r_{t-1} - \underset{(0.010)}{0.015} I(r_{t-1} > 0) r_{t-1} \\
 &- \underset{(0.000)}{0.023} I(r_{t-1,5} < 0) r_{t-1,5} - \underset{(0.000)}{0.004} I(r_{t-1,5} > 0) r_{t-1,5} \\
 &- \underset{(0.000)}{0.010} I(r_{t-1,22} < 0) r_{t-1,22} + \underset{(0.000)}{0.003} I(r_{t-1,22} > 0) r_{t-1,22} \\
 &+ \text{Controls} + \hat{\varepsilon}_t \\
 \hat{h}_t &= \underset{(0.000)}{0.002} + \underset{(0.017)}{0.828} h_{t-1} + \underset{(0.010)}{0.063} \hat{\varepsilon}_t^2 - \underset{(0.020)}{0.137} I(r_{t-1} < 0) r_{t-1} h_{t-1} \\
 &- \underset{(0.020)}{0.010} I(r_{t-1} > 0) r_{t-1} h_{t-1} \\
 \hat{\alpha} &= \underset{(0.151)}{2.061}, \hat{\beta} = \underset{(0.134)}{1.299}, R^2 = 0.698, BIC = -1391.2
 \end{aligned}
 \tag{13}$$

To conclude this section, we examine the effects of directly modeling the relation between the conditional expectation of the realized volatility and the volatility of volatility by considering the conditional heteroskedasticity model $h_t = \theta_0 + \theta_1 h_{t-1} + \theta_2 \varepsilon_t^2 + \theta_3 (RV_t - \hat{\varepsilon}_t) + \theta_4 (RV_t - \hat{\varepsilon}_t)^2$. The results are given in equation (14), confirming the relevant relation between the two variables. Perhaps surprisingly, our estimations indicate that the predicted value of RV_t for a given day is more informative for the volatility of volatility than the *ex post* realized quarticity measure.

$$\begin{aligned}
\text{Model 4: } RV_t = & \underset{(0.022)}{-0.072} + \underset{(0.010)}{0.141} RAV_{t-1} - \underset{(0.028)}{0.029} RV_{t-1} \\
& + \underset{(0.020)}{0.058} RV_{t-2} + \underset{(0.039)}{0.347} RV_{t-1,5} - \underset{(0.026)}{0.282} RV_{t-1,22} \\
& - \underset{(0.010)}{0.075} I(r_{t-1} < 0)r_{t-1} - \underset{(0.000)}{0.013} I(r_{t-1} > 0)r_{t-1} \\
& - \underset{(0.000)}{0.026} I(r_{t-1,5} < 0)r_{t-1,5} - \underset{(0.000)}{0.006} I(r_{t-1,5} > 0)r_{t-1,5} \\
& - \underset{(0.000)}{0.010} I(r_{t-1,22} < 0)r_{t-1,22} + \underset{(0.000)}{0.005} I(r_{t-1,22} > 0)r_{t-1,22} \\
& + \text{Controls} + \hat{\varepsilon}_t \\
\hat{h}_t = & \underset{(0.005)}{0.016} + \underset{(0.131)}{0.177} h_{t-1} + \underset{(0.023)}{0.073} \hat{\varepsilon}_t^2 - \underset{(0.015)}{0.042} (RV_t - \hat{\varepsilon}_t) \\
& - \underset{(0.015)}{0.072} (RV_t - \hat{\varepsilon}_t)^2 \\
\hat{\alpha} = & \underset{(0.193)}{2.207}, \hat{\beta} = \underset{(0.163)}{1.355}, R^2 = 0.702, BIC = -1449.8
\end{aligned}
\tag{14}$$

3.4. A Monte Carlo Method for Return Density Forecasting. In this section, we propose a method that will enable the application of the modeling framework described previously when the ex-ante density of daily returns is of interest. As the above discussion depicts, realized volatility series are subject to intense variation from day to day; the typical unconditional coefficient of variation for the models analyzed previously for the S&P 500 index lies in the $1/4 - 1/3$ range. As we will illustrate below, this implies that the distribution of returns standardized by the conditional mean of realized volatility departs significantly from the normal distribution verified when the same returns are standardized by the realized volatility.

Unfortunately, an analytical solution for the density implied by our flexible normal variance-mean mixture hypothesis (realized volatility is distributed normal inverse gaussian and returns given volatility are normally distributed) is not available. We then turn to the following Monte Carlo method, which can be easily implemented with any mixing hypothesis and that can be made very accurate with realistic computational power. Conditional on information up to day $t - 1$, the forecasted empirical density function for day t for the case of no intraday leverage effects are calculated as follows:

- (1) In the first step, the functional form of the model (equation 1) is used for the evaluation of predictions of the realized volatility and the volatility conditional on past realized volatility observations, returns, the estimated volatility of volatility series and shocks, and other variables.
- (2) We randomly generate n shocks distributed as the standardized NIG with the parameters estimated from the data as described in section 2.2.6, which multiplied by $\hat{h}_t^{1/2}$ and added to \widehat{RV}_t originate a vector of n simulated realized volatilities for day t , where \hat{x}_t denotes a prediction of a generic variable x for day t .
- (3) Under the hypothesis that standardized returns are normally distributed, we employ each of these n simulated volatilities to simulate m associated returns. Antithetic variables techniques can be used, so that m should be at least 2. The empirical density function of the set of *all* the $n \times m$ simulated returns yield our final density forecast.

For the case with intraday leverage effects, the procedure can be summarized as follows:

- (1) In the first step, n standardized shocks distributed as $N(0,1)$ are generated. The functional form of the model (equation 1, see also equation 3) is used for the evaluation of predictions of the realized volatility and the volatility conditional on past realized volatility observations, returns, the estimated volatility of volatility series and shocks, and other variables.
- (2) We randomly generate n shocks distributed as the standardized NIG with the parameters estimated from the data as described in section 2.2.6, which multiplied by $\tilde{h}_t^{1/2}$ and added to \widetilde{RV}_t originate a vector of n simulated realized volatilities for day t .
- (3) The simulated returns are given by the product of each simulated realized volatility with the respective standardized shock. The empirical density function of the set of all the n simulated returns yield our final density forecast.

In Figures 6, 7 below we illustrate the impact of the volatility of volatility for the ex-ante distribution of returns without the presence of intraday leverage effects. The assumed parameters for the NIG distribution correspond to the ones estimated on (11) and in the second figure we arbitrarily assume the conditional mean of realized volatility to be 1.2 (this is to rule out imperfections caused by the support of the NIG distribution). For low values of the variance of volatility (or more generally, for a low conditional coefficient of variation) the distribution of returns is still very close to the gaussian case. This is consistent with the evidence that returns standardized by the realized volatility are nearly normally distributed: since the impact of the volatility of volatility is non-linear and grows slowly with the variable, the effect of errors in the Barndorff-Nielsen et al. (2007a) realized volatility estimator are not enough to generate pronounced non-gaussianity on the scaled returns. Figure 6 also clarifies the consequences of the non-gaussianity on the distribution of realized volatility shocks, showing that the excess kurtosis on the volatility amplify the excess kurtosis on returns for a given volatility of volatility level, specially for higher values of the latter.

The case with intraday leverage effects is illustrated on Figure 8, where the parameters have again been estimated from the data and the conditional mean of realized volatility is 1.2. This setting directly implies that the ex-ante distribution of returns is visibly non-gaussian even for a very low volatility of volatility. As the model is specified, the coefficient of skewness grows with the volatility of volatility, jumping from -0.07 in the first panel (volatility of volatility = 0.1) to -0.93 in the last one (volatility of volatility = 0.55). We have also estimated alternative models to relax this restriction and verified that the effect is in fact present in the data, even though the skewness does not grow as much with the volatility of volatility.

4. APPLICATIONS

In this section we test our methods in a variety of applications favored in the literature. We base the out-of-sample analysis on the eight last years of the sample, ranging from 2000 to 2007 and covering 1557 trading days. All implementations re-estimate the models daily using the full past data to calculate the desired forecasts, which are all one step ahead.

4.1. Point Forecasting. Even though in the introduction to this paper we argued that point forecasting may not be a powerful tool for comparing volatility models, predictions have been the main basis of comparison in the volatility literature and are the subject of extensive analysis (e.g., Hansen and Lunde, 2005). Therefore, we start by analyzing the performance of the models described previously in this aspect as a first layer of comparison. The evaluation of forecasts is based on the mean absolute error (MAE), the root

FIGURE 6. Kurtosis of the simulated distribution under the assumption that shocks to realized have the NIG (black) and normal (grey) distributions.

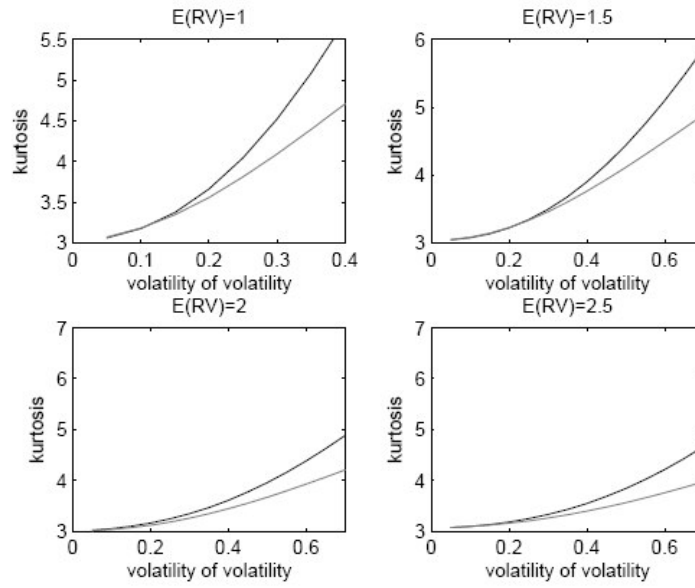
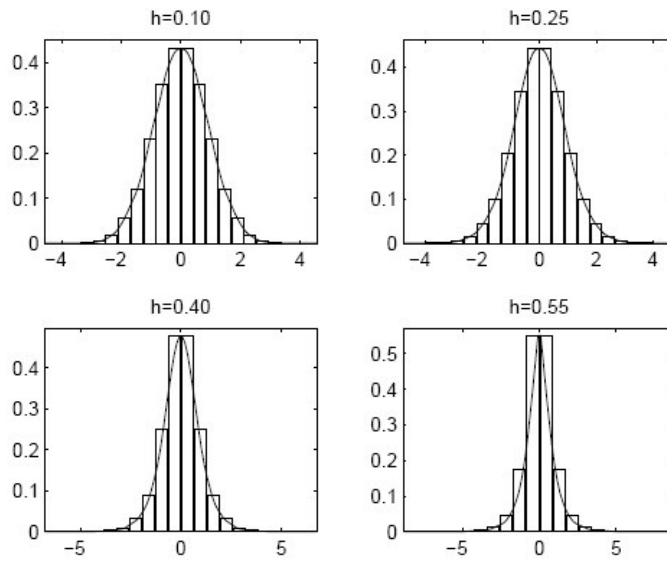


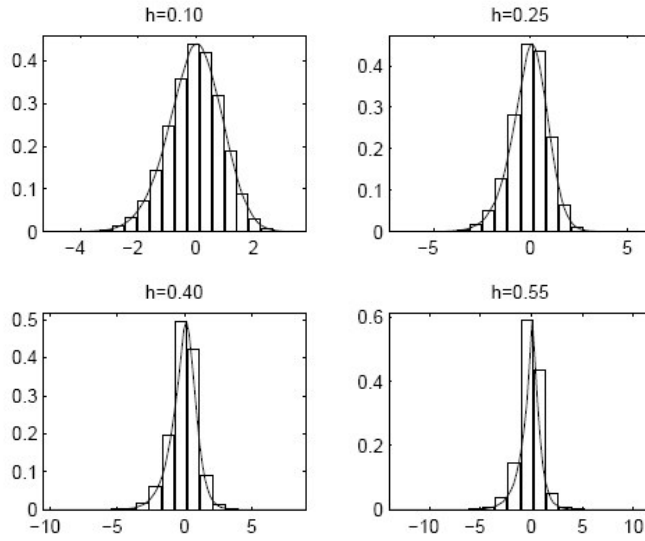
FIGURE 7. Densities for the simulated distributions (no volatility feedback).



mean squared error (RMSE) and the estimation of the Mincer-Zarnowitz regression

$$RV_t = \alpha + \beta \widetilde{RV}_{t-1,i} + \varepsilon_{t,i}$$

FIGURE 8. Densities for the simulated distributions (with volatility feedback).



where RV_t is the observed realized volatility on day t and $\widetilde{RV}_{t|t-1,i}$ is the one-step-ahead forecast of model i for the volatility on day t . If the model i is correctly specified then $\alpha = 0$ and $\beta = 1$. We report the (corrected) R^2 of the regression as a measure of the ability of the model to track variance over time and test of superior predictive ability (SPA) test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any other competing models in terms of a given loss function.

The forecasting statistics are displayed on Table 3. The results suggest that whenever cumulated returns are included in the conditional mean specification the respective model fares unambiguously better than its counterpart without leverage effects (on all criteria). The same is valid for the inclusion of the realized absolute variation (even though we include only one specification to our results to save space), though by a smaller margin than has been reported in the literature.⁴ In terms of the long-memory/quasi long-memory specification, the results indicate that the HAR model performs unambiguously the best, especially when leverage effects are not taken into account. Finally, the enhanced efficiency of the estimates that take time-varying volatility of volatility into account does not prevent these specification from adverse results: models with constant volatility of volatility perform significantly better, except in the case where we allow for the relation between the forecasted volatility and the volatility of volatility (labeled augmented GARCH model, which is (14) without RAV_t), which is the best one overall. Recalling one the argument from the introduction, we remark that despite the statistical significance of the comparison between some models, the typical size of the volatility of the volatility is more than 13 (and up to 30) times as large as the difference in the MAE between the best model and the simplest HAR specification; forecasting differences are swamped by the size of the noise.

⁴We hypothesize that jump issues as currently modeled are less relevant to forecasting when more efficient/less noisy realized volatility estimators are implemented.

TABLE 3. FORECASTING RESULTS.

The table reports the out-of-sample forecasting results for the S&P 500 daily realized volatility for the period between 2000 and 2007 (1557 trading days), where each model is re-estimated daily and used for one day ahead predictions. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is an explanatory variable for the volatility of volatility. RMSE is the root mean squared error and MAE the mean absolute error. R^2 is the (corrected) R-squared of $RV_t = \alpha + \beta \widehat{RV}_{t|t-j,i} + \varepsilon_{t,i}$, where $\widehat{RV}_{t|t-j,i}$ is the prediction of model i for the realized volatility on day t and RV_t is the observed realized volatility on that day. SPA is the p -value of the Superior Predictive Ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any other competing models in terms of a given loss function.

Model	R-Squared	RMSE	MAE	SPA in R^2	SPA in MSE
HAR	0.719	0.231	0.152	0.002	0.003
HAR/AE	0.763	0.212	0.143	0.333	0.431
HAR-GARCH	0.714	0.244	0.167	0.002	0.005
HAR/AE-GARCH	0.750	0.227	0.160	0.000	0.111
HAR/AE/RAV-GARCH	0.760	0.213	0.141	0.492	0.603
ARFIMA	0.682	0.245	0.158	0.003	0.007
ARFIMA/AE	0.743	0.221	0.147	0.023	0.039
ARFIMA-GARCH	0.685	0.244	0.158	0.003	0.004
ARFIMA/AE-GARCH	0.721	0.231	0.154	0.002	0.008
HAR/AE/VF-GARCH	0.746	0.229	0.162	0.000	0.036
HAR/AE-Augmented Garch	0.765	0.211	0.141	0.981	0.983
Combination (all)	0.766	0.213	0.143	–	–

4.2. **Value-at-Risk.** We now proceed to analyze the ability of different models and the Monte Carlo method in forecasting adequate quantiles over the lower tail of the return distribution. We also introduce a more standard method of density forecasting into our analysis, namely considering $r_t \sim N(0, \widehat{RV}_t)$, henceforth labeled point forecast method. We remark, however, that this is not intended as a “models race” but rather as basis for reference that illustrates the dangers of relying excessively on point forecasts and ignoring the volatility of volatility.

The evaluation of value-at-risk forecasts is based on the likelihood ratio tests for unconditional coverage and independence of Christoffersen (1998). Our analysis is similar to Beltratti and Morana (2005), who study the benefits of value-at-risk with long memory. Let $\widehat{q}_{t|t-1}^i(\alpha)$ be the $(1 - \alpha)$ interval forecast of model i for day t conditional on information on day $t - 1$. In our application, we consider 95% and 99% value-at-risk measures, i.e., $\alpha = 0.05$ and $\alpha = 0.01$, respectively. We construct the sequence of coverage failures for the lower α tail as:

$$F_{t|t-1} = \begin{cases} 1 & \text{if } r_{t+1} < \widehat{q}_{t+1|t}^i(\alpha) \\ 0 & \text{if } r_{t+1} > \widehat{q}_{t+1|t}^i(\alpha) \end{cases}$$

where r_t is the return observed on day t . The unconditional coverage (UC) is a test of the null $\mathbb{E}(F_{t+1|t}) = \alpha$ against $\mathbb{E}(F_{t+1|t}) \neq \alpha$. The test of independence is constructed against a first-order Markov alternative.

The value-at-risk performance of the models are organized and presented in Table 4. The results show that as expected from the analysis of section 3.4, the method of calculating VaRs based only on the point forecast of volatility is clearly biased towards underestimating the value-at-risk, failing to provide adequate coverage at all intervals. On the hand, the tables provide some weak evidence that taking the time varying

volatility of volatility into account in the estimation helps in reducing the bias of the point forecast method. The Monte Carlo method, in turn, significantly reduces or eliminates the problem of excess violations for all models (even though the experiment has in general no power in ranking them). In the 1% VaR case we cannot reject the hypothesis that most models considered produce accurate VaRs, with the difference in relation to the point forecasting method being significant in most cases (not reported in the table). For the 2.5% and 5% intervals, however, it becomes clear that the only model capable of generating a negatively skewed distribution (the model with intraday leverage effects) is adequate, with the rest underestimating possible losses.

4.3. Expected Shortfall. It is well known that value-at-risk measures suffer from a series of shortcomings as a tool for assessing the risk of assets or portfolios, for example by failing to provide an estimate of how large losses can be on the lower tail of the return distribution, relying instead on a threshold linked to a certain probability. With roots in the theory of coherent risk measures (Artzner et al., 1997, 1999), the expected shortfall (or conditional value-at-risk) is one of the concepts introduced to overcome this deficiency. The expected shortfall for day t is defined as

$$(15) \quad ES_t = E(r_t | r_t < Var_t^{0.99})$$

where $Var_t^{0.99}$ denotes the 99% value-at-risk for day t . Hence, the expected shortfall is defined as the expected value of the tail, making it a particularly attractive means of assessing the modeling qualities of different models with respect to the volatility of volatility, which as we have seen is directly linked to size of the tails and consequently the conditional value-at-risk. To do so, we construct a specialized sequence of standardized returns r_t/\bar{RV}_t , where \bar{RV}_t is the average from all models of the predicted value of the realized volatility, and compare the predicted expected shortfalls against the returns (r_t) associated to the lowest percentile of this sequence. The reason why we do so, and do not consider say $r_t/RV_t < \phi^{-1}(0.01)$ (where $\phi^{-1}(\cdot)$ is the inverse of the standard normal), is that it would make little sense to compare a statistic based on an ex-ante distribution (the forecasted expected shortfall) with another one based on an ex-post distribution. Therefore, low values of r_t/\bar{RV}_t represent the most unpredictably low returns in our sample, maximizing the degree of uncertainty that the models are being tested against.

The results for the point forecast based method and the Monte Carlo based method are displayed in tables 5 and 6 respectively, where the the evaluation of forecasts is again based on the mean absolute error (MAE), the mean squared error (MSE) and the estimation of the Mincer-Zarnowitz regression. We note that the small number of observations that are left for this experiment (15) leaves the SPA method with low power for statistically ranking models. For the point forecasted method, the analysis is related to that of forecasting, with the inclusion of RAV_t and asymmetric effects enhancing the quality of the specification (significant in the case of the MSE) and the inclusion of the GARCH(1,1) effect being detrimental on the R_2 criterion and practically ineffective on the other criteria. On the other hand, the ARFIMA have a relatively superior performance compared to the point forecasting analysis. The Monte Carlo method, in its turn, importantly reduces the size of the forecasting errors (for the best model the reduction of the MAE compared to the point forecasting method is of the order of 35%). In this case, asymmetric effects (cumulated returns and volatility feedback effects), time varying volatility of volatility and the relation between the level and the volatility of volatility improve the results, albeit the differences do not stand out

TABLE 4. VALUE AT RISK ANALYSIS.

The table reports the out-of-sample value-at-risk results for the S&P 500 daily realized volatility for the period between 2000 and 2007 (1557 trading days), where each model is re-estimated daily and used for calculating 1%, 2.5% and 5% value-at-risk thresholds by the traditional VaR method based only on the forecast of realized volatility and the Monte Carlo Method described in section 3.4. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is a explanatory variable for the volatility of volatility. The column failures indicates the proportion of days when returns over the next day in the α lower tail of the predicted distribution. UC and IND are the p -values of the likelihood ratio tests for unconditional coverage and independence (against a first order Markov alternative) developed by Christoffersen (1998).

	1% VaR					
	Point Forecast VaR			Monte Carlo Method		
	Failures	UC	IND	% Failures	UC	IND
HAR	0.021	0.000	0.239	0.015	0.077	0.417
HAR/AE	0.023	0.000	0.835	0.014	0.123	0.427
HAR-GARCH	0.019	0.001	0.587	0.011	0.720	0.552
HAR/AE-GARCH	0.017	0.008	0.472	0.014	0.189	0.271
HAR/AE/RAV-GARCH	0.021	0.000	0.668	0.014	0.123	0.427
ARFIMA	0.024	0.000	0.012	0.017	0.015	0.069
ARFIMA/AE	0.023	0.000	0.226	0.014	0.123	0.438
ARFIMA-GARCH	0.023	0.000	0.008	0.016	0.027	0.400
ARFIMA/AE-GARCH	0.021	0.000	0.181	0.015	0.047	0.386
HAR/AE/VF-GARCH	0.019	0.002	0.548	0.007	0.219	0.692
HAR/AE-Augmented Garch	0.022	0.000	0.202	0.014	0.123	0.366

	2.5% VaR					
	Point Forecast VaR			Monte Carlo Method		
	Failures	UC	IND	% Failures	UC	IND
HAR	0.048	0.000	0.417	0.036	0.009	0.059
HAR/AE	0.051	0.000	0.596	0.041	0.000	0.796
HAR-GARCH	0.041	0.000	0.364	0.032	0.085	0.293
HAR/AE-GARCH	0.039	0.002	0.326	0.032	0.085	0.614
HAR/AE/RAV-GARCH	0.048	0.000	0.417	0.037	0.004	0.929
ARFIMA	0.051	0.000	0.341	0.033	0.043	0.032
ARFIMA/AE	0.049	0.000	0.485	0.037	0.004	0.078
ARFIMA-GARCH	0.050	0.000	0.260	0.034	0.030	0.037
ARFIMA/AE-GARCH	0.046	0.000	0.354	0.033	0.061	0.027
HAR/AE/VF-GARCH	0.038	0.002	0.350	0.019	0.092	0.548
HAR/AE-Augmented Garch	0.048	0.000	0.371	0.033	0.061	0.580

	5% VaR					
	Point Forecast VaR			Monte Carlo Method		
	Failures	UC	IND	% Failures	UC	IND
HAR	0.084	0.000	0.472	0.076	0.000	0.489
HAR/AE	0.083	0.000	0.364	0.077	0.000	0.419
HAR-GARCH	0.079	0.000	0.261	0.077	0.000	0.194
HAR/AE-GARCH	0.076	0.000	0.483	0.073	0.000	0.662
HAR/AE/RAV-GARCH	0.084	0.000	0.794	0.076	0.000	0.250
ARFIMA	0.082	0.000	0.145	0.078	0.000	0.133
ARFIMA/AE	0.084	0.000	0.941	0.080	0.000	0.945
ARFIMA-GARCH	0.082	0.000	0.145	0.077	0.000	0.105
ARFIMA/AE-GARCH	0.082	0.000	0.569	0.076	0.000	0.707
HAR/AE/VF-GARCH	0.074	0.000	0.587	0.048	0.739	0.096
HAR/AE-Augmented Garch	0.082	0.000	0.880	0.076	0.000	0.729

statistically due to the testing limitations commented on previously (except for comparing specifications with or without asymmetric effects).

TABLE 5. POINT FORECAST BASED EXPECTED SHORTFALL.

The table reports the out-of-sample expected shortfall results for the S&P 500 daily realized volatility for the period between 2000 and 2007, where each model is re-estimated daily. The expected shortfall of model i is defined as $ES_{t,i} = E(r_t | r_t < Var_{t,i}^{0.99})$, where $Var_{t,i}^{0.99}$ denotes the 99% value-at-risk of model i for day t . The forecasts are compared only against returns that satisfy being below the first percentile of the sequence r_t/\overline{RV}_t , where \overline{RV}_t is the average from all models of fitted in-sample value of the realized volatility. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is a explanatory variable for the volatility of volatility. RMSE is the root mean squared error and MAE the mean absolute error. R^2 is the R-squared of $r_t = \alpha + \beta \widetilde{ES}_{t-j,i} + u_{t,i}$, where $\widetilde{ES}_{t-j,i}$ is the prediction of model i for the expected shortfall on day t . SPA is the p -value of the Superior Predictive Ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any other competing models in terms of a given loss function.

Model	R-Squared	RMSE	MAE	SPA in R^2	SPA in MSE
HAR	0.528	1.085	0.792	0.147	0.030
HAR/AE	0.643	0.973	0.730	0.128	0.438
HAR-GARCH	0.509	1.058	0.763	0.221	0.040
HAR/AE-GARCH	0.633	0.947	0.686	0.136	0.316
HAR/AE/RAV-GARCH	0.555	1.093	0.822	0.096	0.108
ARFIMA	0.609	0.887	0.619	0.602	0.411
ARFIMA/AE	0.673	0.841	0.604	0.808	0.858
ARFIMA-GARCH	0.595	0.889	0.624	0.538	0.290
ARFIMA/AE-GARCH	0.645	0.847	0.613	0.829	0.725
HAR/AE/VF-GARCH	0.644	0.928	0.667	0.633	0.499
HAR/AE-Augmented Garch	0.686	0.908	0.678	0.460	0.921

4.3.1. *Plausibility Check.* On table 7 below we conduct a brief counterfactual analysis of the lower tail performance of selected models under the Monte Carlo density forecasting. We display the inverse of the forecasted cumulative density functions evaluated at the ten lowest observed returns on the out-of-sample period (1557 trading days). This quantity can be interpreted as the expected number of days the model predicts for such a CDF (or lower) to occur. If no lower simulated returns were observed on the conditional Monte Carlo sample, $> 10^5$ is indicated. This simple analysis proves most realized volatility models to be implausible and provide further evidence that time accounting for the volatility of volatility and the possibility of next-day leverage effects(conditional skewness) is crucial for robustness.

4.4. **Whole Density Forecasts.** After evaluating the specific region of the forecasted return densities more critical to risk management, we turn now to a diagnostic test for the complete densities. We rely on the theory of density forecasting evaluation developed by (Diebold et al., 1998). The method consists of considering the predicted cumulative density functions evaluated at the observed returns, which originate a sequence of random variables which would be distributed as $U(0,1)$ and i.i.d. if the forecasts were identical to the true densities. Table 8 shows the p -values of Kolmogorov-Smirnov tests of the $U(0,1)$ hypothesis for the forecasted densities of all models, with a strong result: the model with intraday leverage effects is

TABLE 6. MONTE CARLO METHOD BASED EXPECTED SHORTFALL.

The table reports the out-of-sample expected shortfall results for the S&P 500 daily realized volatility for the period between 2000 and 2007, where each model is re-estimated daily. The expected shortfall of model i is defined as $ES_{t,i} = E(r_t | r_t < Var_{t,i}^{0.99})$, where $Var_{t,i}^{0.99}$ denotes the 99% value-at-risk of model i for day t . The forecasts are compared only against returns that satisfy being below the first percentile of the sequence r_t/\overline{RV}_t , where \overline{RV}_t is the average from all models of fitted in-sample value of the realized volatility. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). RMSE is the root mean squared error and MAE the mean absolute error. R^2 is the R-squared of $r_t = \alpha + \beta \widetilde{ES}_{t-j,i} + u_{t,i}$, where $\widetilde{ES}_{t-j,i}$ is the prediction of model i for the expected shortfall on day t . SPA is the p -value of the Superior Predictive Ability test developed by Hansen (2005). The null hypothesis is that a given model is not inferior to any other competing models in terms of a given loss function.

Model	R-Squared	RMSE	MAE	SPA in R^2	SPA in MSE
HAR	0.536	0.852	0.586	0.133	0.036
HAR/AE	0.648	0.769	0.496	0.173	0.373
HAR-GARCH	0.525	0.800	0.538	0.249	0.045
HAR/AE-GARCH	0.688	0.656	0.462	0.887	0.090
HAR/AE-GARCH/RAV	0.607	0.829	0.564	0.198	0.169
ARFIMA	0.611	0.773	0.504	0.318	0.438
ARFIMA/AE	0.673	0.695	0.464	0.678	0.664
ARFIMA-GARCH	0.523	0.965	0.562	0.222	0.124
ARFIMA/AE-GARCH	0.582	0.861	0.510	0.482	0.228
HAR/AE/VF-GARCH	0.717	0.624	0.530	0.708	0.836
HAR/AE-Augmented Garch	0.721	0.645	0.434	0.873	0.932

TABLE 7. Plausibility Check: Inverse of the Forecasted Empirical CDF at Low Observed Returns

The table reports the inverse of the forecasted cumulative density functions (using the Monte Carlo Method) evaluated at then lowest observed returns in the period between 2000 and 2007 (1557 trading days). The specification for the conditional mean and conditional heteroskedasticity are separated by dashes, with a G indicating a GARCH model (that is, time varying volatility of volatility). AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is a explanatory variable for the volatility of volatility.

Date	Return	HAR	HAR/AE	HAR-G	HAR/AE-G	HAR/AE/VF-G	HAR/AE-AG
3/12/2001	-4.53	1333	606	952	667	189	488
4/3/2001	-3.60	1538	690	645	426	256	198
9/17/2001	-4.97	> 10 ⁵	> 10 ⁵	> 10 ⁵	1538	606	833
7/10/2002	-3.46	46	27	69	50	31	48
7/19/2002	-4.22	56	19	100	35	22	34
7/22/2002	-3.39	28	8	47	11	10	13
8/5/2002	-3.56	37	18	59	32	21	22
9/3/2002	-4.28	156	168	235	323	129	190
3/24/2003	-3.58	204	714	351	952	230	426
2/27/2007	-3.38	> 10 ⁵	> 10 ⁵	> 10 ⁵	> 10 ⁵	4355	7540

the only one to pass this test, with all the rest being strongly rejected both at the point forecast and Monte Carlo methods.

4.5. **Excessive Conservativeness? An Application with Capital Charges.** The 1980s and 1990s were characterized by a series of financial disasters, many of which could be attributed, entirely or in part, to

TABLE 8. Density Forecasts

The table reports the p-values of Kolmogorov-Smirnov tests for the null hypothesis that the sequence of forecasted cumulative density functions evaluated at observed returns in the period between 2000 and 2007 (1557 trading days) are distributed as $U(0,1)$, where each model is re-estimated daily and used for one day ahead predictions. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is an explanatory variable for the volatility of volatility.

Model	PF Method	Monte Carlo
HAR	0.002	0.000
HAR/AE	0.001	0.000
HAR-GARCH(1,1)	0.021	0.007
HAR/AE-GARCH(1,1)	0.043	0.011
HAR/AE-GARCH(1,1)/RAV	0.001	0.001
ARFIMA	0.004	0.001
ARFIMA/AE	0.004	0.003
ARFIMA-GARCH(1,1)	0.003	0.002
ARFIMA/AE-GARCH(1,1)	0.006	0.005
HAR/AE/VF-GARCH	0.055	0.227
HAR/AE-Augmented Garch	0.001	0.000

poor risk management. The high levels of integration in modern financial markets do not permit a *laissez-faire* approach to the regulation of financial institutions, as systemic risk could lead to serious problems in the financial system. The groundbreaking Basel Capital Accord, originally signed by the Group of Ten (G10) countries in 1988, but since largely adopted by over 100 countries, requires Authorized Deposit-taking Institutions (ADIs) to hold sufficient capital to provide a cushion against unexpected losses, where the value-at-risk is the procedure designed to forecast the maximum expected loss over a target horizon. Initially, the Basel Accord stipulated a standardized approach which all institutions were required to adopt in calculating their VaR thresholds. This approach suffered from several deficiencies, the most notable of which were its conservatism (or lost opportunities) and its failure to reward institutions with superior risk management expertise.

Following much industry criticism, the Basel Accord was amended in April 1995 to allow institutions to use internal models to determine their VaR and the required capital charges. However, institutions wishing to use their own models are required to have the internal models evaluated by the regulators using the backtesting procedure. The Basel Accord penalizes institutions which use models with a greater number of violations than would be expected, given the statistical 1% level of confidence, this procedure is known as backtesting. Daily capital charges are given by

$$(16) \quad (3 + k) \text{Min} \left(\text{Var}_t, \sum_{i=t-60}^{t-1} \text{Var}_i / 60 \right)$$

where Var_t is the reported value-at-risk at day t and the penalties k applicable to violations of this quantity are given in Table 9.

TABLE 9. Basel Accord Penalty Zones.

Zone	Number Violations	k
Green	0 to 4	0.00
Yellow	5	0.40
	6	0.50
	7	0.65
	8	0.75
	9	0.85
Red	10+	1.00

Since the incorporation of the volatility of volatility invariably expands the forecasted value-at-risk, it is not immediately clear that this approach would be interesting in the case of a risk manager also interested in maximizing the amount of capital that could be freed for profitable investments. In fact, it could be the case that the better lower tail properties of our model are simply explained by forced conservativeness. To examine this possibility, we make a further illustration of the methods discussed previously by considering what would happen if we apply the volatility of volatility framework for the decision of choosing a VaR threshold optimally (in the sense of minimizing capital charges) for the S&P 500 index. Suppose that we start from a $(\gamma \times 100)\%$ VaR for day t (denoted Var_t^γ) and consider raising the threshold to $Var_t^{\gamma+\rho}$. The expected cost of doing this is determined by the increase in the probability of the returns falling below the VaR (ρ) and the higher future penalties:

$$(17) \quad \rho(\Delta k)E \left(\sum_{i=t+1}^{t+252} Var_i \right)$$

And benefit is the amount of capital that is freed at this particular date (in percentage points) and the expected benefit of a reduced average for the next 60 days:

$$(18) \quad (Var_t^\gamma - Var_t^{\gamma+\rho}) \left\{ 1 + (1/60)E \left[\sum_{i=t+1}^{60} I \left(Var_i < \sum_{j=i-60}^{i-1} Var_j/60 \right) \right] \right\}$$

Empirically, our optimization method consist in obtaining value-at-risks for different values of γ and applying the above comparison sequentially, where the size of the sequence depends on a choice of ρ . Of course, when $\rho \rightarrow 0$ the solution of the discrete problem converges to the solution of the optimization of capital charges that would arise if we had a continuous density. The expectations in (17) and (18) introduce a difficulty in that they require that we consider the same optimization problem in the future. To circumvent this problem in the case of the first equation we just make a forecast based on the time series of past VaRs, where the in-sample thresholds are set as the 1% VaRs, and exclude the out-of-sample first year as an initialization period. In the case of the second equation, we substitute the expectation for the historical average of the number of days where $\sum_{i=t-60}^{t-1} Var_i/60 < Var_t$.

The results for the point forecast and Monte Carlo methods are organized in Tables 10 and 11 respectively. In the first case, we can see that modeling refinements that were seem to be an improvement on previous analysis, considering asymmetric effects and time varying volatility of volatility, become counter-productive in most cases. Overall, the model that allows for correlation between the level and the volatility of volatility is the best over the point forecast method, even though all models are very similar. We remark

that the superior performance of the optimal method in this case rests on there being small punishments for excess violations. As might be expected, the Monte Carlo without optimization leads to an increase in capital charges. However, the optimal method unambiguously improves the result, providing evidence against the conjecture of excessive conservativeness.

TABLE 10. Capital Charges: Point Forecast Based Method.

The table reports the out-of-sample capital charges results for the S&P 500 daily realized volatility for the period between 2000 and 2007 (1557 trading days), where each model is re-estimated daily and used for one day ahead predictions. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is a explanatory variable for the volatility of volatility. The capital charges are calculated according to equation (16). The second column displays the average daily capital charges with the 99% VaR and the last column shows capital charges when the optimal method of equations (17) and (18) is implemented.

Model	Mean Daily CC	Optimal
HAR	6.426	6.417
HAR/AE	6.592	6.521
HAR-GARCH(1,1)	6.658	6.433
HAR/AE-GARCH(1,1)	6.607	6.443
HAR/AE-GARCH(1,1)/RAV	6.467	6.408
ARFIMA	6.830	6.390
ARFIMA/AE	6.652	6.472
ARFIMA-GARCH(1,1)	6.752	6.541
ARFIMA/AE-GARCH(1,1)	6.675	6.340
HAR/AE/VF-GARCH	6.763	6.522
HAR/AE-Augmented Garch	6.381	6.266

5. CONCLUSION

As exemplified by the long memory property case, the volatility literature grows in the middle of a theoretical gap. There currently exists no fully compelling theoretically parametric model of asset returns and the theoretical underpinnings of the patterns observed in volatility processes are poorly understood. In this paper, we embrace the view that the uncertainty surrounding tail events significantly affects economic decisions and focus on the objective of trying to measure this risk adequately by starting from the simplest paradigm (returns standardized by the realized volatility are normal), again without examining the underlying theoretical implications of time varying volatility of volatility. Despite the evidence about outcomes on the tails being by definition limited, large tails days are far more frequent than actual extreme events (the probability of a realization on the tail is low after all), so that by modeling the volatility of volatility with the methods discussed in the paper we are able to understand and manage this risk better.

We have argued that the existence of a large and time varying volatility of volatility is an important stylized times series fact of realized volatility. Far from exhausting the analysis of the empirical properties of this volatility of volatility, we have documented that this variable is subject to strong leverage effects

TABLE 11. Capital Charges: Monte Carlo Method.

The table reports the out-of-sample capital charges results for the S&P 500 daily realized volatility for the period between 2000 and 2007 (1557 trading days), where each model is re-estimated daily and used for one day ahead predictions. The specification for the conditional mean and conditional heteroskedasticity are separated by dashes. AE, VF and RAV mean that the model is being estimated with asymmetric effects, volatility feedback effects and the realized absolute variation respectively (see equation 3). In the augmented GARCH model the predicted value of the realized volatility is a explanatory variable for the volatility of volatility. The capital charges are calculated according to equation (16). The second column displays the average daily capital charges with the 99% VaR and the last column shows capital charges when the optimal method of equations (17) and (18) is implemented.

Model	Mean Daily CC	Optimal
HAR	6.745	6.107
HAR/AE	6.790	6.025
HAR-GARCH(1,1)	6.869	6.114
HAR/AE-GARCH(1,1)	7.004	6.150
HAR/AE-GARCH(1,1)/RAV	6.749	6.108
ARFIMA	6.833	6.039
ARFIMA/AE	7.050	6.078
ARFIMA-GARCH(1,1)	6.929	6.049
ARFIMA/AE-GARCH(1,1)	7.021	6.137
HAR/AE/VF-GARCH	7.623	6.371
HAR/AE-Augmented Garch	6.763	6.003

and short-lived explosive regimes and is highly positively related to the conditional mean of volatility. We have also shown that the size of the day to day unpredictability in volatility is such that it swamps typical differences on average in out of sample forecasting performances, adding to the skepticism over the practice of evaluating and comparing volatility models by means of their explanatory power (Fleming et al., 2001, 2003). We have shown that when high-frequency data is available significant modeling gains are obtained when the realized volatility framework is extended with the concept of volatility of volatility so as to model both the conditional volatility and the kurtosis of returns. Our analysis also indicates that accounting for asymmetric effects (both those related to the recent history of asset prices and volatility feedback effects) are extremely important for the accuracy of the general model.

Our analysis has also important potential implications for the asset pricing literature. First, at the same time our analysis may relate to the interest in the volatility premium (see for example Bakshi and Kapadia, 2003, who finds a negative volatility risk premium using a sample of S&P 500 index options), the strong positive relation between the conditional mean and volatility of volatility suggests that additional work may be needed to disentangle it from the risk premium. Second, our findings raise several questions about the implications of the observed properties of the volatility of volatility (time variation, long memory, leverage effects, correlation with the level of volatility) and the non-gaussianity of the shocks to realized volatility for options pricing, all of which are steps moving away from the Heston (1993) workhorse model of stochastic volatility. Finally, the pervasiveness of leverage effects, which significantly impacts the volatility, skewness and kurtosis of the return distribution of S&P 500 returns and are inextricably linked to virtually

all episodes of high volatility and tail events on our sample, may call for more theoretical investigation on the mechanisms behind asymmetric effects.

REFERENCES

- ANDERSEN, T. AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–906.
- ANDERSEN, T., T. BOLLERSLEV, AND F. DIEBOLD (2007): "Roughing it up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–720.
- 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., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.
- AREAL, N. AND S. R. TAYLOR (2002): "The Realized Volatility of FTSE-100 Futures Prices," *Journal of Futures Markets*, 22, 627–648.
- ARTZNER, P., F. DELBAEN, J. EBER, AND D. HEATH (1997): "Thinking Coherently," *RISK*, 10, 68–71.
- (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9, 203–228.
- AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2005): "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," NBER Working Papers 11380, National Bureau of Economic Research.
- BAILLIE, R., T. BOLLERSLEV, AND H. O. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74, 3–30.
- BAILLIE, R. T. (1996): "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.
- BAKSHI, G. AND N. KAPADIA (2003): "Delta-Hedged Gains and the Negative Market Volatility Risk Premium," *Review of Financial Studies*, 16, 527–566.
- BANDI, F. AND J. RUSSELL (2005): "Microstructure noise, realized volatility and optimal sampling," unpublished paper, Graduate School of Business, University of Chicago.
- BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2005): "Realised kernels can consistently estimate integrated variance: correcting realised variance for the effect of market frictions," Discussion paper, Nuffield College, Oxford.
- (2007a): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," Unpublished manuscript, Stanford University.
- (2007b): "Subsampling Realized Kernels," Working paper, Stanford University.
- BARNDORFF-NIELSEN, O. AND N. SHEPHARD (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society B*, 64, 253–280.
- (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.
- (2006): "Econometrics of Testing for Jumps in Financial Economics using Bipower Variation," *Journal of Financial Econometrics*, 4, 1–30.
- BELTRATTI, A. AND C. MORANA (2005): "Statistical Benefits of Value-At-Risk with Long Memory," *Journal of Risk*, 7.

- (2006): “Breaks and Persistency: Macroeconomic Causes of Stock Market Volatility,” *Journal of Econometrics*, 131, 151–177.
- BIAIS, B., L. GLOSTEN, AND C. SPATT (2005): “Market microstructure: a survey of microfoundations, empirical results, and policy implications,” *Journal of Financial Markets*, 8, 217–264.
- BOLLERSLEV, T., J. LITVINOVA, AND G. TAUCHEN (2006): “Leverage and Volatility Feedback Effects in High-Frequency Data,” *Journal of Financial Econometrics*, 4, 353–384.
- BREIDT, F., N. CRATO, AND P. 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.
- CHRISTOFFERSEN, P. (1998): “Evaluating Interval Forecasts,” *International Economic Review*, 39, 841–62.
- COMTE, F. AND E. RENAULT (1996): “Long memory continuous time models,” *Journal of Econometrics*, 73, 101–149.
- CORSI, F. (2004): “A Simple Long Memory Model of Realized Volatility,” Manuscript, University of Southern Switzerland.
- CORSI, F., U. KRETSCHMER, S. MITTNIK, AND C. PIGORSCH (2008): “The Volatility of Realized Volatility,” *Econometric Reviews*, 27, 46–78.
- DEO, R., C. HURVICH, AND Y. LU (2006): “Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility,” *Journal of Econometrics*, 131, 29–58.
- DIEBOLD, F. AND A. INOUE (2001): “Long Memory and Regime Switching,” *Journal of Econometrics*, 105, 131–159.
- DIEBOLD, F. X., T. A. GUNTHER, AND A. S. TAY (1998): “Evaluating density forecasts with applications to financial risk management,” *International Economic Review*, 39, 863–883.
- FLEMING, J., C. KIRBY, AND B. OSTDIEK (2001): “The Economic Value of Volatility Timing,” *Journal of Finance*, 56, 329–352.
- (2003): “The Economic Value of Volatility Using Realized Volatility,” *Journal of Financial Economics*, 67, 473–509.
- GATHERAL, J. AND R. OOMEN (2007): “Zero Intelligence Variance Estimation,” Working paper, Warwick Business School.
- GEWEKE, J. AND S. PORTER-HUDAK (1983): “The estimation and application of long memory time series models,” *Journal of Time Series Analysis*, 4, 221–238.
- GHYSELS, E., A. SINKO, AND R. VALKANOV (2007): “MIDAS Regressions: Further Results and New Directions,” *Econometric Reviews*, 26, 53–90.
- GRANGER, C. AND Z. DING (1996): “Varieties of Long Memory Models,” *Journal of Econometrics*, 73, 61–77.
- GRANGER, C. W. J. AND N. HYUNG (2004): “Occasional Structural Breaks and Long Memory with an Application to the SP 500 Absolute Stock Returns,” *Journal of Empirical Finance*, 11, 399–421.
- HANSEN, P. (2005): “A Test for Superior Predictive Ability,” *Journal of Business and Economic Statistics*, 23, 365–380.
- HANSEN, P., J. LARGE, AND A. LUNDE (2008): “Moving average-based estimators of integrated variance,” *Econometric Reviews*, 27, 79–111.

- HANSEN, P. R. AND A. LUNDE (2005): "A Forecast Comparison of Volatility Models: Does Anything beat a GARCH(1,1) Model?" *Journal of Applied Econometrics*, 20, 873–889.
- (2006): "Realized variance and market microstructure noise (with discussion)," *Journal of Business and Economic Statistics*, 24, 127–218.
- HESTON, S. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.
- HILLEBRAND, E. (2005): "Neglecting Parameter Changes in GARCH Models," *Journal of Econometrics*, 129, 121–138.
- HILLEBRAND, E. AND M. MEDEIROS (2007): "Forecasting Realized Volatility Models: The Benefits of Bagging and Nonlinear Specifications," Discussion paper, Pontifical Catholic University of Rio de Janeiro.
- HURVICH, C. AND R. DEO (1999): "Plug-in Selection of the Number of Frequencies in Regression Estimates of the Memory Parameter of a Long-Memory Time Series," *Journal of Time Series Analysis*, 20, 331–341.
- HYUNG, N. AND P. FRANSES (2002): "Inflation rates: long-memory, level shifts, or both?" Report 2002-08, Econometric Institute, Erasmus University.
- KOOPMAN, S., B. JUNGBACKER, AND E. HOL (2005): "Forecasting Daily Variability of the SP 100 Stock Index Using Historical, Realised and Implied Volatility Measurements," *Journal of Empirical Finance*, 23, 445–475.
- LIU, C. AND J. MAHEU (2007): "Forecasting Realized Volatility: A Bayesian Model Averaging Approach," Working paper.
- LOBATO, I. N. AND N. E. SAVIN (1998): "Real and Spurious Long-Memory Properties of Stock-Market Data," *Journal of Business and Economic Statistics*, 16, 261–268.
- 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.
- MARTENS, M., D. VAN DIJK, AND M. DE POOTER (2004): "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," Discussion Paper 04-067/4, Tinbergen Institute.
- MCALÉER, M. AND M. MEDEIROS (2008a): "A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries." *Journal of Econometrics*, forthcoming.
- (2008b): "Realized Volatility: A Review," *Econometric Reviews*, 27, 10–45.
- MIKOSCH, T. AND C. STARICA (2004): "Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects," *Review of Economics and Statistics*, 86, 378–390.
- MÜLLER, U., M. DACOROGNA, R. DAV, R. OLSEN, O. PICTET, AND J. VON WEIZSACKER (1997): "Volatilities of Different Time Resolutions – Analysing the Dynamics of Market Components," *Journal of Empirical Finance*, 4, 213–239.
- MORANA, C. AND A. BELTRATTI (2004): "Structural Change and Long Range Dependence in Volatility of Exchange Rates: Either, Neither or Both?" *Journal of Empirical Finance*, 11, 629–658.
- OHANISSIAN, A., J. RUSSELL, AND R. TSAY (2004): "True or Spurious Long Memory in Volatility: Does it Matter for Pricing Options?" Working paper, Graduate School of Business, University of Chicago.
- SCHARTH, M. AND M. MEDEIROS (2006): "Asymmetric Effects and Long Memory in the Volatility of Dow Jones Stocks," Working paper, Department of Economics, Pontifical Catholic University of Rio de

Janeiro.

- TAUCHEN, G. AND H. ZHOU (2005): "Identifying Realized Jumps on Financial Markets," Working paper, Duke University, Department of Economics.
- THOMAKOS, D. AND T. WANG (2003): "Realized Volatility in the Futures Market," *Journal of Empirical Finance*, 10, 321–353.
- W. CHAN, R. J. AND M. KALIMIPALLI (2006): "The economic value of trading with realized volatility in the SP 500 index options market," Working paper.
- ZHANG, L., P. 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.
- ZHOU, B. (1996): "High frequency data and volatility in foreign exchange rates," *Journal of Business and Economic Statistics*, 14, 45–52.