Edith Cowan University

Research Online

2009

Pricing options by simulation using realized volatility

David E. Allen
Edith Cowan University

Michael McAleer
Complutense University of Madrid

Marcel Scharth
University of Amsterdam

PRICING OPTIONS BY SIMULATION USING REALIZED VOLATILITY

David E. Allen

School of Accounting, Finance and Economics
Edith Cowan University

Michael McAleer

Department of Quantitative Economics
Complutense University of Madrid

Marcel Scharth

University of Amsterdam

ABSTRACT. A growing literature advocates the use of high-frequency data for the purpose of volatility estimation. However, despite the successes in modeling the conditional mean of realized volatility empirical evaluations of this class of models outside the realm of short run forecasting is limited. How can realized volatility be used for pricing options? What are the modeling qualities introduced by realized volatility models for pricing derivatives? In this short paper, we propose an options pricing framework based on a new realized volatility model that captures all the relevant empirical regularities of the realized volatility series of the S&P 500 index. We emphasize two main empirical regularities for our volatility model and that are potentially very relevant for option pricing purposes.

Fist, realized variation measures constructed from high-frequency returns reveal a large degree of time series unpredictability in the volatility of asset returns. Even though returns standardized by (ex-post) quadratic variation measures are nearly gaussian, this unpredictability brings substantially more uncertainty to the empirically relevant (ex-ante) distribution of returns. In this setting carefully modeling the stochastic structure of the time series disturbances of realized volatility is fundamental. Second, there is evidence of very large leverage effects; large falls (rises) in prices being associated with persistent regimes of high (low) variance in the index returns.

We propose a model for the conditional volatility, skewness and kurtosis of daily index and stocks returns. The main new feature of this model is to recognize that volatility is itself more volatile and more persistent in high volatility periods. Contrary to "peso problem" considerations, we show that when volatility is (nearly) observable it is not necessary to rely on rare realizations on past return data to learn about the tails of the return distribution, an unexplored and large modeling gain enabled by high frequency data.

We conduct a brief empirical illustration analysis of the pricing performance of this approach against some benchmark models using data from the S&P 500 options in the 2001-2004 period. The results indicate that as expected the superior forecasting accuracy of realized volatility translates into significantly smaller pricing errors when compared to models of the GARCH family. More significantly, our results indicate that modeling leverage effects and the volatility of volatility are paramount reducing common pricing anomalies.

KEYWORDS: Realized volatility, option pricing, volatility of volatility, forecasting.

1. INTRODUCTION

The advent of high frequency stock market data and the subsequent introduction of realized volatility measures represented a substantial step forward in the accuracy with which econometric models of volatility could be evaluated 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) or stochastic volatility family of models (Andersen et al., 2003) on forecasting future volatility, to the point that the comparison has been dropped altogether in recent papers.

Contributions to the realized volatility modeling and forecasting literature are exemplified by Andersen et al. (2003) and the HAR (heterogeneous autoregressive) model of Corsi (2004). Martens et al. (2004) 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 otheres argue that the inclusion of jump components significantly improves forecasting performance. Scharth and Medeiros (2009) introduce multiple regime models linked to asymmetric effects. Bollerslev et al. (2009) propose a full system for returns, jumps and continuous time for components of price movements using realized variation measures.

Despite these successes in modeling the conditional mean of realized volatility, empirical evaluations of this class of models outside the realm of short run forecasting is limited. Fleming et al. (2003) examines the economic value of volatility timing using realized volatility. Bandi et al. (2008) evaluates and compares the quality of several recently-proposed realized volatility estimators in the context of option pricing and trading of short term options on a stylized setting. Stentoft (2008b) derive an appropriate return and volatility dynamics to be used for option pricing purposes in the context of realized volatility and perform an empirical analysis using stock options for three large American companies.

In this short paper, we propose an options pricing framework based on a new realized volatility model that captures all the relevant empirical regularities of the realized volatility series of the S&P 500 index and conduct a brief empirical analysis of the pricing performance of this approach against some benchmark models using data from the S&P 500 options in the 2001-2004 period. The results indicate that as expected the superior forecasting accuracy of the proposed realized volatility model translates into significantly smaller pricing errors when compared to models of the GARCH family. More significantly, our results indicate that modeling leverage effects and the volatility of volatility are paramount reducing common pricing anomalies.

2. DATA AND STYLIZED FACTS OF REALIZED VOLATILITY

2.1. **Realized volatility and Data.** Suppose that at day t the logarithmic prices of a given asset follow a continuous time diffusion:

$$dp(t+\tau) = \mu(t+\tau) + \sigma(t+\tau)dW(t+\tau), \ 0 \leq \tau \leq 1, \ t = 1, 2, 3...$$

where $p(t+\tau)$ is the logarithmic price at time $t+\tau$, is the drift component, $\sigma(t+\tau)$ is the instantaneous volatility (or standard deviation), and $dW(t+\tau)$ is a standard Brownian motion. Andersen et al. (2003) (and others) showed that the daily compound returns, defined as $r_t = p(t) - p(t+1)$, are Gaussian conditionally on $\mathcal{F}_t = \sigma(p(s), s \leq t)$, the σ-algebra (information set) generated by the sample paths of p, such that

$$r_t | \mathcal{F}_t \sim N \left(\int_0^1 \mu(t-1+\tau)d\tau, \int_0^1 \sigma^2(t-1+\tau)d\tau \right)$$

The term $IV_t = \int_0^1 \sigma^2(t - 1 + \tau)d\tau$ is known as the integrated variance, which is a measure of the day t *ex post* volatility. In this sense, the integrated variance is the object of interest. In practical applications prices are observed at discrete intervals. If we set $p_{i,t}$, $i = 1, .., n$ to be the ith price observation during day t, realized variance is defined as $\sum_{i=1}^n r_{i,t}^2$. The realized volatility is the square-root of the realized variance and we shall denote it by RV_t. Ignoring the remaining measurement error, this *ex post* volatility measure can modeled as an "observable" variable, in contrast to the latent variable models.

In real data, however, high frequency measures are contaminated by microstructure noise such as bid-ask bounce, asynchronous trading, infrequent trading, price discreteness, among others. In this paper, we turn to the theory developed by Barndorff-Nielsen et al. (2008) and implement a realized kernel estimator based on one minute returns and the modified Tukey-Hanning kernel, which is consistent in the presence of microstructure noise.

The empirical analysis focuses on the realized volatility of the S&P 500 index (SPX) and the S&P 500 options traded the Chicago Board Options Exchange (CBOE). The raw intraday data was obtained from the Taqtiq/SIRCA (Securities Industry Research Centre of Asia-Pacific) database. To calculate the realized volatility series we use tick-by-tick open to close quotes originated in the E-Mini S&P500 futures market of the Chicago Mercantile Exchange.[1] The period of analysis for the realized volatility starts in January 2, 1996, and ends in November 28, 2008. For the options, we record daily intraday option prices close to the $14 : 30$ time mark between January 2, 2001, and ends in March 15, 2004. The data for the risk-free rate is obtained from the three month T-Bill.

3. MODELING REALIZED VOLATILITY AND PRICING EUROPEAN OPTIONS

3.1. **The HAR/AE-VL Model.** Our model for returns, realized volatility and the volatility of realized volatility is given by:ecalling that we shall be working with the realized volatility (RV_t), consider the notation $RV_{j,t} = \sum_{i=t-j+1}^t RV_{t,j}/j$. We can then write our HAR model with daily, weekly and monthly components as:

$$
\begin{aligned}
r_t &= \mu_t + RV_t \varepsilon_t, \\
RV_t &= \phi_0 + \phi_1 RV_{t-1} + \phi_2 RV_{5,t-1} + \phi_3 RV_{22,t-1} \\
&\quad + \lambda_1 I(r_{t-1} < 0)r_{t-1} + \lambda_2 I(r_{5,t-1} < 0)r_{5,t-1} \\
&\quad + \lambda_3 I(r_{22,t-1} < 0)r_{22,t-1} + h_t \nu_t, \\
h_t^2 &= \theta_0 + \theta_1 E(RV_t | \mathcal{F}_{t-1})^2 + \theta_2 \nu_{t-1}^2,
\end{aligned}
$$
(1)

where r_t is the log return at day t, μ_t is the conditional mean for the returns, RV_t is the realized volatility, ε_t is $N.i.d.(0, 1)$, ψ_t shifts the unconditional mean of realized volatility, d denotes the fractional differencing parameter, L the lag operator, I is the indicator function, $r_{j,t-1}$ is a notation for the cumulated returns $\sum_{i=t-j}^{t-1} r_{t-i}$, h_t is the volatility of the realized volatility, ν_t is $i.i.d.$ and distributed normal inverse gaussian with $E(\nu_t) = 0$ and $E(\nu_t^2) = 1$, ε_t and ν_t are allowed to be dependent, $\mathcal{F}_{t-1}$ is the information set of the end of day $t - 1$ and $E(RV_t | \mathcal{F}_{t-1})$ is given by the model for the conditional mean of realized volatility. We discuss below each part of the general model separately.

3.1.1. *HAR specification.* The HAR (Heterogeneous Autoregressive) model proposed by Corsi (2004) is an unfolding of the Heterogeneous ARCH (HARCH) model developed by earlier in. It is specified as a multi-component volatility model with an additive hierarchical structure, leading to an additive time

---

[1]These fully electronic contracts feature among the most liquid derivatives contracts in the world, therefore closely tracking price movements of the S&P 500 index.

series model of the realized volatility which specifies the volatility as a sum of volatility components over different horizons.

We can see that the HAR specification is an $AR(22)$ model rendered parsimonious by several parameter restrictions. Simulations reported in Corsi (2004) that the HAR model is 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. For its estimation simplicity, the HAR-RV has been commonly favored in the high frequency econometrics literature (e.g, Andersen et al., 2007).

3.1.2. *Asymmetric Effects.*

Scharth and Medeiros (2009) (among others) 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. We also consider the relation between returns and volatility on a same day (which comes from the dependence between ν_t and ε, discussed soon).

3.1.3. *Volatility of Volatility.*

Allen et al. (2009) show that the time series of the volatility of the realized volatility of the S&P500 index displays evidence of long memory, leverage effects and high correlation with the level of volatility. To account for all these aspects simultaneously and parsimoniously, we take the "level of volatility" to be the conditional mean of volatily and set the variance of the errors to be a function of this variable.

3.1.4. *The Distribution of ν_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). The density of the NIG distribution is given by:

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

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. μ and δ are always set so that the distribution has mean 0 and variance 1.

3.1.5. *The Dependence between ν_t and ε_t.*

To account for the asymmetry in the ex-ante return distribution, we let ν_t and ε_t be dependent and model this dependence via a bivariate Clayton copula. Let $U = \Phi(\varepsilon_t)$ and $V = 1 - \Upsilon(\nu_t)$, where $\Phi(.)$ and $\Upsilon(.)$ are the corresponding normal and NIG cdfs for ε_t and ν_t respectively. The joint CDF or copula of U and V is given by:

$$(3) \qquad C_\kappa \equiv P(U \le u, V \le v) = \left(u^{-\kappa} + v^{-\kappa} - 1\right)^{-1/\kappa}$$

3.2. **Estimation.** To estimate the model we maximize the following log-likelihood function:

$$\ell(\hat{d}, \hat{\phi}, \hat{\psi}, \hat{\lambda}, \hat{\theta}, \hat{\alpha}, \hat{\beta}; RV_{1...T}, X_{1...T}) = T\log(\hat{\alpha}) - T\log(\pi) + \sum_{t=1}^{T} \log\left[K_1(\hat{\alpha}\hat{\delta}(1 + \hat{y}_t^2)^{1/2})\right]$$

$$(4) \qquad\qquad -0.5\sum_{t=1}^{T}\log(1 + \hat{y}_t^2) + T\hat{\delta}(\hat{\alpha}^2 - \hat{\beta}^2)^{1/2} + \hat{\delta}\hat{\beta}\sum_{t=1}^{T}\hat{y}_t$$

$$-0.5\sum_{t=1}^{T}\log(\hat{h}_t)$$

where X collects the additional explanatory variables, $\gamma = (\hat{\alpha}^2 - \hat{\beta}^2)^{1/2}$ and $\hat{y}_t = \frac{\hat{\nu}_t/\hat{h}_t - \hat{\mu}}{\hat{\delta}}$

3.3. **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. 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. Conditional on information up to day $t - 1$, the forecasted empirical density function for day t can be calculated as follows:

(1) The functional form of the model is used for the evaluation of predictions of the realized volatility and the volatility of volatility conditional on past realized volatility observations, returns, the estimated volatility of volatility series and shocks, and other variables. We randomly generate n shocks distributed as the standardized NIG with the parameters estimated from the data as described in section, which multiplied by $\tilde{h}_t$ and added to $\widetilde{RV}_t$ originate a vector of n simulated realized volatilities for day t.

(2) We compute the CDF of the simulated volatility shocks. Using the estimated Clayton copula, we generate n standardized shocks for the returns conditional on the volatility shocks by the inverse CDF method.

(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.

3.4. **Pricing European Options with Realized Volatility.** To price european options on the S&P 500 index using the framework discussed above, we simulate returns and volatility under the risk neutral distribution. As it is well known, the existence of a risk neutral dynamics follows from absence of arbitrage and mild regularity conditions. To keep the pricing framework tractable, we follow the approach of Stentoft (2008b) and assume that investors require no premium for being exposed to realized volatility risk. In this case the risk neutral RV dynamics are the same as the physical dynamics. The risk neutral system is:

$$
\begin{aligned}
(5) \quad & r_t = \mu_t^* + RV_t \varepsilon_t^*, \\
& RV_t = \phi_0 + \phi_1 RV_{t-1} + \phi_2 RV_{5,t-1} + \phi_3 RV_{22,t-1} \\
& \qquad + \lambda_1 I(r_{t-1} < 0) r_{t-1} + \lambda_2 I(r_{5,t-1} < 0) r_{5,t-1} \\
& \qquad + \lambda_3 I(r_{22,t-1} < 0) r_{22,t-1} + h_t \nu_t, \\
& h_t^2 = \theta_0 + \theta_1 E(RV_t | \mathcal{F}_{t-1})^2 + \theta_2 \nu_{t-1}^2,
\end{aligned}
$$

where ε_t^* is distributed $N(0,1)$. μ_t^* is such that $E^Q(\exp[\mu_t + RV_t \varepsilon_t^*]) = \exp rf_t$, where rf_t is the daily risk-free rate (assumed constant during the life of the option) and $E^Q(.)$ is the expectation under the risk neutral measure.

4. EMPIRICAL ILLUSTRATION

In this section we perform a brief empirical analysis of our option pricing model. For conciseness we focus on put options with between 9 and 60 calendar days to expiration. Defining moneyness by $M = S_t/X$, where S_t is the underlying index price at the time when the option is observed and X is the strike price, we divide the options into the following groups: at-the-money ($0.98 < M < 1.02$), out-the money ($1.02 < M < 1.05$), in-the-money ($0.95 < M < 0.98$), deep out-of-the-money ($1.05 < M < 1.1$) and the deep in-the-money ($0.9 < M < 0.95$). We consider some alternative models. (i) the Black and Scholes price (where the volatility is given by the mean of realized volatility over the last month) (ii) GARCH/NIG prices (for the GARCH, EGARCH, GJR-GARCH and NGARCH models). See (for example) Stentoft (2008a) for the theoretical background and details. (iii) two variations of our realized model: in both cases the specification is exactly as in the main model, except that we consider a GARCH(1,1) model for the volatility of realized volatility (h_t in the notation of the last sections) and first alternative does not allow for lagged leverage effects.

The results are summarized in Table 1 below, where we focus on the mean absolute pricing error metric. As expected, both GARCH and RV prices are substantial improvements over the Black and Scholes prices. Among the GARCH models, the EGARCH stands out as having a much superior performance than the others, followed by the GJR, NGARCH and GARCH specifications. Perhaps surprisingly, the RV-HAR-GARCH models are inferior to the EGARCH model. Nevertheless, the HAR/AE-VL model presented in this paper substantially reduces pricing errors at all moneyness categories when compared to the best GARCH model. Table 2 illuminates the reason for better performance of the model: contrary to all the other specifications it does not underprice in general these put options (and in particular the out-of-the-money ones, a well known and deficiency of the BS model). The reason why the model achieves the results is the conjunction of leverage effects and the specification of the volatility and volatility: increases in the volatility may breed even more volatility through a higher volatility of volatility– the model generates much fatter tails than the alternatives.

REFERENCES

ALLEN, D., M. MCALEER, AND M. SCHARTH (2009): "How Predictable is Volatility?" Working paper.

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. DIEBOLD, AND P. LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

TABLE 1. MEAN ABSOLUTE PRICING ERRORS.

	ATM	OTM	ITM	DOTM	DITM
Black and Scholes	4.26	4.89	3.13	5.11	2.33
GARCH/NIG	3.69	4.07	2.86	4.16	2.04
EGARCH/NIG	2.68	2.59	2.31	2.36	1.89
GJR-GARCH/NIG	3.23	3.09	2.77	2.65	2.04
NGARCH/NIG	3.61	3.75	3.06	3.92	2.36
RV-HAR-GARCH	3.31	4.13	2.30	4.92	1.60
RV-HAR/AE-GARCH	2.90	3.24	2.32	3.58	1.78
RV-HAR/AE-VL	1.87	1.77	1.61	1.71	1.31

TABLE 2. MEAN PRICING ERRORS.

	ATM	OTM	ITM	DOTM	DITM
Black and Scholes	-2.27	-3.63	-0.45	-4.13	0.42
GARCH/NIG	-1.40	-2.39	0.01	-2.80	0.51
EGARCH/NIG	-1.80	-1.78	-1.62	-1.74	-1.67
GJR-GARCH/NIG	-2.06	-2.01	-1.58	-1.64	-1.41
NGARCH/NIG	-3.08	-3.35	-2.45	-3.61	-1.96
RV-HAR-GARCH	-2.69	-3.81	-1.59	-4.74	-1.00
RV-HAR/AE-GARCH	-2.29	-2.85	-1.75	-3.33	-1.35
RV-HAR/AE-VL	0.25	0.23	0.04	0.30	-0.28

BANDI, F., J. RUSSELL, AND C. YANG (2008): "Realized volatility forecasting and option pricing," *Journal of Econometrics*, 134–146.

BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

BOLLERSLEV, T., U. KRETSCHMER, C. PIGORSCH, AND G. TAUCHEN (2009): "A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects," *Journal of Econometrics*, forthcoming.

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.

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2003): "The Economic Value of Volatility Timing Using Realized Volatility," *Journal of Financial Economics*, 67, 473–509.

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.

SCHARTH, M. AND M. MEDEIROS (2009): "Asymmetric Effects and Long Memory in the Volatility of Dow Jones Stocks," *International Journal of Forecasting*, forthcoming.

STENTOFT, L. (2008a): "American Option Pricing using GARCH models and the Normal Inverse Gaussian distribution," Working paper.

——— (2008b): "Option Pricing using Realized Volatility," Working paper.