Queen's Economics Department Working Paper No. 1186

The Implied-Realized Volatility Relation with Jumps in Underlying Asset Prices

Bent Jesper Christensen
University of Aarhus and CREATES

Morten Ørregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

11-2005

The Implied-Realized Volatility Relation with Jumps in Underlying Asset Prices*

Bent Jesper Christensen
University of Aarhus

Morten Ørregaard Nielsen†
Cornell University

November 2, 2005

Abstract

Recent developments allow a nonparametric separation of the continuous sample path component and the jump component of realized volatility. The jump component has very different time series properties than the continuous component, and accounting for this allows improved forecasting of future realized volatility. We investigate the potential forecasting role of implied volatility backed out from option prices in the presence of these new separate realized volatility components. We show that implied volatility has incremental information relative to both the continuous and jump components of realized volatility when forecasting subsequently realized return volatility, and it appears to be an unbiased forecast. Furthermore, implied volatility has predictive power for future values of each component of realized volatility separately, showing in particular that even the jump component of realized volatility is, to some extent, predictable.

JEL Classifications: C1, C32, G1.

Keywords: Bipower variation, implied volatility, instrumental variables, jumps, options, realized volatility, stock prices, vector autoregressive model, volatility forecasting.

*We are grateful to Per Frederiksen, Neil Shephard, and participants at the Nordic Econometric Meeting 2005 in Helsinki, and the Econometric Society World Congress 2005 in London for comments, to Tim Bollerslev for supplying the realized volatility, bipower variation, and tripower quarticity data used in this study, and to Rolf Karlsen and Christian Sønderup for research assistance. This research was supported by the Center for Analytical Economics (CAE) at Cornell, the Center for Analytical Finance (CAF) at Aarhus, and the Danish Social Science Research Council (SSF/FSE).

†Corresponding author. Please address correspondence to: Morten Ørregaard Nielsen, Department of Economics, Cornell University, 482 Uris hall, Ithaca, NY 14853, USA; phone: +1 607 255 6338; fax: +1 607 255 2818; email: mon2@cornell.edu

1 Introduction

The analysis of financial market volatility is of the utmost importance to asset pricing, derivative pricing, hedging, and risk management. Several different sources of information may be invoked in generating forecasts of unknown future volatility. Besides measurements based on historical return records, observed derivative prices are known from finance theory to be highly sensitive to and hence informative about future volatility. It is therefore natural to consider data on both asset prices and associated derivatives when measuring, modelling and forecasting volatility. Earlier literature has shown that implied volatility backed out from option prices provides a better volatility forecast than sample volatility based on past daily returns, but more recent literature shows that volatility forecasting based on past returns may be improved dramatically by using high-frequency (e.g., 5-minute) returns, and explicitly allowing for jumps in asset prices when computing forecasts. The important question addressed in the present paper is whether implied volatility from option prices continues to be the dominant volatility forecast, even when comparing to these new improved return based alternatives, using high-frequency data and accommodating a jump component in asset prices.

The recent realized volatility literature has focussed much on the summation of high-frequency squared returns as a robust way of consistently estimating conditional volatility, see e.g. French, Schwert & Stambaugh (1987), Schwert (1989), Andersen & Bollerslev (1998*a*), Andersen, Bollerslev, Diebold & Ebens (2001), Andersen, Bollerslev, Diebold & Labys (2001), and Barndorff-Nielsen & Shephard (2002*a*), and for a recent survey, see Andersen, Bollerslev & Diebold (2004). In particular, Andersen, Bollerslev, Diebold & Labys (2003) and Andersen, Bollerslev & Meddahi (2004) show that simple reduced form time series models for realized volatility thus constructed from historical returns outperform the commonly used GARCH and related stochastic volatility models in forecasting future volatility. This is good news for volatility forecasters, academic as well as practitioners, since realized volatility is relatively easy to compute, given the increasing availability of high-frequency historical return records.

Two main drawbacks to the approach of summing high-frequency squared returns are that, firstly, the reliance on past returns only completely ignores the presumably considerable information content on volatility available through observable prices of traded derivative securities. Implied volatility from option prices is commonly among practitioners expected to be a much more precise forecast of future volatility than anything based on past returns, as current option prices avoid obsolete information and are assumed to incorporate all relevant information efficiently. Secondly, recent studies have stressed the importance of explicitly allowing separate treatment of the jump and continuous sample path components, both in estimating parametric stochastic volatility models (e.g. Andersen, Benzoni & Lund (2002), Chernov, Gallant, Ghysels & Tauchen (2003), Eraker, Johannes & Polson (2003), and Ait-Sahalia (2004)), in nonparametric realized volatility modelling (e.g. Barndorff-Nielsen & Shephard (2003*a*, 2004*b*), Huang & Tauchen (2005), and Andersen, Bollerslev & Diebold (2005)), and in empirical option pricing (e.g. Bates (1991) and Bakshi, Cao & Chen (1997)). In particular, in the stochastic volatility and realized volatility literatures, the jump component is found to be far less predictable than the continuous sample path component, clearly indicating separate roles for these in a forecasting context.

Complete reliance on return data only may not provide an efficient volatility forecast, given investors' information set. If option market participants are rational and markets are efficient, then implied volatility backed out from traded option prices should reflect available information about future volatility through expiration of the options, including that contained in past returns. Ignoring option price data in forecasting volatility therefore does not seem natural. In fact, Christensen & Prabhala (1998) consider more than a decade of return and option price data for the S&P 100 index and find that implied index option (OEX) volatility is an unbiased and efficient forecast of future realized volatility, subsuming the information content of past realized volatility as a forecast. Other studies documenting the incremental information in implied volatility relative to past realized volatility include Day & Lewis (1992), Lamoureux & Lastrapes (1993), Jorion (1995), Fleming (1998), and Blair, Poon & Taylor (2001). It is therefore clearly of interest to examine the role of implied volatility from option prices in the context of the most recent realized volatility modelling and forecasting literature.

The consistency of realized volatility as an estimate of true total volatility as the frequency of return observations is increased extends to the case of asset price processes including both stochastic volatility and jump components. However, for forecasting purposes, the behavior of and information content in the continuous sample path and jump components of total volatility may be very different (Andersen et al. (2005)). Recent theoretical developments by Barndorff-Nielsen & Shephard (2003*a*, 2003*b*, 2004*a*, 2004*b*) allow a fully nonparametric separation of the continuous sample path and jump components of realized volatility. This separation into two series is exactly the tool required to model and analyze the underlying continuous and jump components individually, and lends itself to empirical application. Using this methodology, Andersen et al. (2005) extend results of Andersen et al. (2003) and Andersen, Bollerslev & Meddahi (2004) by including both the continuous and jump components of past realized volatility as separate regressors in the forecasting of future realized volatility. Andersen et al.'s (2005) results show that significant gains in forecasting performance may be achieved by splitting the explanatory variables into the separate continuous and jump components, compared to using only total past realized volatility. While the continuous component of past realized volatility is strongly serially correlated, the jump component is found to be distinctly less persistent, and hence the two components play very different roles in forecasting. The explanatory power, judged in terms of adjusted R^2, improves even more when forecasting volatility over a full month than for shorter horizon forecasts. Thus, for one month forecasts of S&P 500 return volatility, R^2 more than doubles when replacing explanatory variables based on daily or lower frequency returns with their high-frequency counterparts and accounting for jumps, and R^2 triples resp. quadruples when considering variances resp. log-volatilities[1] instead of raw volatilities (standard deviations). Clearly, the enhanced forecasting performance begs the question of whether implied volatility from option prices continues to come in as the dominant forecast, even in the presence of past realized volatility measurements based on high-frequency return data and differentiating the continuous and jump components.

In this paper, we include implied volatility from option prices in the analysis, thus expanding the set of variables from the information set used for forecasting purposes. Given that Andersen et al. (2005) show that splitting past realized volatility into its separate components yields an improved forecast, adding implied volatility allows examining whether the continuous and jump components of past realized volatility span the relevant part of the information set. Similarly, as previous literature shows that implied volatility outperforms past realized volatility as a forecast, it is of interest to test whether this conclusion holds up after allowing the two components of past realized volatility to act separately. In addition, the earlier literature on the relation between implied and realized volatility has considered realized volatility constructed from daily return observations, due to data limitations, and this could be one reason for imprecise measurement of realized volatility and might have biased the results on forecasting performance in favor of implied volatility from option prices. In sum, by providing a joint analysis of the forecasting power of both implied volatility and the separate continuous and jump components of realized volatility, we are able to address a host of issues from the literature in the present paper.

[1] There are several advantages to using log-volatilities. Of the three transformations considered, it is closest to Gaussianity, and so, as emphasized by Andersen et al. (2005, p. 6), it is more amenable to the use of standard time series procedures. Furthermore, using log-volatilities automatically imposes non-negativity of the fitted and forecasted volatilities.

We study the S&P 500 index and the associated SPX options, which have the advantages of being heavily traded and European style, i.e., there is no early exercise issue. We compute alternative volatility measures from the separate data segments: Realized volatility and its continuous and jump components from high-frequency index returns, and implied volatility from option prices. We show that implied volatility contains incremental information relative to both the continuous and jump components of realized volatility when forecasting subsequently realized index return volatility. Indeed, implied volatility is found to be an unbiased forecast, and in a few of our specifications even subsumes the information content of both components of realized volatility. This shows that there is volatility information in option prices which is not contained in return data, and that the continuous and jump components of realized volatility do not span investors' information set. Furthermore, implied volatility from option prices retains its dominant role in a forecasting context even when compared to realized volatility split into its separate components and even when using high-frequency (as opposed to daily) returns in constructing these. As an additional novel contribution, we consider separate forecasting of the continuous and jump components of future realized volatility. Our results show that implied volatility has predictive power for both components, and in particular that even the jump component of realized volatility is, to some extent, predictable.

To verify the robustness of our conclusions, we conduct an number of additional analyses. Since implied volatility is the new variable added in our study, compared to the realized volatility literature, and since it may potentially be measured with error stemming from non-synchronicity between sampled option prices and corresponding index levels, stale prices of individual stocks making up the index, bid-ask spreads, model error, etc., we take special care in handling this variable. In particular, we consider an instrumental variables approach, using lagged values of implied volatility along with the separate components of past realized volatility as instruments. In addition, we provide a structural vector autoregressive (VAR) analysis of the systems consisting of implied volatility in conjunction with realized volatility or its two separate components. Both the instrumental variables analysis and the structural VAR analysis control for possible endogeneity of implied volatility in the forecasting regression. Furthermore, the simultaneous system approach allows testing interesting cross-equation restrictions. The results from these additional analyses reinforce our earlier conclusions, namely, in particular, that implied volatility is the dominant forecasting variable in investors' information set.

The results are interesting and complement both of the above mentioned strands of literature. First of all, although implied volatility had earlier been found to forecast better than past realized volatility, it might have been speculated that by measuring past realized volatility more precisely, using high-frequency return data (Poteshman (2000) and Blair et al. (2001) suggest this in the context of the implied-realized volatility relation), or by combining its separate continuous and jump components optimally, e.g. with unequal coefficients, it would be possible to construct an even better forecast of future volatility than that contained in option prices. We find that this is not so. Secondly, since recent high-frequency data analysis shows that forecasts are improved by splitting realized volatility into its separate components, it might have been anticipated that these together summarize the relevant information set. Again, we reject the conjecture, showing that incremental information may be obtained in the option market.

The remainder of the paper is laid out as follows. In the next section, we describe the development of realized volatility and the nonparametric identification of its separate continuous sample path and jump components. In Section 3, we discuss the implied-realized volatility relation. Section 4 presents our data and empirical results and Section 5 concludes.

2 Estimation of Jumps in Financial Markets

In a typical asset pricing model the log-price of a financial asset, $p(t)$, is assumed to follow a continuous time stochastic volatility model (see e.g. Ghysels, Harvey & Renault (1996) or Barndorff-Nielsen & Shephard (2001) and the references therein for overviews of the vast literature on this topic) with an additive jump process. In particular, we assume $p(t)$ follows the general jump diffusion model expressed in stochastic differential equation form as

$$dp(t) = \mu(t)\,dt + \sigma(t)\,dw(t) + \kappa(t)\,dq(t), \quad t \geq 0, \tag{1}$$

where the mean process $\mu(\cdot)$ is continuous and locally bounded and the instantaneous volatility process $\sigma(\cdot) > 0$ is càdlàg, and both are assumed to be independent of the standard Brownian motion $w(\cdot)$. The simple counting process $q(t)$ is normalized such that $dq(t) = 1$ corresponds to a jump at time t and $dq(t) = 0$ otherwise. With this normalization, $\kappa(t)$ is the size of the jump at time t in case $dq(t) = 1$. Furthermore, we denote by $\lambda(t)$ the possibly time varying intensity of the arrival process for jumps.[2]

Allowing instantaneous volatility to be random (note that the càdlàg assumption in fact allows for jumps in the instantaneous volatility process) and serially correlated, the model (1)

[2] Formally, $\Pr(q(t) - q(t-h) = 0) = 1 - \int_{t-h}^{t} \lambda(s)\,ds + o(h)$, $\Pr(q(t) - q(t-h) = 1) = \int_{t-h}^{t} \lambda(s)\,ds + o(h)$, and $\Pr(q(t) - q(t-h) \geq 2) = o(h)$. Note that this assumption rules out infinite activity Lévy processes, e.g. the normal inverse Gaussian process, with infinitely many jumps in finite time intervals.

will generate returns with (unconditional) distributions that are fat-tailed and exhibit volatility clustering. This replicates more closely actually observed processes than constant volatility, and e.g. allows the model to overcome some of the shortcomings of the basic Black & Scholes (1973) and Merton (1973) option pricing model, see Hull & White (1987).

An important feature of the model (1) is that, in the absence of jumps,

$$p\left(t\right)|\int_0^t \mu\left(s\right)ds, \sigma^{2*}\left(t\right) \sim N\left(\int_0^t \mu\left(s\right)ds, \sigma^{2*}\left(t\right)\right), \tag{2}$$

where

$$\sigma^{2*}\left(t\right) = \int_0^t \sigma^2\left(s\right)ds \tag{3}$$

is called the integrated volatility (or integrated variance) and is an object of primary interest. For instance, in pricing options this is the relevant volatility measure, see Hull & White (1987), and for the econometrician this is an object to be estimated, see also Andersen & Bollerslev (1998*a*).

Another related important quantity is the quadratic variation (or notional volatility/variance in the terminology of Andersen, Bollerslev & Diebold (2004)) process of $p(t)$, denoted $[p](t)$, which is defined for any semimartingale (see e.g. Protter (1990)) by

$$[p]\left(t\right) = p^2\left(t\right) - 2\int_0^t p\left(s-\right)dp\left(s\right), \tag{4}$$

or equivalently

$$[p]\left(t\right) = p\lim \sum_{j=1}^{M} \left(p\left(s_j\right) - p\left(s_{j-1}\right)\right)^2, \tag{5}$$

where $0 = s_0 < s_1 < ... < s_M = t$ and the limit is taken for $\max_j |s_j - s_{j-1}| \to 0$ as $M \to \infty$.

Under some very general regularity conditions, which allow the instantaneous volatility process to exhibit many irregularities such as jumps, it is well known that the quadratic variation process for the model (1) is given by

$$[p]\left(t\right) = \sigma^{2*}\left(t\right) + \sum_{s=0}^{q(t)} \kappa^2\left(s\right), \tag{6}$$

i.e., the sum of integrated volatility and the squared jumps that have occured through time t, see e.g. the discussion in Andersen, Bollerslev, Diebold & Labys (2001, 2003).

Direct modeling of price processes via jump diffusion models such as (1) is standard in the financial asset pricing literature, and direct estimation of (1) has been considered recently by e.g. Andersen & Bollerslev (1998*b*) and Andersen, Bollerslev, Diebold & Vega (2004) for information arrivals, see also Andersen et al. (2002), Chernov et al. (2003), Eraker et al. (2003), Eraker (2004), Ait-Sahalia (2004), and Johannes (2004). These studies all find that jumps are an integral part of the price process and thus point towards the importance of incorporating jumps in the estimation of the parameters of the price process.

Instead of directly modeling (1), we take a different nonparametric approach to identifying the two components (integrated volatility resp. the jump term) of the quadratic variation process (6) using high-frequency data, following Barndorff-Nielsen & Shephard (2003*a*, 2003*b*, 2004*a*, 2004*b*), Andersen et al. (2005), and others. Assume that T months of intra-monthly observations are available and denote the intra-monthly observations for month t on the log-price of the asset by $p_{t,j}$. The time period could be any arbitrary period (e.g. intra-daily or intra-weekly), but our empirical analysis below is based on intra-monthly observations in order to estimate volatility components at a monthly frequency. Suppose n_t intra-monthly observations are available in month t. It is often desirable to have observations that are evenly spaced in time, i.e. M evenly spaced observations may be desired based on the n_t intra-monthly and possibly irregularly spaced observations. To avoid the problem of irregularly spaced data in high-frequency data sets, either linear interpolation or imputation (using the last observed price) is typically used. In this way, a data set of M evenly spaced intra-monthly price observations can be constructed based on an irregularly spaced high-frequency data set.

Using these M evenly spaced log-price observations we denote the (continuously compounded) intra-monthly returns for month t by

$$r_{t,j} = p_{t,j} - p_{t,j-1}, \quad j = 1, ..., M, \quad t = 1, ..., T. \tag{7}$$

Using (5), quadratic variation can be estimated by

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2, \quad t = 1, ..., T, \tag{8}$$

which is denoted the realized volatility of the process $p(\cdot)$ for month t. Thus, a time series of observations on realized volatility is obtained. Some authors refer to the quantity (8) as the realized variance and reserve the name realized volatility for the square root of (8), e.g. Barndorff-Nielsen & Shephard (2001, 2002*a*, 2002*b*), but we shall use the more conventional name realized volatility.

Note that the coarseness of the realized volatility estimator is governed by the choice of M. Choosing a higher number of intra-monthly returns improves the precision of the estimator, but at the same time makes it more sensitive to microstructure effects in the market, e.g. measurement errors, bid-ask bounces, etc., see Campbell, Lo & MacKinlay (1997). Several studies have considered the practical choice of M, e.g. Ait-Sahalia, Mykland & Zhang (2003), Bandi & Russell (2003), Zhang, Mykland & Ait-Sahalia (2003), and Nielsen & Frederiksen (2004), among many others. In our implementation, we follow the majority of the literature and use 5-minute returns. The 5-minute sampling frequency is close to optimal in the presence of market microstructure noise, as argued theoretically by Bandi & Russell (2003) and in simulations by Nielsen & Frederiksen (2004). For related (theoretical) results on optimal sampling schemes for maximum likelihood estimation of diffusions in the presence of market microstructure noise, see Ait-Sahalia et al. (2003).

As argued by Andersen & Bollerslev (1998*a*), Andersen, Bollerslev, Diebold & Labys (2001) and Barndorff-Nielsen & Shephard (2002*a*, 2002*b*), RV_t in (8) is by definition a consistent (in probability and uniformly in t) estimator of the increment to the quadratic variation process (6), using (5). The consistency result does not require that the observations are evenly spaced, only that the maximum distance between observations goes to zero in the limit. Thus, as $M \to \infty$,

$$RV_t \to_p \int_{t-1}^{t} \sigma^2(s)\, ds + \sum_{s=q(t-1)}^{q(t)} \kappa^2(s) = \sigma_t^{2*} + \sum_{s=q(t-1)}^{q(t)} \kappa^2(s), \tag{9}$$

defining the month t integrated volatility as $\sigma_t^{2*} = \int_{t-1}^{t} \sigma^2(s)\, ds$. The latter is the component of the quadratic variation process that is due to the continuous sample path element of the price process (1). Therefore, realized volatility is a consistent estimator of the key integrated volatility measure, σ_t^{2*}, only in the absence of jumps. Furthermore, Barndorff-Nielsen & Shephard (2002*a*) showed that (in the absence of jumps) RV_t converges to σ_t^{2*} in probability at rate $\sqrt{M}$ and satisfies a mixed Gaussian asymptotic distribution theory.

In a recent series of papers, Barndorff-Nielsen & Shephard (2003*a*, 2003*b*, 2004*a*, 2004*b*) have shown that separate nonparametric identification of the two components in (6), i.e., the continuous sample path and jump components, is possible using what is termed bipower and tripower variation measures. In particular, the (first lag) realized bipower variation is defined as

$$BV_t = \mu_1^{-2} \sum_{j=2}^{M} |r_{t,j}|\, |r_{t,j-1}|, \quad t = 1, ..., T, \tag{10}$$

where $\mu_1 = \sqrt{2/\pi}$. Barndorff-Nielsen & Shephard (2004b) showed that

$$BV_t \rightarrow_p \int_{t-1}^{t} \sigma^2(s)\, ds = \sigma_t^{2*}, \quad \text{as } M \rightarrow \infty, \tag{11}$$

i.e. realized bipower variation is a consistent estimator of integrated volatility, that is, the continuous component of the quadratic variation process (see Barndorff-Nielsen, Graversen & Shephard (2004) and Barndorff-Nielsen & Shephard (2004a) for surveys of further results on power variation). It follows that the jump component of the quadratic variation process can be estimated consistently as

$$RV_t - BV_t \rightarrow_p \sum_{s=q(t-1)}^{q(t)} \kappa^2(s). \tag{12}$$

Two issues immediately arise in relation to the estimation of the jump component by the difference between realized volatility and bipower variation. First, it is desirable in applications to ensure non-negativity of the estimate of the jump component, and this can be done simply by imposing a non-negativity truncation on $RV_t - BV_t$. Secondly, $RV_t - BV_t$ can be positive due to sampling variation even if there is no jump during month t, and thus we need the notion of a "significant jump". Barndorff-Nielsen & Shephard (2003a, 2004b) show that in the absence of jumps,

$$\sqrt{M} \frac{RV_t - BV_t}{\left(\left(\mu_1^{-4} + 2\mu_1^{-2} - 5\right) \int_{t-1}^{t} \sigma^4(s)\, ds\right)^{1/2}} \rightarrow_d N(0,1), \quad \text{as } M \rightarrow \infty, \tag{13}$$

where $\int_{t-1}^{t} \sigma^4(s)\, ds$ is called the integrated quarticity and needs to be estimated to make this a feasible test of the significance of the jump component for month t. The integrated quarticity may be estimated by the realized tripower quarticity measure

$$TQ_t = \frac{1}{M}\mu_{4/3}^{-3} \sum_{j=3}^{M} |r_{t,j}|^{4/3} |r_{t,j-1}|^{4/3} |r_{t,j-2}|^{4/3}, \quad t = 1, ..., T, \tag{14}$$

where $\mu_{4/3} = 2^{2/3}\Gamma(7/6)/\Gamma(1/2)$. This estimator is consistent, i.e.,

$$TQ_t \rightarrow_p \int_{t-1}^{t} \sigma^4(s)\, ds, \quad \text{as } M \rightarrow \infty. \tag{15}$$

Combining the results in (13) and (15), a feasible test for the significance of jumps may be obtained by replacing the integrated quarticity by TQ_t. However, simulations by Huang

& Tauchen (2005) suggest that the resulting test statistic has poor properties in finite samples. Following Huang & Tauchen (2005) and Andersen et al. (2005), we therefore employ a logarithmically transformed test statistic. Thus, in the absence of jumps,

$$Z_t = \sqrt{M} \frac{\ln RV_t - \ln BV_t}{\left(\left(\mu_1^{-4} + 2\mu_1^{-2} - 5 \right) TQ_t BV_t^{-2} \right)^{1/2}} \rightarrow_d N(0,1), \quad \text{as } M \rightarrow \infty, \tag{16}$$

where the distribution follows by application of the delta rule. A feasible test for jumps is obtained by identifying extreme (positive) values of Z_t with a significant jump during month t. In particular, we define the jump component of realized volatility by

$$J_t = I_{\{Z_t > \Phi_{1-\alpha}\}} (RV_t - BV_t), \quad t = 1, ..., T, \tag{17}$$

where $I_{\{A\}}$ is the indicator function of the set A, $\Phi_{1-\alpha}$ is the $100(1-\alpha)\,\%$ point of the standard normal distribution, and α is the chosen significance level. With J_t the estimator of the jump component of quadratic variation, we lastly formally define the estimator of the continuous component of quadratic variation as

$$C_t = RV_t - J_t, \quad t = 1, ..., T, \tag{18}$$

which is chosen to ensure that the estimators of the jump and continuous sample path components add up to realized volatility (otherwise we could have just used the realized bipower variation defined in (10)). In other words, the month t continuous component of realized volatility is equal to realized volatility if there is no jump in month t and equal to realized bipower variation if there is a jump in month t, i.e. $C_t = I_{\{Z_t \leq \Phi_{1-\alpha}\}} RV_t + I_{\{Z_t > \Phi_{1-\alpha}\}} BV_t$.

Thus, J_t and C_t from (17) and (18) are called the jump component respectively the continuous component of realized volatility, and are estimators of the corresponding components of quadratic variation in (6). It follows that the two components of quadratic variation may be consistently estimated by the corresponding components of realized volatility, i.e. $C_t \rightarrow_p \sigma_t^{*2}$ and $J_t \rightarrow_p \sum_{s=q(t-1)}^{q(t)} \kappa^2(s)$ as $M \rightarrow \infty$ if also $\alpha \rightarrow 0$ (possibly as a function of M).

Note that using standard significance levels (or any $\alpha < 1/2$) ensures that both J_t and C_t from (17) and (18) are automatically positive since $\Phi_{1-\alpha} > 0$ for $\alpha < 1/2$. Hence, this high-frequency data approach allows for month-by-month separate nonparametric consistent (as $M \rightarrow \infty$) estimation of both components of quadratic variation, i.e. the jump component and the continuous sample-path or integrated volatility component, as well as the quadratic variation process itself.

3 The Implied-Realized Volatility Relation

Existing empirical work on the continuous and jump components of realized volatility from the previous section (e.g. Andersen et al. (2005)) has not considered the role of implied volatility from option prices in forecasting future realized volatility and its components. Other previous work has shown that implied volatility provides a rather precise forecast of future realized volatility itself. Thus, if option market participants are rational and markets are efficient, then the price of a financial option should reflect all publicly available information about expected future return volatility of the underlying asset over the life of the option, and empirical evidence supports this notion. Tests of this hypothesis have typically employed the option pricing formula of Black & Scholes (1973) and Merton (1973) - henceforth the BSM formula. According to this, the fair (arbitrage free) price of a European call option with τ periods to expiration and strike price k is given by

$$c(s,k,\tau,r,\sigma) = s\Phi(\delta) - e^{-r\tau}k\Phi\left(\delta - \sigma\sqrt{\tau}\right), \tag{19}$$
$$\delta = \frac{\ln(s/k) + \left(r + \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}},$$

where s is the price of the underlying asset, r is the riskless interest rate, Φ is the standard normal c.d.f., and σ is the return volatility of the underlying asset through expiration of the option (τ periods hence).

Given an observation c of the price of the option, the implied volatility IV may be determined by inverting (19), i.e. solving the nonlinear equation

$$c = c\left(s,k,\tau,r,\sqrt{IV}\right) \tag{20}$$

numerically with respect to IV, for given data on s,k,τ, and r. If this is done every period t, a time series IV_t results. Each implied volatility IV_t may now be considered as the market's forecast of the actually realized return volatility of the underlying asset, from the time t where IV_t is calculated and until expiration of the option at $t+\tau$.

Various versions of this approach have been adopted empirically e.g. by Day & Lewis (1992), Lamoureux & Lastrapes (1993), Jorion (1995), Christensen & Prabhala (1998), Fleming (1998), and Blair et al. (2001). For instance, Christensen & Prabhala (1998) considered regression specifications of the type

$$y_t = \alpha + \beta x_t + \varepsilon_t, \tag{21}$$

where y_t is the chosen measure of realized volatility, either $RV_t^{1/2}$, $\ln RV_t$, or RV_t, measured over the course of month t, and x_t is the corresponding transformation of IV_t, measured at the beginning of month t. The unbiasedness hypothesis of interest is that $\beta = 1$. In addition, the hypotheses $\alpha = 0$, no serial correlation in the residuals, or no correlation with other variables in the information set at time t, can be tested. A monthly sampling frequency was employed for x_t and y_t. The underlying asset was the S&P 100 stock market index, and y_t was calculated from daily returns. The options were at-the-money ($k_t = s_t$) one-month ($\tau = 1/12$) OEX contracts. Basic ordinary least squares regression in (21) produced a β-estimate less than unity for both the square-root and log-transform. Correction for the potential errors-in-variable problem in implied volatility (due to bid-ask spreads, nonsynchronicities, model error, etc.) using an instrumental variables approach yielded a β-estimate insignificantly different from unity, consistent with the unbiasedness hypothesis. These results are broadly consistent with others from the more recent literature, including Jorion (1995), who considered the foreign exchange market, Fleming (1998), and Blair et al. (2001).

The results of Christensen & Prabhala (1998) suggest that when adding other variables from the information set on the right-hand side of the regression, in particular lagged realized volatility, then this does not improve the forecast, i.e., implied volatility is efficient in this sense. The new approach to decomposition of realized volatility, RV_t, into its continuous and jump components, C_t and J_t, allows refining this analysis, inspecting whether implied volatility carries incremental information relative to each of the two components of lagged realized volatility separately, or only relative to their simple sum (the case from the literature). Furthermore, computing realized volatility from high-frequency (5-minute) returns instead of daily returns allows examining whether poor measurement of realized volatility had biased the forecasting results in favor of implied volatility in the earlier literature, as argued by Poteshman (2000) and Blair et al. (2001). Finally, by splitting the left-hand side variable into its continuous and jump components, differences with respect to how each of these is best forecast may be investigated. It is these new analyses to which we turn in the empirical section below.

By focusing on the standard BSM formula, which is commonly used by practitioners, it is possible to gauge whether implied volatility backed out from this formula is in fact calibrated to incorporate jump risk, rather than merely forecasting the continuous component of future volatility. On the other hand, explicitly accounting for stochastic volatility, the Hull & White (1987) option pricing formula takes the form of an expected BSM formula with $\sigma^2\tau$ replaced by integrated volatility, σ^{*2}. Since the BSM formula is approximately linear in σ for near-the-money options (Feinstein (1988)), $IV_t^{1/2}$ in this case approximately estimates expected integrated volatility and so should forecast $C_{t+1}^{1/2}$. Again, our empirical analysis now allows investigating whether in fact IV_t incorporates jump risk, too.

4 Empirical Results

4.1 Data and Summary Statistics

We analyze the U.S. stock market as represented by the S&P 500 index. The S&P 500 options (SPX options) are traded frequently and heavily. When using the BSM formula (19), the SPX options have the advantage over the OEX contracts considered by Day & Lewis (1992), Christensen & Prabhala (1998), and Fleming (1998) that they are European style as assumed in the formula. Thus, unlike the OEX options, the SPX options have no early exercise feature and no built-in wildcard option stemming from nonsynchronicity between exercise price (index level at NYSE close, 4:00 PM EST) and index level when exercising (until CBOE close at 3:15 PM CST), see Harvey & Whaley (1992). In addition, hedging using the heavily traded SPX futures contracts should work better for SPX options than for OEX options, for which there is no directly associated futures contract. Consequently, there is reason to believe that arbitrage pricing is more precise for SPX than for OEX contracts.

To calculate the implied volatilities we use at-the-money calls with one month to expiration, which were traded on the Chicago Board of Options Exchange on a monthly basis starting in May 1986 (expiration month). By convention, SPX options expire on the third Saturday of every month. We sample the call which is closest to being at-the-money at the following Monday close and which expires the following month. We record the price as the bid-ask midpoint. This produces a sequence of sampled options covering nonoverlapping intervals. We have option data from the Berkeley Options Data Base (BODB) (see the BODB User's Guide or Rubinstein & Vijh (1987) for a description) until December 1995, and supplement with data collected from Wall Street Journal and Financial Times until December 2002 (expiration month). Thus, we analyze more recent time periods than Day & Lewis (1992) who consider data only until 1989, Christensen & Prabhala (1998) (data until May 1995), Fleming (1998) (data until April 1992), and Blair et al. (2001) (data until 1999). From each sampled quote, an implied volatility is backed out using the BSM formula (19), as described in the previous section. We correct for dividends on the S&P 500 index as described in Merton (1973) and Hull (2002, pp. 268-269). Dividend yields and risk-free interest rates (US Eurodollar deposit 1 month (bid, 11 AM, London) middle rate) are obtained from Datastream. We use two different measures of time to expiration to reflect that calendar days are relevant for interest and dividends, and trading days for volatilities, following French (1984) and Hull (2002, p. 252).

Realized volatility and its components are based on 5-minute S&P 500 returns using linear interpolation following Müller, Dacorogna, Olsen, Pictet, Schwarz & Morgenegg (1990), Dacorogna, Müller, Nagler, Olsen & Pictet (1993), and Barucci & Reno (2002), among others. We consider a monthly sampling frequency for each volatility measure (covering the same nonoverlapping intervals as the implied volatilities), resulting in roughly 2,000 intra-monthly return observations for each realized volatility (97 per day and approximately 20 trading days per month). The remaining volatility measures are then constructed as in section 2 above using a 0.1% significance level, i.e. $\Phi_{1-\alpha} = 3.090$, to detect jumps. For a more detailed description of the construction of realized volatility and its components, see Andersen, Bollerslev, Diebold & Vega (2004) or Andersen et al. (2005).

All variables are stated on a monthly basis resulting in a total of $T = 200$ observations. We use the convention that our time index refers to the month where implied volatility is sampled. Thus, implied volatility, IV_t, is measured on the Monday immediately following the time interval over which RV_t and its components C_t and J_t are calculated. For example, suppose t refers to the month of June in a given year. Implied volatility, IV_t, is backed out from an option price sampled on the Monday following the third Saturday of June. Realized volatility, RV_t, and its components, C_t and J_t, are measured over the interval starting the Monday following the third Saturday of May and ending on expiration date t, i.e. the Friday preceding the third Saturday of June. With this convention, we may consider IV_t as a forecast of RV_{t+1}, since implied volatility is sampled at the beginning of the month covered by realized volatility.

We conduct our entire analysis for three different transformations of the data, the raw variance form, the standard deviation form, and logarithmically transformed (log-volatility) variables. To avoid taking the logarithm of zero, the jump component J_t, which equals zero in the case of no significant jumps during the month, is in the case of log-transformed volatility measures replaced by $\tilde{J}_t$, obtained by substituting the smallest non-zero observation from the time series for each zero observation. It turns out that 10% of the J_t observations are zero, corresponding to 20 out of 200 months without significant jumps, and that the smallest non-zero observation is 0.0075 in standard deviation form. For comparison, Andersen et al. (2005) found no significant jumps in 91.8% of daily observations. Obviously, there are more true jumps in a month than in a day on average under any model, e.g. the Poisson arrival model (1), and so it is reasonable to find a larger proportion of significant jumps at the monthly frequency than at the daily.

Table 1 about here

Table 1 presents summary statistics for the four annualized volatility measures under each of the three transformations we consider. Confirming the results of Andersen, Bollerslev, Diebold & Ebens (2001), the logarithmic transform is the one bringing realized volatility closest to Gaussianity (Panel A). It is seen from Table 1 that the same is true for both the continuous component and the jump component, see also Andersen et al. (2005). Our results in Table 1 establish that it is also the logarithmic transform that leaves implied volatility, the new variable in our study, closest to normal. All results in the remainder of the paper are shown for all three transformations of the variables and presented in the same order as in Table 1, with the logarithmic form being closest to Gaussianity and the variance form being most different from Gaussianity for all variables. From Panel C, the jump component constitutes 11.11% of realized volatility, on average, and exhibits considerable variation itself. The coefficient of variation (standard deviation divided by mean) is close to 3 for J_t, compared to only about 2 for RV_t.

Comparing across volatility measures, implied and realized volatility have roughly equal means, consistent with the notion that the former is the unbiased expectation of the latter. Both realized volatility and its continuous component have greater standard deviations than implied volatility, further supporting the interpretation of implied volatility as a conditional expectation. Finally, implied volatility is closer to normal than realized, even under the logarithmic transform, where realized is quite close to Gaussian (c.f. Andersen, Bollerslev, Diebold & Ebens (2001)).

Figure 1 about here

Figure 1 shows time series plots of the four volatility measures under each of the three transformations. The October 19, 1987 stock market crash greatly affects the November 1987 (expiration month) observation, which is only shown in the top panel (log-volatility representation) in order not to distort the axes in the other panels, and implied volatility is seen to hover above realized volatility for more than a year after the crash, possibly reflecting increased investor fears of a second crash, as suggested by Bates (1991). As expected, the measured jump component for the crash month is very large. Still, it does not explain the entire movement in realized volatility for the month, which makes sense, as a large jump may reasonably be expected to be accompanied by an increase in continuous sample path stochastic volatility. Furthermore, towards the end of the sample period, there are several equally large jump components, possibly associated with the September 11, 2001 terrorist attacks and subsequent uncertainty, and possibly the Enron scandal. We choose to include the full time series in our empirical analysis, rather than excluding the crash month and other months with large jumps and introducing ad hoc subperiod analysis, since, in particular, our approach of nonparametrically separating the continuous sample path and jump components is explicitly geared towards accommodating jumps. In general, realized volatility and its continuous component follow each other closely and are nearly indistinguishable in parts of the figure. Implied volatility is slightly above the two, especially in the earlier part of the sample period but nonetheless exhibits a very similar pattern, even though it is calculated from a completely separate data set on option prices rather than high-frequency returns. The jump component clearly behaves differently, as expected from Table 1 and Andersen et al. (2005), but is by no means negligible, hence reinforcing the importance of treating the individual components of realized volatility separately.

4.2 The Information Content of the Continuous and Jump Components

Table 2 shows the results of regression of future realized volatility on variables in the information set at the beginning of the period. Panel A of the table is for the logarithmically transformed (log-volatility) series. Panel B shows results for the case where the square-root transform is applied to all variables, i.e. volatilities are in standard deviation form, and Panel C for the raw (variance) form of the variables. Within each of these three panels of the table, each line corresponds to a single regression. Numbers reported are coefficient estimates (estimated standard errors in parentheses), adjusted R^2, and the Breusch-Godfrey (henceforth BG) test statistic for residual autocorrelation (up to lag 12), which is used instead of the standard Durbin-Watson statistic due to the presence of lagged endogenous variables in some of our specifications. The BG statistic is asymptotically χ^2 with 12 degrees of freedom under the null of no residual autocorrelation.

Table 2 about here

The first four rows of each panel show the results of univariate regression on each of the explanatory variables. The regression specifications corresponding to the first three lines take the form

$$RV_{t+1} = \alpha + \gamma x_t + \varepsilon_{t+1}, \tag{22}$$

where x_t is the lagged value of realized volatility, RV_t, or its continuous component, C_t, or jump component, J_t (note that $\tilde{J}_t$ replaces any J_t term in x_t in the log-regressions, as explained in Section 4.1 above). Here and in the following we leave out the explicit transformations (logarithm or square-root) in our equations which are used to refer the all three transformations, but the explicit transformations are used in the tables. Focussing first on Panel A (volatility in logarithmic form), the first line shows that, as expected, lagged realized volatility, RV_t, does have significant explanatory power for the future realization, RV_{t+1}. The first-order autocorrelation coefficient is .77, and the regression explains 59% of the variation in future realized volatility. It is now of interest to examine whether it is the continuous component, C_t, or the jump component, J_t, of lagged realized volatility that has the majority of the forecasting power. The second and third lines of the table show that both the continuous and jump components of lagged realized volatility enter significantly in the univariate regressions, but that they get different parameter estimates. The coefficient on the continuous component is .78, similar to that on lagged realized volatility from the first line, at .77. The coefficient on the jump component is much lower, .24, and in contrast to the previous two regressions, the BG statistic shows clear signs of misspecification. Neither the continuous nor the jump component explains as much of the variation in future realized volatility as the 59% explained by lagged realized volatility. The results suggest that the continuous and jump components should not be combined in the form of raw realized volatility for the purposes of volatility forecasting, but may instead have different coefficients. This indicates the potential usefulness of the general approach of decomposing volatility into its various components in a forecasting context.

4.3 The Information Content of Implied Volatility

The next (fourth) line of Table 2 shows the regression of realized on implied volatility,

$$RV_{t+1} = \alpha + \beta IV_t + \varepsilon_{t+1}, \tag{23}$$

where implied volatility IV_t is measured at the beginning of the time interval covered by realized volatility RV_{t+1}. The slope coefficient is much higher, at 1.07, than those on realized volatility or its continuous component in the first two lines of the table. It is strongly significant, with a t-statistic of 21.68, and the regression explains 70% of the variation in future realized volatility, the maximum among the four univariate regressions. In fact, the coefficient on implied volatility is insignificantly different from unity at conventional levels (t-statistic of 1.35). Furthermore, in contrast to the previous three regressions, the intercept (first column of the table) is insignificantly different from zero in the regression of realized on implied volatility. These results indicate that implied volatility is an unbiased predictor of future realized volatility, and that it contains more predictive power than each of lagged realized volatility and the continuous and jump components of this. While this confirms earlier findings with respect to lagged realized volatility, and extends these to recent time periods and to the case of measuring realized volatility from high-frequency (5-minute) rather than daily returns, the comparison between implied volatility and the separate components of lagged realized volatility is novel. One caveat though, in this specification, is that the BG statistic, at 32.27, shows signs of misspecification.

The remainder of Panel A shows the results of multivariate regression on various groups of explanatory variables, of the form

$$RV_{t+1} = \alpha + \beta IV_t + \gamma x_t + \varepsilon_{t+1}, \tag{24}$$

where x_t is one of the lagged realized volatility measures RV_t, C_t, J_t, or the vector (C_t, J_t), and $\beta = 0$ is imposed if implied volatility is not included in the regression. The first question to examine is whether implied volatility indeed carries incremental information relative to lagged realized volatility and its separate continuous and jump components. This is addressed in the results in the fifth line of the panel.[3] It is seen that implied volatility gets a much higher coefficient, .81 (t-statistic of 9.64), than lagged realized volatility, whose coefficient is rather low in the bivariate regression, at .25, and less significant (t-statistic of 3.80). Furthermore, adjusted R^2 only increases by 1.9 percentage points when adding lagged realized volatility to the regression on implied volatility. These results suggest that implied volatility does carry incremental information relative to lagged realized volatility. While this confirms earlier results, and shows that they carry over to our more recent data and to the case of high-frequency realized volatility, we are now in a position to examine whether the conclusion holds even when splitting lagged realized volatility into its separate continuous and jump components.

The next (sixth) line in Panel A focusses on the basic split between the two components of lagged realized volatility. The coefficients on the continuous component, .75, and the jump

[3]Results are shown for once-lagged realized volatility and the components of this. We experimented with including further lags and found that these were all insignificant when implied volatility was included. In particular, including only IV_t, but not necessarily further lags, IV_{t-1} etc., suffices for rendering the additional lags of x_t insignificant.

component, .04, are still very different from each other. Adjusted R^2 is only 58.4%, i.e., 0.6 percentage point lower relative to the regression on lagged realized volatility (first line of the table), and 11.8 percentage points less than in the regression on implied volatility alone (fourth line). The latter finding shows that it is not possible to combine the separate continuous and jump components of lagged realized volatility to form a better forecast than implied volatility.

The bivariate regressions in the next two lines show that implied volatility subsumes the information content in the jump component of lagged realized volatility, which has a t-statistic of less than one, while the continuous component of lagged realized volatility remains significant. These results are confirmed in the last line of Panel A, which shows the results from regression of realized volatility on implied volatility together with both components of lagged realized volatility. The continuous component, C_t, remains significant and the jump component, J_t, enters insignificantly. Implied volatility, IV_t, remains the dominant forecasting variable, with a coefficient of .81, and a t-statistic of 9.77. Furthermore, BG shows no sign of misspecification. It is also remarkable that throughout the panel, the intercept is statistically insignificant if and only if implied volatility enters the regression. All the results are consistent with the notion that implied volatility carries the majority of the information about future realized volatility, even when compared against the separate continuous and jump components of lagged realized volatility. In this case, the regression on all three explanatory variables leaves both C_t and IV_t significant, i.e., they both have incremental information.

For completeness, we follow Andersen et al. (2005) and show also the corresponding results for the cases where each volatility measure is in standard deviation form (Panel B of Table 2) or in variance form (Panel C). The regression specifications are the same as (22)-(24) above, keeping in mind the new definitions of RV_t, IV_t, and x_t (standard deviations resp. variances replace the logarithmic measures, and J_t is used instead of $\tilde{J}_t$). In Panels B and C the same pattern for adjusted R^2 as in Panel A emerges. If implied volatility is included in the regression, the adjusted R^2 is orders of magnitude higher than when this variable is excluded. Thus, in Panel B (standard deviations) adjusted R^2 ranges between 19% and 29% when implied volatility is excluded and is 74-75% when implied volatility is included. The corresponding adjusted R^2's for Panel C (variances) are 3-5% when implied volatility is excluded and 86-89% when it is included. Furthermore, in the standard deviation specification (Panel B), the coefficients on implied volatility are very close to unity, and indeed implied volatility subsumes the information content of the other variables whose coefficients are insignificant. In fact, if the continuous and jump components are excluded and only implied volatility is used to forecast subsequent realized volatility, the BG test no longer shows signs of misspecification (fourth line of Panels B and C).

The results obtained so far contribute to the existing literature in interesting ways. It has been known that lagged realized volatility carries little incremental information relative to implied volatility (see Christensen & Prabhala (1998) and Blair et al. (2001)). The new finding is that when basing realized volatility on high-frequency rather than daily returns and splitting lagged realized volatility into its continuous and jump components, implied volatility continues to play the dominant role in a forecasting context. In fact, implied volatility is significant whenever it is included, and in the standard deviation specification subsumes the information content of both components of realized volatility. These would seem important results of the relatively new approach to decomposition of realized volatility.

When interpreting the results, it should be recalled that implied volatility here is backed out from the BSM formula, as is standard among practitioners and in the empirical literature. Since the BSM formula does not account for jumps in asset prices, although it is consistent with a time-varying volatility process for a continuous sample path asset price process, it would perhaps be natural to expect that exactly the jump component would not be fully captured by BSM implied volatility. In fact, there are signs in this direction (Table 2, line 8 of Panel C). However, implied volatility retains the majority of the explanatory power, suggesting that option prices have been calibrated to incorporate the effect of jumps to some extent. Of course, an alternative line of attack would be to say that the BSM model is misspecified and therefore eliminate it altogether, and apply a more general option pricing formula allowing explicitly for jumps in asset prices, e.g. as in Bates (1991) and Bakshi et al. (1997). Such an approach would entail estimating additional parameters, including prices of volatility and jump risk, in contrast to our simply backing out implied volatility directly from the BSM formula. Thus, our approach yields a conservative estimate of the information content on future quadratic variation of the log-price process contained in option prices. Our results show that simple BSM implied volatility does play an important role in a forecasting context in the presence of jumps in asset prices, more important than past realized volatility and its separate continuous and jump components, and we leave the alternative, more complicated analysis for future research.

4.4 The Role of Implied Volatility in Forecasting the Continuous and Jump Components

Besides asking about the consequences of splitting lagged realized volatility into its separate components when forecasting realized volatility, we may also split realized volatility, RV_{t+1}, on the left hand side of the regression and examine which variables forecast each of the components, C_{t+1} and J_{t+1}. If it is really the case that implied volatility is more closely related to the continuous component of realized volatility than to the jump component, then it may be expected that implied volatility plays a bigger role in forecasting C_{t+1} than in forecasting J_{t+1}. More generally, it is of interest to investigate which variables in the information set carry incremental information in forecasting the continuous and jump components of future volatility separately. This issue has not been addressed before in a setting including the implied volatility from option markets.

Table 3 about here

Table 3 shows the results for forecasting the continuous component, C_{t+1}, of realized volatility. The format is the same as in Table 2. The general regression specification is of the form

$$C_{t+1} = \alpha + \beta IV_t + \gamma x_t + \varepsilon_{t+1} \tag{25}$$

in the variance representation, with x_t containing lagged realized volatility or one or both of its components, and similarly in the other two representations, applying the square-root respectively the log-transformation to all variables. The results using the logarithmic form of volatility (Panel A) are very similar to the corresponding results in Table 2. This suggests that realized volatility and its continuous component share important features, which seems natural. Again, implied volatility gets higher coefficients and t-statistics than the other variables (lagged realized volatility and its continuous and jump components), and adjusted R^2 is highest when implied volatility is included in the regression. When forecasting the continuous component of realized volatility it is the lagged value of this and not of the jump component which potentially carries incremental information relative to implied volatility. As in Table 2 Panel A, intercepts are statistically insignificant if and only if implied volatility enters the regression.

Comparing with Panels B and C of Table 3 (standard deviations and variances), the same pattern as in Table 2 emerges. In standard deviation form (Panel B) implied volatility subsumes the information content of the other variables and is nearly unbiased. Indeed, when eliminating the insignificant variables it can not be rejected at conventional significance levels that implied volatility is an unbiased estimate of subsequent realized volatility. The t-statistic for a test of unit coefficient takes the value .74 (the similar statistic in Panel A is insignificant as well, at 1.06).

Table 4 about here

To further investigate whether implied volatility does have predictive power with respect to the jump component of realized volatility, we turn to Table 4, which reports results from regression of the future jump component, J_{t+1}, on the same explanatory variables as in the two previous tables. The general regression specification is therefore

$$J_{t+1} = \alpha + \beta IV_t + \gamma x_t + \varepsilon_{t+1} \tag{26}$$

in the variance specification, and similarly in the other two representations, operating the square-root respectively the log-transform on all variables. The univariate regression results of Table 4 show that all the forecasting variables considered have explanatory power for the jump component of realized volatility (except in Panel C where lagged realized volatility and its continuous component are insignificant). The multivariate regressions in Table 4 show that both implied volatility and the lagged jump component of realized volatility remain significant whenever they are included. Indeed, in the standard deviation and variance forms (Panels B and C), they subsume the information content of the other two variables, lagged realized volatility and its continuous component. However, the BG statistics show clear signs of misspecification in all the regressions of Panels B and C, but not in Panel A, the logarithmic regressions.

In general, coefficient estimates are clearly different from the previous two tables, showing that the jump component is quite different from the continuous component, and that the latter is most similar to realized volatility. Furthermore, implied volatility does not suffice for forecasting the jump component, but it does have incremental information relative to this. Thus, it would appear that implied volatility to some extent does forecast something more than the continuous component of realized volatility, consistent with the notion that option prices are calibrated to incorporate at least some of the jump information.

4.5 Instrumental Variables Analysis

Given that the continuous and jump components of realized volatility evidently have different properties, and that implied volatility has been shown to possess incremental information in the forecasting of both, it is of separate interest to investigate which variables in the information set implied volatility itself depends on.

Table 5 about here

Table 5, laid out as the previous three tables, shows results from the relevant regressions, which take the form

$$IV_{t+1} = \alpha + \delta z_{t+1} + \varepsilon_{t+1} \tag{27}$$

in the variance representation, where z_{t+1} contains the lagged value of implied volatility, IV_t, one of the three realized volatility measures RV_{t+1}, C_{t+1}, or J_{t+1}, or a combination of these four variables, and similarly in the other two representations, applying the appropriate square-root or log-transform to all variables. Each of the four explanatory variables considered is highly significant in univariate regression, showing that it is possible to predict not only future realized volatility and its separate components, but also future option values. The multivariate regressions show that all variables considered retain incremental information in explaining implied volatility, with the lagged jump component being the least significant regressor (t-statistic of 1.79 in the last line of Panel A). Furthermore, the BG statistic shows some signs of misspecification throughout Panel A, although least so in the fifth and last lines (p-values of 3.2% and 3.1%, respectively). Results from standard deviation and variance regressions (Panels B and C of Table 5) largely confirm these conclusions.

The results from Table 5 suggest that a combination of variables in the information set would provide a good instrument for implied volatility. This is relevant in an errors-in-variables (EIV) context. In particular, implied volatility, which is the new regressor introduced in our study relative to Andersen et al. (2005), may be measured with error and/or be correlated with the regression error term. Possible reasons for such effects include misspecification of the BSM formula, bid-ask spreads, stale prices of individual stocks making up the index, nonsynchronicities between sampled option prices and corresponding index levels due to delays in time-stamping or later closing of the CBOE compared to the NYSE, and deviations from the at-the-money target. See Christensen & Prabhala (1998) and Poteshman (2000) for further details.

Table 6 about here

Thus, we consider in Table 6 the results from two-stage least squares (2SLS) estimation of the regression specifications (23) and (24) from Table 2, i.e., forecasting realized volatility from variables in the information set. Throughout, we use lagged implied volatility along with both the continuous and jump components of realized volatility as the (three-dimensional) vector of additional instruments for implied volatility. In particular, these enter along with the explanatory variables from (23) resp. (24) as z_t in first stage regressions based on (27). Therefore, only second stage regression results from (23) and (24) (replacing IV_t by the fitted values from the first stage regression) are reported in Table 6. We report 2SLS standard errors which account for the first stage regressions, and drop adjusted R^2's which do not have the usual interpretation in the present instrumental variables framework.

The 2SLS results in Table 6 are more clear-cut than the corresponding OLS results from Table 2. Implied volatility gets a coefficient that is very close to unity, whether considering the log-volatility, standard deviation, or variance forms (first line of each panel) and intercepts are insignificant throughout the table. All other explanatory variables are insignificant at conventional levels in Panels B and C, including past realized volatility and its separate continuous and jump components, whether included individually or jointly. In all but two cases (first and fourth lines of Panel A), the coefficient on implied volatility is insignificantly different from unity, although the standard errors in some specifications are very high possibly due to multicollinearity problems. Finally, the BG statistics show the strongest signs of misspecification when the coefficient on implied volatility is significantly different from unity (lines 1 and 4 of Panel A). Thus, the results (particularly Panels B and C) suggest more strongly than the previous OLS results that implied volatility could be an unbiased and efficient forecast of subsequently realized volatility, possibly completely subsuming the information content of past realized volatility and its separate continuous and jump components.

Table 7 about here

Table 7 shows the analogous 2SLS results when substituting the continuous component, C_{t+1}, of realized volatility for RV_{t+1} as the dependent variable, corresponding to the ordinary least squares results for (25) reported in Table 3. The first stage regressions are the same as in Table 6. The results in Table 7 are similar to those in Table 6. The intercept is insignificant throughout the table as is the coefficient on the jump component of lagged realized volatility. In Panels B and C all other explanatory variables except implied volatility are insignificant, too, and the coefficient on implied volatility is insignificantly different from unity. In Panel A the other explanatory variables, lagged realized volatitity and its continuous component, are only barely significant (t-statistics of 2.06 to 2.33). Thus, implied volatility is a powerful forecast not only of realized volatility, but also of the continuous component of the latter. This begs the question of which role implied volatility from option prices plays in predicting future jump components, the question to which we turn next.

Comparing across Tables 6 and 7, particularly Panels B and C, it is noted that the coefficient on implied volatility is higher and closer to unity when forecasting realized volatility (Table 6) than when forecasting the continuous component of this (Table 7). Thus, it would appear that implied volatility to some extent does forecast something more than the continuous component of realized volatility, consistent with the notion that option prices are calibrated to incorporate at least some of the jump information.

Table 8 about here

Table 8 shows the 2SLS results from the forecasting regressions with the future jump component, J_{t+1}, of realized volatility as the dependent variable, corresponding to the ordinary least squares results for (26) reported in Table 4. Here, the results are somewhat different compared to Tables 6 and 7. Implied volatility once again proves to have predictive power, which is of course in itself striking, given that it is backed out from a formula not assuming jumps in stock prices, and this is consistent with the calibration interpretation and the ordinary least squares results from Table 4. The result goes counter to the conclusions of Andersen et al. (2005), who find that the jump component of realized volatility is essentially unpredictable. What we find is that a particular and well known variable in the information set at t, namely, implied volatility from the option market, IV_t, does in fact have statistically significant predictive power relative to J_{t+1}. Thus, the t-statistics range from 4.2 to 7.0 in univariate regressions of J_{t+1} on IV_t (first line of each panel).

When other explanatory variables are added to the specification, the significance of implied volatility is reduced, and in some cases disappears. Thus occurs mostly when the lagged jump component, J_t, is included as a regressor, and the coefficient on the latter is significant or nearly significant at conventional levels in all but one specification.

All in all, the instrumental variables results support the notion that implied volatility from option prices is the most powerful of the predictors we consider, and that it forecasts not only the continuous component of realized volatility, but may in fact be calibrated to forecast also to some extent the expected future jump component. Here, the very forecastability of the jump component is a rather novel result (see Andersen et al. (2005)). Further, if the goal is to forecast future realized volatility in standard deviation or variance form as well as possible, the predictor of choice is implied volatility itself, not some combination of this with past realized volatility and its separate continuous and jump components. The preferred coefficient on implied volatility in constructing the best forecast is unity, and so needs not be set as a function of instrumental variables, although these were employed (in Table 6) in reaching the conclusion.

4.6 Structural Vector Autoregressive Analysis

Essentially, the instrumental variables approach above is designed to handle potential endogeneity of implied volatility in the regressions of realized volatility or its components on implied volatility. In particular, even simple mismeasurement of implied volatility would leave this variable correlated with the error term in the equation. An alternative and possibly more efficient method for handling endogeneity is a simultaneous system approach. Hence, we consider the structural vector autoregressive (VAR) system

$$\begin{pmatrix} 1 & 0 \\ B_{21} & 1 \end{pmatrix} \begin{pmatrix} RV_{t+1} \\ IV_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} A_{11} & \beta \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} RV_t \\ IV_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \end{pmatrix}. \tag{28}$$

The first equation in the bivariate system corresponds to the regression of realized volatility on its own lag and implied volatility. The specification of the second equation reflects our findings that implied volatility depends on both realized volatility and its own lag, and is related in this sense to the use of these variables as additional instruments in a 2SLS treatment of the first equation. The sources of simultaneity are two: First, the leading matrix of coefficients on the left-hand side of the system is not simply the identity matrix. The off-diagonal term $-\beta$ accommodates the relation between realized and implied volatility. Secondly, the system errors $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+1}$ may be contemporaneously correlated. Indeed, since RV_{t+1} is measured over the interval ending on the Friday preceding the third Saturday in month $t+1$ and IV_{t+1} is measured already on the following Monday, accomodating such a correlation would seem a necessary extension of the modeling framework.

Table 9 about here

We estimate the structural VAR system by (Gaussian) full information maximum likelihood (FIML) and construct likelihood ratio (LR) tests of relevant hypotheses. The estimation results appear in Table 9. Panel A shows the results for variables in logarithmic form, and the results for the standard deviation and variance forms are shown in Panels B and C, respectively. Each panel has two lines, corresponding to the two equations of system (28). The results in the first line, the regression of realized volatility on implied and past realized volatility, differ from the ordinary least squares results in Table 2 (fifth line of each panel), due to the correction for endogeneity in the structural system estimation. The results also differ from the instrumental variables results in Table 6 (second line of each panel) for two reasons. Firstly, as mentioned above, system (28) is more closely related to a different instrumentation, using RV_t and IV_{t-1} as instruments for IV_t instead of the three-dimensional vector of instruments used in the Table 6 results. Secondly, system (28) accommodates potentially important correlation between the structural system errors, $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+1}$.

Across all three panels of Table 9, the (single equation) BG test statistics show virtually no signs of misspecification, which seems to be an improvement in the VAR system relative to the previous univariate specifications.[4] In the first equation of the system, the estimated coefficients on implied volatility are similar to those obtained in the ordinary least squares regressions in Table 2. We turn next to testing the joint hypotheses of interest, which are H_1: $\beta = 1$, H_2: $\beta = 1, \alpha_1 = 0$, and H_3: $\beta = 1, \alpha_1 = 0, A_{11} = 0$. These hypotheses represent increasingly restrictive versions of the general hypothesis that implied volatility from option markets is an unbiased forecast of subsequently realized return volatility in the underlying. In particular, H_3 entails that implied volatility is not only an unbiased, but also an efficient forecast of realized volatility, in the sense that it subsumes the information content of past realized volatility.

Table 10 about here

LR tests of the three hypotheses appear in Table 10. The LR test of H_1 in Panel A (variables in logarithmic form) takes the value 5.39, which in the asymptotic χ^2_1-distribution corresponds to a p-value of 2%. This may be compared to a p-value of 71% from the asymptotic t-test of the hypothesis $\beta = 1$ in the second line of Table 6. The difference is due to the different instrumentation, structural error correlation, and the fact that FIML controls efficiently for the endogeneity of implied volatility. The LR test of H_2 takes the value 7.57 on two degrees of freedom, i.e. a p-value of 2.3%. The tests of H_1 and H_2 show some evidence against the unbiasedness hypothesis at the 5% level but not the 1% level, whether or not the intercept restriction is included. Therefore we proceed to the test of H_3, i.e. whether or not implied volatility is an efficient forecast in the sense of subsuming the information content of lagged realized volatility. From the table, this test rejects strongly, indicating that implied volatility is a nearly unbiased but inefficient forecast of subsequently realized volatility.

[4] The LR test for zero correlation between the structural errors $\varepsilon_{1,t+1}$ and $\varepsilon_{2,t+1}$ in (28) takes the values 49.81, 207.99, and 439.27 in the three panels, on one degree of freedom in the asymptotic χ^2-distribution. This points to the importance of controlling for endogeneity using the structural VAR system.

The results in Panel B (variables in standard deviation form) confirm that implied volatility is nearly unbiased at the 1% level but again reject efficiency, and all three hypotheses are rejected in Panel C (variance form).[5] These results are at odds with the findings from Table 6, Panels B and C, where unbiasedness and efficiency are not rejected.

Given the differences between instrumental variables and VAR system results, and the fact that the logarithmic transformation (Panel A) leaves the variables closest to Gaussian and hence improving statistical efficiency of the FIML procedure, we place most weight on the results suggesting near unbiasedness albeit inefficiency of the implied volatility forecast.[6] This also extends the findings from the instrumental variables analysis in Christensen & Prabhala (1998), who considered data through May 1995, to more recent time periods through 2002, and to the case of realized volatility based on high-frequency rather than daily returns.

The decomposition of realized volatility into its continuous and jump components leads naturally to a trivariate generalization of the above system (28), namely,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ B_{31} & B_{32} & 1 \end{pmatrix} \begin{pmatrix} C_{t+1} \\ J_{t+1} \\ IV_{t+1} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} & \beta_1 \\ A_{21} & A_{22} & \beta_2 \\ 0 & 0 & A_{33} \end{pmatrix} \begin{pmatrix} C_t \\ J_t \\ IV_t \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t+1} \\ \varepsilon_{2,t+1} \\ \varepsilon_{3,t+1} \end{pmatrix}. \tag{29}$$

Here, the last line of the system corresponds to regression (27) from the last line of each panel of Table 5, that is, the instrumentation for implied volatility in the second stage regressions from Tables 6-8. The first two equations exactly facilitate splitting the second stage regression into its continuous and jump components in a simultaneous system framework.

Table 11 about here

The results from FIML estimation of the structural system (29) appear in Table 11. Again, Panels A, B, and C show the results for variables in log-volatility, standard deviation, and

[5]The non-diagonal structural error correlation matrix (see the above footnote) is not an additional source of inefficiency of the implied volatility forecast, since the error in the second equation, $\varepsilon_{2,t+1}$, is not known before RV_{t+1} is realized. Serial correlation in $\varepsilon_{1,t+1}$ would reinforce inefficiency, but there is very little evidence in this direction from the BG tests.

[6]Andersen et al. (2005, p. 6) also emphasize that "... from a modeling perspective, the logarithmic realized volatilities are more amenable to the use of standard time series procedures", noting in addition that "[m]odeling and forecasting log volatility also has the virtue of automatically imposing non-negativity of fitted and forecasted volatilities."

variance forms, respectively.[7] It is the coefficients on implied volatility (β_1 and β_2 in (29)) that in the larger system replace the single coefficient β on implied volatility in the first equation of (28). The results show that when forecasting the continuous and jump components of realized volatility separately, β_1 becomes closer to unity in three panels of Table 11 than the corresponding coefficient β in Table 9. The coefficient β_2 is considerably smaller than β from the smaller system under all three transformations of the variables. The implication is that implied volatility plays rather different roles in forecasting the continuous respectively the jump components of future realized volatility. Of course these differences can only be recovered due to the new approach of separating the components of realized volatility In Panel A (variables in logarithmic form) the lagged left-hand side variable has incremental forecasting power in both the first equation (forecasting $\ln C_{t+1}$) and the second equation (forecasting $\ln \tilde{J}_{t+1}$), whereas neither component adds to the forecasting of the other. From the third equation, implied volatility is forecast by the continuous component of realized volatility, and all three equations appear well specified according to the BG test. Looking across all three panels, the lagged jump component J_t always comes in as an important predictor in the second equation (forecasting J_{t+1}), and has the highest t-statistics in this equation in all panels. Indeed, jump components prove quite predictable and the other explanatory variables in the J_{t+1} equation are significant, too, except that the continuous component drops out in Panel A. It must be noted, though, that the BG tests show signs of misspecification of the jump equation in the last two panels of the table. The other two are well specified in all three panels and implied volatility comes in as the only significant forecasting variable in both the first equation of Panel B (the equation for the continuous component) and the third equation of Panel C (implied volatility equation).

Table 12 about here

The results show that the continuous and jump components of realized volatility are forecast in different ways. Although β_1 is close to unity, β_2 is not, and the jump component is forecast to a large extent by its own past. We test the hypotheses $\mathrm{H}_1 : \beta_1 = 1$, $\mathrm{H}_2 : \beta_1 = 1, \alpha_1 = 0$, and $\mathrm{H}_3 : \beta_1 = 1, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$, now with 4 degrees of freedom in the test of H_3. Here, the restriction $\beta_1 = 1$ is unbiasedness of implied volatility as a forecast of the continuous component of realized volatility, only. In H_3 the coefficients A_{11} and A_{22} on both components of realized

[7] In the structural system (29) the LR test for diagonal error correlation matrix takes the values 46.08, 167.86, and 460.37 in the three panels, now on three degrees of freedom. Again, this shows the importance of controlling for endogeneity using the structural VAR approach, although contemporaneous correlation per se does not constitute inefficiency of the implied volatility forecast.

volatility are restricted to zero to be compared to the single zero restriction added in H_3 in the bivariate system (28). LR test results appear in Table 12. The results for Panel A (variables in logarithmic form) show that H_1 is not rejected at the 1% level, i.e., implied volatility appears unbiased as a predictor of the continuous component, C_{t+1}, further supporting the conclusions from Table 7. The hypothesis H_2, adding the zero constraint on the intercept, is not rejected either (5% level). Adding the efficiency conditions, H_3 is rejected. Similarly, the results for the standard deviation form in Panel B support near unbiasedness but not efficiency of implied volatility as a forecaster of the continuous component of future realized volatility. For variables in variance form, Panel C, unbiasedness is rejected, but we place most weight on the Panel A results where variables are closest to Gaussianity.

While the results from (29) are consistent with the findings from Tables 7 and 8, the simultaneous system estimation in addition allows the testing of interesting cross-equation restrictions. In particular, we consider the additional hypothesis $H_4 : \beta_1 = 1, \beta_2 = 0$, inspecting whether implied volatility indeed only forecasts the continuous component of realized volatility, and not the jump component at all. Since it imposes restrictions on parameters in both the first and second equation of system (29), this hypothesis is only testable in the simultaneous system framework. Adding as before the zero intercept and efficiency conditions in the first equation (C_{t+1}) leads to the additional hypotheses $H_5 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0$ and $H_6 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$. Thus, H_4, H_5, and H_6 correspond to H_1, H_2, and H_3, respectively, but with the additional cross-equation restriction that implied volatility does not forecast the jump component of realized volatility (viz. $\beta_2 = 0$). The results, which are also included in Table 12, suggest that this is not the case. Thus, in Panel A (variables in logarithmic form), all the new hypotheses H_4 through H_6 are rejected at the 1% level or better, even though H_1 and H_2 were not. This indicates that implied volatility does not only forecast the continuous component of realized volatility, but also to some extent the jump component (we have not only $\beta_1 > 0$, but in fact $\beta_1 \approx 1$ and $\beta_2 > 0$). In Panels B and C, the new hypotheses, H_4 through H_6, are rejected throughout, but again we place considerable weight on the results in Panel A with log-volatilities.

Considering the nested hypotheses and testing H_4 against H_1, H_5 against H_2, respectively H_6 against H_3, instead of always testing against the unrestricted model, the difference in LR values is at least 6.37 in Panel A, 35.44 in Panel B, and 7.37 in Panel C, for p-values of 1.2%, 0.0%, and 0.7%, respectively, on one degree of freedom. Thus, maintaining unbiasedness of implied volatility (in the sense of either H_1 or H_2) or maintaining unbiasedness and efficiency (in the sense of H_3), we reject in all three panels at the 5% level (and nearly at the 1% level) that implied volatility carries no incremental information about future jump components. This supports our findings from Tables 4 and 8, and is consistent with the notion that option market participants to some extent calibrate implied volatility to accommodate expected future jumps.

5 Concluding Remarks

In this paper, we consider simultaneously two different types of forecasts of future financial market volatility. One is implied volatility, backed out from data on option prices. The other type of forecast uses volatility measurements based on high-frequency historical return data, in particular the separate continuous sample path and jump components of past realized volatility, invoking recent nonparametric identification methodology. We find that SPX option implied volatility is the dominant forecasting variable for future realized S&P 500 return volatility. In particular, implied volatility seems to be an unbiased forecast and it carries incremental information about future volatility relative to both the continuous and jump components of past realized volatility. Indeed, in a few of our specifications it is even an efficient forecast, subsuming all the information content of both components of past realized volatility. Furthermore, we find that implied volatility carries incremental information about both the continuous and the jump components of future realized volatility, thus showing in particular that even the jump component is to some extent predictable, and that option market participants calibrate prices to incorporate expected future jumps.

The dominant role of option implied volatility as a forecaster holds up to tests allowing the individual continuous and jump components of past realized volatility to act separately, and to measurement of realized volatility and its components from returns of higher frequency than the daily, which has been employed in past literature on the implied-realized volatility relation. Thus, while reduced form time series models for realized volatility and its components work reasonably well, option prices should not be ignored when forecasting financial market volatility, including for the purposes of asset pricing, derivative pricing, hedging, risk management, etc.

The inclusion of implied volatility in the modelling and forecasting of realized volatility and its components opens several avenues for future research. For example, in addition to S&P 500 realized volatility, Andersen et al. (2005) considered also realized volatility of foreign exchange and interest rates. It would be interesting to apply our methodology to such data, too, examining whether implied volatility retains its dominant role in forecasting volatility in these other markets. In addition, it would be of interest to study the information content of prices of other options, including put options and options of longer maturity, as well as out-of-the-money options, thus investigating the forecasting role of the observed smile/smirk patterns in these markets. Finally, the realized volatility computations could be based on other frequencies than the 5-minutes used here, and jumps could be identified using other significance levels than our 0.1%, thus allowing further investigation of the generality of our new results.

References

Ait-Sahalia, Y., Mykland, P. A. & Zhang, L. (2003), 'How often to sample a continuous-time process in the presence of market microstructure noise', *Forthcoming in Review of Financial Studies* .

Ait-Sahalia, Y. (2004), 'Disentangling diffusion from jumps', *Journal of Financial Economics* **74**, 487–528.

Andersen, T., Bollerslev, T., Diebold, F. X. & Vega, C. (2004), 'Real-time price discovery in stock, bond and foreign exchange markets', *Working Paper, University of Pennsylvania* .

Andersen, T. G., Benzoni, L. & Lund, J. (2002), 'An empirical investigation of continuous-time equity return models', *Journal of Finance* **57**, 1239–1284.

Andersen, T. G. & Bollerslev, T. (1998*a*), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

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

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2004), Parametric and nonparametric volatility measurement, *in* L. P. Hansen & Y. Ait-Sahalia, eds, 'Handbook of Financial Econometrics (Forthcoming)', North-Holland, Amsterdam.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2005), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *Working Paper, University of Pennsylvania* .

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

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of exchange rate volatility', *Journal of the American Statistical Association* **96**, 42–55.

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

Andersen, T. G., Bollerslev, T. & Meddahi, N. (2004), 'Analytical evaluation of volatility forecasts', *International Economic Review* **45**, 1079–1110.

Bakshi, G., Cao, C. & Chen, Z. (1997), 'Empirical performance of alternative option pricing models', *Journal of Finance* **52**, 2003–2049.

Bandi, F. M. & Russell, J. R. (2003), 'Microstructure noise, realized volatility, and optimal sampling', *Working paper, University of Chicago Graduate School of Business* .

Barndorff-Nielsen, O. E., Graversen, S. E. & Shephard, N. (2004), 'Power variation and stochastic volatility: A review and some new results', *Journal of Applied Probability* **41A**, 133–143.

Barndorff-Nielsen, O. E. & Shephard, N. (2001), 'Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion)', *Journal of the Royal Statistical Society Series B* **63**, 167–241.

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

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

Barndorff-Nielsen, O. E. & Shephard, N. (2003*a*), 'Econometrics of testing for jumps in financial economics using bipower variation', *Forthcoming in Journal of Financial Econometrics* .

Barndorff-Nielsen, O. E. & Shephard, N. (2003*b*), 'Realised power variation and stochastic volatility', *Bernoulli* **9**, 243–265.

Barndorff-Nielsen, O. E. & Shephard, N. (2004*a*), 'Multipower variation and stochastic volatility', *Working paper, Oxford University* .

Barndorff-Nielsen, O. E. & Shephard, N. (2004*b*), 'Power and bipower variation with stochastic volatility and jumps (with discussion)', *Journal of Financial Econometrics* **2**, 1–48.

Barucci, E. & Reno, R. (2002), 'On measuring volatility and the GARCH forecasting performance', *Journal of International Financial Markets, Institutions and Money* **12**, 183–200.

Bates, D. S. (1991), 'The crash of '87: Was it expected? the evidence from options markets', *Journal of Finance* **46**, 1009–1044.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Blair, B. J., Poon, S. & Taylor, S. J. (2001), 'Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns', *Journal of Econometrics* **105**, 5–26.

Campbell, J. Y., Lo, A. W. & MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton.

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

Christensen, B. J. & Prabhala, N. R. (1998), 'The relation between implied and realized volatility', *Journal of Financial Economics* **50**, 125–150.

Dacorogna, M. M., Müller, U. A., Nagler, R. J., Olsen, R. B. & Pictet, O. V. (1993), 'A geographical model for the daily and weekly seasonal volatility in the foreign exchange market', *Journal of International Money and Finance* **12**, 413–438.

Day, T. E. & Lewis, C. M. (1992), 'Stock market volatility and the information content of stock index options', *Journal of Econometrics* **52**, 267–287.

Eraker, B. (2004), 'Do stock prices and volatility jump? Reconciling evidence from spot and option prices', *Journal of Finance* **59**, 1367–1403.

Eraker, B., Johannes, M. & Polson, N. (2003), 'The impact of jumps in volatility and returns', *Journal of Finance* **58**, 1269–1300.

Feinstein, S. P. (1988), 'A source of unbiased implied volatility forecasts', *Working Paper 88-9, Federal Reserve Bank of Atlanta* .

Fleming, J. (1998), 'The quality of market volatility forecasts implied by S&P 100 index option prices', *Journal of Empirical Finance* **5**, 317–345.

French, D. W. (1984), 'The weekend effect on the distribution of stock prices: Implications for option pricing', *Journal of Financial Economics* **13**, 547–559.

French, K. R., Schwert, G. W. & Stambaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 3–30.

Ghysels, E., Harvey, A. C. & Renault, E. (1996), Stochastic volatility, *in* C. R. Rao & G. S. Maddala, eds, 'Statistical Methods in Finance', North-Holland, Amsterdam, pp. 119–191.

Harvey, C. R. & Whaley, R. E. (1992), 'Market volatility prediction and the efficiency of the S&P 100 index option market', *Journal of Financial Economics* **31**, 43–73.

Huang, X. & Tauchen, G. (2005), 'The relative contribution of jumps to total price variance', *Journal of Financial Econometrics* **3**, 456–499.

Hull, J. C. (2002), *Options, Futures, and Other Derivatives*, 5th edn, Prentice-Hall, Englewood Cliffs, New Jersey.

Hull, J. C. & White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *Journal of Finance* **42**, 281–300.

Johannes, M. (2004), 'The statistical and economic role of jumps in interest rates', *Journal of Finance* **59**, 227–260.

Jorion, P. (1995), 'Predicting volatility in the foreign exchange market', *Journal of Finance* **50**, 507–528.

Lamoureux, C. G. & Lastrapes, W. D. (1993), 'Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities', *Review of Financial Studies* **6**, 293–326.

Merton, R. C. (1973), 'Theory of rational option pricing', *Bell Journal of Economics and Management Science* **4**, 141–183.

Müller, U. A., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M. & Morgenegg, C. (1990), 'Statistical study of foreign exchange rates, empirical evidence of a price scaling law, and intraday analysis', *Journal of Banking and Finance* **14**, 1189–1208.

Nielsen, M. Ø. & Frederiksen, P. H. (2004), 'Finite sample accuracy of integrated volatility estimators', *Working paper, Cornell University* .

Poteshman, A. M. (2000), 'Forecasting future volatility from option prices', *Working Paper, University of Illinois at Urbana-Champaign* .

Protter, P. (1990), *Stochastic Integration and Differential Equations: A New Approach*, Springer-Verlag, New York.

Rubinstein, M. & Vijh, A. M. (1987), 'The Berkeley options data base: A tool for empirical research', *Advances in Futures and Options Research* **2**, 209–221.

Schwert, G. W. (1989), 'Why does stock market volatility change over time?', *Journal of Finance* **44**, 1115–1153.

Zhang, L., Mykland, P. A. & Ait-Sahalia, Y. (2003), 'A tale of two time scales: Determining integrated volatility with noisy high-frequency data', *Forthcoming in Journal of the American Statistical Association* .

Table 1: Summary statistics

Panel A: Variables in logarithmic form				
Statistic	$\ln RV_t$	$\ln C_t$	$\ln \tilde{J}_t$	$\ln IV_t$
Mean	-3.9098	-4.0022	-6.8241	-3.6474
Std. dev.	0.8347	0.8168	1.4655	0.6580
Skewness	0.6467	0.5637	-0.0929	0.5756
Kurtosis	4.3520	4.3729	3.6877	3.1489
Panel B: Variables in std. dev. form				
Statistic	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$
Mean	0.1559	0.1481	0.0425	0.1711
Std. dev.	0.0851	0.0785	0.0407	0.0669
Skewness	4.1123	4.4573	3.3834	3.1195
Kurtosis	32.4320	38.5738	18.2111	23.2883
Panel C: Variables in variance form				
Statistic	RV_t	C_t	J_t	IV_t
Mean	0.0315	0.0281	0.0035	0.0337
Std. dev.	0.0627	0.0569	0.0099	0.0401
Skewness	10.1374	11.0538	6.2254	8.5354
Kurtosis	124.1903	142.9701	45.4996	98.4287

Note: The monthly realized volatility RV_t and its continuous component C_t and jump component J_t are constructed from 5-minute S&P 500 index returns spanning the period from May 1986 through December 2002, for a total of 200 monthly observations, each based on about 2,000 5-minute returns. The monthly implied volatility IV_t is backed out from the BSM formula adjusted for dividends and applied to the at-the-money SPX call option expiring on the Friday immediately preceding the third Saturday of the given month and sampled on the Monday following the expiration date of the previous month. Each of the four volatility measures covers the same one-month interval between two consecutive expiration dates.

Table 2: Realized volatility regressions

Panel A: Dependent variable is $\ln RV_{t+1}$

Const.	$\ln RV_t$	$\ln C_t$	$\ln \tilde{J}_t$	$\ln IV_t$	Adj R^2	BG
−0.8851 (0.1827)	0.7726 (0.0456)	–	–	–	59.0%	20.126
−0.7743 (0.1922)	–	0.7827 (0.0470)	–	–	58.2%	19.913
−2.2795 (0.2607)	–	–	0.2384 (0.0373)	–	16.8%	102.700**
−0.0198 (0.1824)	–	–	–	1.0666 (0.0492)	70.2%	32.268**
0.0115 (0.1771)	0.2507 (0.0659)	–	–	0.8050 (0.0835)	72.1%	17.595
−0.6392 (0.2180)	–	0.7499 (0.0533)	0.0390 (0.0299)	–	58.4%	15.973
0.0606 (0.1783)	–	0.2494 (0.0661)	–	0.8147 (0.0821)	72.1%	18.903
0.0541 (0.1994)	–	–	0.0232 (0.0251)	1.0432 (0.0555)	70.2%	30.900**
0.0766 (0.1935)	–	0.2465 (0.0676)	0.0053 (0.0248)	0.8122 (0.0831)	71.9%	19.015

Panel B: Dependent variable is $RV_{t+1}^{1/2}$

Const.	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	Adj R^2	BG
0.0719 (0.0107)	0.5405 (0.0601)	–	–	–	28.7%	15.031
0.0724 (0.0110)	–	0.5654 (0.0659)	–	–	26.8%	21.164*
0.1169 (0.0079)	–	–	0.9305 (0.1360)	–	18.8%	34.320**
−0.0324 (0.0084)	–	–	–	1.1012 (0.0457)	74.4%	17.156
−0.0316 (0.0084)	−0.0725 (0.0484)	–	–	1.1630 (0.0616)	74.6%	26.646**
0.0746 (0.0110)	–	0.4410 (0.0841)	0.3859 (0.1645)	–	28.4%	15.043
−0.0309 (0.0084)	–	−0.0898 (0.0515)	–	1.1706 (0.0605)	74.7%	25.794*
−0.0323 (0.0085)	–	–	0.0009 (0.0887)	1.1008 (0.0533)	74.3%	18.824
−0.0296 (0.0085)	–	−0.1140 (0.0580)	0.0902 (0.0991)	1.1618 (0.0613)	74.7%	23.681*

Panel C: Dependent variable is RV_{t+1}

Const.	RV_t	C_t	J_t	IV_t	Adj R^2	BG
0.0255 (0.0049)	0.1936 (0.0699)	–	–	–	3.3%	3.325
0.0263 (0.0049)	–	0.1901 (0.0773)	–	–	2.5%	8.246
0.0265 (0.0046)	–	–	1.4990 (0.4428)	–	5.0%	2.819
−0.0174 (0.0022)	–	–	–	1.4534 (0.0416)	86.0%	12.449
−0.0154 (0.0020)	−0.1722 (0.0259)	–	–	1.5529 (0.0406)	88.5%	27.058**
0.0252 (0.0049)	–	0.0745 (0.0899)	1.2706 (0.5219)	–	4.9%	2.237
−0.0153 (0.0020)	–	−0.1894 (0.0283)	–	1.5474 (0.0402)	88.5%	23.778*
−0.0171 (0.0021)	–	–	−0.5286 (0.1768)	1.4964 (0.0433)	86.5%	26.727**
−0.0153 (0.0020)	–	−0.1874 (0.0320)	−0.0243 (0.1848)	1.5484 (0.0410)	88.5%	25.605*

Note: The table shows ordinary least squares estimation results for the regression specifications (22), (23), and (24) and the corresponding standard deviation and log-volatility regressions. Standard errors are in parentheses, Adj R^2 is the adjusted R^2 for the regression, and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 3: Continuous component regressions

Panel A: Dependent variable is $\ln C_{t+1}$

Const.	$\ln RV_t$	$\ln C_t$	$\ln \tilde{J}_t$	$\ln IV_t$	Adj R^2	BG
−1.0524 (0.1796)	0.7536 (0.0449)	–	–	–	58.7%	15.872
−0.9330 (0.1879)	–	0.7663 (0.0460)	–	–	58.3%	22.568**
−2.4840 (0.2575)	–	–	0.2221 (0.0368)	–	15.1%	105.160**
−0.1712 (0.1758)	–	–	–	1.0504 (0.0474)	71.1%	30.484**
−0.1438 (0.1715)	0.2248 (0.0638)	–	–	0.8167 (0.0809)	72.7%	17.795
−0.8512 (0.2138)	–	0.7464 (0.0522)	0.0236 (0.0293)	–	58.2%	17.742
−0.0970 (0.1722)	–	0.2322 (0.0638)	–	0.8158 (0.0793)	72.8%	17.879
−0.1497 (0.1925)	–	–	0.0069 (0.0242)	1.0435 (0.0536)	71.0%	30.263**
−0.1280 (0.1868)	–	0.2378 (0.0653)	−0.0104 (0.0240)	0.8206 (0.0802)	72.7%	18.506

Panel B: Dependent variable is $C_{t+1}^{1/2}$

Const.	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	Adj R^2	BG
0.0716 (0.0099)	0.4921 (0.0558)	–	–	–	28.0%	13.924
0.0711 (0.0102)	–	0.5208 (0.0609)	–	–	26.7%	15.343
0.1152 (0.0074)	–	–	0.7855 (0.1278)	–	15.7%	38.904**
−0.0280 (0.0074)	–	–	–	1.0298 (0.0404)	76.5%	14.069
−0.0271 (0.0074)	−0.0922 (0.0425)	–	–	1.1086 (0.0541)	77.0%	22.217*
0.0725 (0.0102)	–	0.4444 (0.0782)	0.2366 (0.1531)	–	27.2%	13.913
−0.0264 (0.0074)	–	−0.0982 (0.0453)	–	1.1058 (0.0533)	77.0%	21.733*
−0.0292 (0.0075)	–	–	−0.1132 (0.0780)	1.0643 (0.0469)	76.7%	17.327
−0.0271 (0.0075)	–	−0.0859 (0.0511)	−0.0459 (0.0874)	1.1103 (0.0541)	76.9%	22.263*

Panel C: Dependent variable is C_{t+1}

Const.	RV_t	C_t	J_t	IV_t	Adj R^2	BG
0.0229 (0.0045)	0.1662 (0.0636)	–	–	–	2.9%	4.136
0.0234 (0.0045)	–	0.1675 (0.0702)	–	–	2.3%	2.566
0.0243 (0.0042)	–	–	1.1410 (0.4056)	–	3.4%	3.213
−0.0168 (0.0019)	–	–	–	1.3317 (0.0356)	87.5%	9.739
−0.0147 (0.0016)	−0.1708 (0.0214)	–	–	1.4304 (0.0335)	90.5%	20.122
0.0227 (0.0044)	–	0.0883 (0.0822)	0.8702 (0.4775)	–	3.4%	2.116
−0.0147 (0.0017)	–	−0.1811 (0.0235)	–	1.4217 (0.0335)	90.4%	19.202
−0.0163 (0.0018)	–	–	−0.7457 (0.1454)	1.3924 (0.0356)	88.9%	19.004
−0.0148 (0.0016)	–	−0.1544 (0.0264)	−0.3301 (0.1520)	1.4353 (0.0338)	90.6%	21.479*

Note: The table shows ordinary least squares estimation results for the general regression specification (25) and the corresponding standard deviation and log-volatility regressions. Standard errors are in parentheses, Adj R^2 is the adjusted R^2 for the regression, and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 4: Jump component regressions

Panel A: Dependent variable is $\ln\tilde{J}_{t+1}$

Const.	$\ln RV_t$	$\ln C_t$	$\ln\tilde{J}_t$	$\ln IV_t$	Adj R^2	BG
−3.6970	0.7950	–	–	–	20.4%	9.381
(0.4426)	(0.1106)					
−3.6720	–	0.7832	–	–	19.0%	12.286
(0.4651)		(0.1138)				
−4.2608	–	–	0.3727	–	13.5%	20.010
(0.4617)			(0.0661)			
−3.2046	–	–	–	0.9924	19.3%	17.684
(0.5267)				(0.1421)		
−3.1268	0.4630	–	–	0.5126	21.7%	12.141
(0.5154)	(0.1918)			(0.2450)		
−2.9382	–	0.6046	0.2119	–	22.1%	7.221
(0.5184)		(0.1267)	(0.0712)			
−3.0583	–	0.3912	–	0.5988	21.1%	14.248
(0.5207)		(0.1930)		(0.2398)		
−2.5149	–	–	0.2117	0.7805	23.0%	8.974
(0.5565)			(0.0700)	(0.1548)		
−2.4886	–	0.2885	0.1908	0.5101	23.5%	7.221
(0.5551)		(0.1939)	(0.0712)	(0.2384)		

Panel B: Dependent variable is $J_{t+1}^{1/2}$

Const.	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	Adj R^2	BG
0.0122	0.1964	–	–	–	16.5%	44.280**
(0.0055)	(0.0310)					
0.0145	–	0.1904	–	–	13.1%	51.826**
(0.0057)		(0.0342)				
0.0218	–	–	0.4984	–	23.9%	30.561**
(0.0036)			(0.0627)			
0.0152	–	–	–	0.3371	30.3%	53.908**
(0.0066)				(0.0360)		
−0.0152	0.0343	–	–	0.3076	30.3%	53.253**
(0.0066)	(0.0382)			(0.0486)		
0.0170	–	0.0494	0.4374	–	24.1%	30.781**
(0.0054)		(0.0412)	(0.0807)			
−0.0150	–	0.0033	–	0.3342	30.1%	55.095**
(0.0067)		(0.0408)		(0.0479)		
−0.0120	–	–	0.2881	0.2491	36.2%	35.602**
(0.0064)			(0.0666)	(0.0400)		
−0.0098	–	−0.0935	0.3613	0.2990	37.3%	35.985**
(0.0064)		(0.0434)	(0.0743)	(0.0459)		

Panel C: Dependent variable is J_{t+1}

Const.	RV_t	C_t	J_t	IV_t	Adj R^2	BG
0.0026	0.0274	–	–	–	2.5%	83.696**
(0.0008)	(0.0110)					
0.0028	–	0.0226	–	–	1.2%	85.008**
(0.0008)		(0.0122)				
0.0023	–	–	0.3580	–	12.2%	66.530**
(0.0007)			(0.0670)			
−0.0006	–	–	–	0.1217	24.0%	110.490**
(0.0008)				(0.0152)		
−0.0006	−0.0014	–	–	0.1225	23.6%	111.050**
(0.0008)	(0.0105)			(0.0165)		
0.0025	–	−0.0138	0.4004	–	12.2%	67.160**
(0.0007)		(0.0136)	(0.0789)			
−0.0005	–	−0.0082	–	0.1257	23.8%	112.80**
(0.0008)		(0.0115)		(0.0163)		
−0.0008	–	–	0.2171	0.1040	27.8%	97.536**
(0.0008)			(0.0644)	(0.0158)		
−0.0005	–	−0.0330	0.3058	0.1131	29.9%	109.970**
(0.0008)		(0.0124)	(0.0717)	(0.0159)		

Note: The table shows ordinary least squares estimation results for the general regression specification (26) and the corresponding standard deviation and log-volatility regressions. Standard errors are in parentheses, Adj R^2 is the adjusted R^2 for the regression, and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 5: Implied volatility regressions

Panel A: Dependent variable is $\ln IV_{t+1}$

Const.	$\ln RV_{t+1}$	$\ln C_{t+1}$	$\ln\tilde{J}_{t+1}$	$\ln IV_t$	Adj R^2	BG
−1.1124 (0.1287)	0.6475 (0.0321)	–	–	–	67.2%	24.680*
−1.0249 (0.1364)	–	0.6547 (0.0334)	–	–	66.0%	24.966*
−2.2370 (0.2004)	–	–	0.2062 (0.0287)	–	20.4%	101.850**
−0.9376 (0.1787)	–	–	–	0.7422 (0.0482)	54.4%	31.608**
−0.9243 (0.1500)	0.5322 (0.0581)	–	–	0.1751 (0.0740)	67.9%	22.531*
−0.8813 (0.1539)	–	0.6198 (0.0376)	0.0415 (0.0212)	–	66.5%	24.109*
−0.8507 (0.1529)	–	0.5297 (0.0614)	–	0.1848 (0.0766)	66.8%	24.294*
−0.6943 (0.1913)	–	–	0.0740 (0.0236)	0.6701 (0.0524)	56.4%	25.443*
−0.7313 (0.1660)	–	0.5056 (0.0625)	0.0376 (0.0210)	0.1736 (0.0764)	67.2%	22.608*

Panel B: Dependent variable is $IV_{t+1}^{1/2}$

Const.	$RV_{t+1}^{1/2}$	$C_{t+1}^{1/2}$	$J_{t+1}^{1/2}$	$IV_t^{1/2}$	Adj R^2	BG
0.0891 (0.0074)	0.5271 (0.0415)	–	–	–	44.7%	29.975**
0.0883 (0.0076)	–	0.5598 (0.0456)	–	–	43.0%	30.785**
0.1356 (0.0059)	–	–	0.8444 (0.1022)	–	25.3%	57.063**
0.0688 (0.0105)	–	–	–	0.5992 (0.0573)	35.4%	19.709
0.0843 (0.0101)	0.4778 (0.0823)	–	–	0.0730 (0.1050)	44.6%	30.157**
0.0897 (0.0076)	–	0.4777 (0.0583)	0.2545 (0.1140)	–	44.1%	30.229**
0.0827 (0.0102)	–	0.4922 (0.0947)	–	0.0910 (0.1116)	42.9%	31.403**
0.0750 (0.0103)	–	–	0.4264 (0.1101)	0.4579 (0.0663)	39.7%	22.078*
0.0841 (0.0101)	–	0.4085 (0.1009)	0.2552 (0.1141)	0.0928 (0.1105)	44.0%	30.513**

Panel C: Dependent variable is IV_{t+1}

Const.	RV_{t+1}	C_{t+1}	J_{t+1}	IV_t	Adj R^2	BG
0.0263 (0.0030)	0.2356 (0.0423)	–	–	–	13.2%	17.024
0.0268 (0.0030)	–	0.2452 (0.0469)	–	–	11.7%	16.049
0.0292 (0.0028)	–	–	1.3550 (0.2742)	–	10.6%	16.570
0.0222 (0.0035)	–	–	–	0.3435 (0.0670)	11.3%	6.765
0.0262 (0.0040)	0.2321 (0.1136)	–	–	0.0059 (0.1781)	12.7%	17.030
0.0262 (0.0029)	–	0.1691 (0.0544)	0.8363 (0.3161)	–	14.3%	12.785
0.0248 (0.0041)	–	0.1577 (0.1340)	–	0.1331 (0.1908)	11.5%	15.917
0.0228 (0.0034)	–	–	0.8789 (0.3081)	0.2374 (0.0756)	14.4%	7.426
0.0241 (0.0041)	–	0.0786 (0.1353)	0.8383 (0.3165)	0.1375 (0.1879)	14.1%	12.806

Note: The table shows ordinary least squares estimation results for the regression specification (27) and the corresponding standard deviation and log-volatility regressions. Standard errors are in parentheses, Adj R^2 is the adjusted R^2 for the regression, and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 6: Realized volatility instrumental variables regressions

Panel A: Dependent variable is $\ln RV_{t+1}$					
Const.	$\ln RV_t$	$\ln C_t$	$\ln \tilde{J}_t$	$\ln IV_t$	BG
0.3753 (0.2239)	–	–	–	1.1748 (0.0607)	27.078**
−0.9688 (0.6501)	0.8214 (0.3651)	–	–	−0.0753 (0.5592)	10.777
−0.5023 (0.4748)	–	0.6090 (0.2853)	–	0.2654 (0.4310)	16.301
0.3674 (0.2281)	–	–	−0.0056 (0.0273)	1.1832 (0.0733)	27.401**
−0.8589 (0.6639)	–	0.9043 (0.4386)	0.0494 (0.0442)	−0.2493 (0.7012)	0.641
Panel B: Dependent variable is $RV_{t+1}^{1/2}$					
Const.	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	BG
−0.0176 (0.0122)	–	–	–	1.0150 (0.0690)	21.643*
0.00269 (0.0535)	0.1307 (0.3166)	–	–	0.7776 (0.5960)	25.174*
−0.0288 (0.0324)	–	−0.0763 (0.2052)	–	1.1464 (0.3599)	27.300**
−0.0049 (0.0154)	–	–	0.1719 (0.1213)	0.8983 (0.1087)	15.486
0.2055 (0.2956)	–	1.1381 (1.5786)	0.7572 (0.8938)	−1.4593 (3.2871)	17.073
Panel C: Dependent variable is RV_{t+1}					
Const.	RV_t	C_t	J_t	IV_t	BG
0.0015 (0.0055)	–	–	–	0.8935 (0.1469)	1.321
−0.0121 (0.0077)	−0.1427 (0.0718)	–	–	1.4276 (0.2862)	16.445
−0.0096 (0.0063)	–	−0.1374 (0.0616)	–	1.3354 (0.2231)	12.813
0.0160 (0.0126)	–	–	1.0098 (0.6622)	0.3611 (0.4131)	6.141
0.0531 (0.0961)	–	0.2546 (0.6363)	2.1612 (3.1778)	−1.0649 (3.6621)	13.081

Note: The table shows results from instrumental variables estimation of the regression specifications (23) and (24), and the corresponding log-volatility and variance regressions. The additional instrumentation of implied volatility consists of both components of realized volatility as well as lagged implied volatility. Standard errors are in parentheses and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 7: Continuous component instrumental variables regressions

Panel A: Dependent variable is $\ln C_{t+1}$					
Const.	$\ln RV_t$	$\ln C_t$	$\ln \tilde{J}_t$	$\ln IV_t$	BG
0.1819 (0.2154)	–	–	–	1.1473 (0.0584)	25.478*
−1.2086 (0.6583)	0.8446 (0.3697)	–	–	−0.1405 (0.5662)	6.310
−0.7997 (0.4863)	–	0.6811 (0.2923)	–	0.1301 (0.4414)	10.965
0.1525 (0.2202)	–	–	−0.0210 (0.0263)	1.1786 (0.0708)	25.615*
−1.0318 (0.6412)	–	0.8734 (0.4236)	0.0321 (0.0427)	−0.2050 (0.6772)	7.274
Panel B: Dependent variable is $C_{t+1}^{1/2}$					
Const.	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	BG
−0.0092 (0.0109)	–	–	–	0.9197 (0.0616)	17.136
0.0356 (0.0586)	0.2790 (0.3468)	–	–	0.4042 (0.6528)	9.904
0.0152 (0.0338)	–	0.1657 (0.2136)	–	0.6343 (0.37464)	18.577
−0.0076 (0.0135)	–	–	0.0208 (0.1059)	0.9056 (0.0949)	15.705
0.2025 (0.2860)	–	1.1365 (1.5273)	0.6053 (0.8648)	−1.4487 (3.1805)	13.942
Panel C: Dependent variable is C_{t+1}					
Const.	RV_t	C_t	J_t	IV_t	BG
0.0037 (0.0053)	–	–	–	0.7251 (0.1427)	4.626
−0.0000 (0.0097)	−0.0392 (0.0899)	–	–	0.8716 (0.3583)	15.606
−0.0003 (0.0074)	–	−0.0489 (0.0731)	–	0.8823 (0.2649)	9.245
0.0117 (0.0106)	–	–	0.5552 (0.5566)	0.4324 (0.3472)	4.284
0.0536 (0.0955)	–	0.2883 (0.6326)	1.8590 (3.1591)	−1.1823 (3.6405)	15.622

Note: The table shows results from instrumental variables estimation of the general regression specification (25) and the corresponding standard deviation and log-volatility regressions. The additional instrumentation of implied volatility consists of both components of realized volatility as well as lagged implied volatility. Standard errors are in parentheses and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 8: Jump component instrumental variables regressions

Panel A: Dependent variable is $\ln\tilde{J}_{t+1}$					
Const.	$\ln RV_t$	$\ln C_t$	$\ln\tilde{J}_t$	$\ln IV_t$	BG
−2.4611	–	–	–	1.1921	17.386
(0.6335)				(0.1718)	
−4.8666	1.4757	–	–	−1.0514	6.994
(1.6624)	(0.9338)			(1.4300)	
−2.6363	–	0.1216	–	1.0105	15.474
(1.2605)		(0.7576)		(1.1442)	
−2.2047	–	–	0.1831	0.9191	10.383
(0.6278)			(0.0750)	(0.2017)	
−4.8002	–	1.9141	0.2995	−2.1129	9.111
(1.7890)		(1.1819)	(0.1192)	(1.8895)	
Panel B: Dependent variable is $J_{t+1}^{1/2}$					
Const.	$RV_t^{1/2}$	$C_t^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	BG
−0.0222	–	–	–	0.3796	65.735**
(0.0095)				(0.0539)	
−0.0633	−0.2503	–	–	0.8474	34.784**
(0.0492)	(0.2911)			(0.5479)	
−0.1129	–	−0.6179	–	1.4440	8.4635
(0.0496)		(0.3139)		(0.5506)	
0.0086	–	–	0.4164	0.0971	38.265**
(0.011590)			(0.0910)	(0.0816)	
0.0613	–	0.2851	0.5630	−0.4935	0.655
(0.1093)		(0.5837)	(0.3305)	(1.2155)	
Panel C: Dependent variable is J_{t+1}					
Const.	RV_t	C_t	J_t	IV_t	BG
−0.0022	–	–	–	0.1684	148.45**
(0.0015)				(0.0398)	
−0.0120	−0.1035	–	–	0.5560	4.295
(0.0065)	(0.0605)			(0.2413)	
−0.0093	–	−0.0885	–	0.4531	3.706
(0.0041)		(0.0409)		(0.1480)	
0.0043	–	–	0.4546	−0.0713	63.849**
(0.0028)			(0.1451)	(0.0905)	
−0.0006	–	−0.0337	0.3023	0.1173	190.05**
(0.0080)		(0.0529)	(0.2641)	(0.3044)	

Note: The table shows results from instrumental variables estimation of the general regression specification (26) and the corresponding standard deviation and log-volatility regressions. The additional instrumentation of implied volatility consists of both components of realized volatility as well as lagged implied volatility. Standard errors are in parentheses and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 9: VAR models using realized volatility

Panel A: Variables in logarithmic form					
Dep. var.	Constant	$\ln RV_{t+1}$	$\ln RV_t$	$\ln IV_t$	BG
$\ln RV_{t+1}$	0.0130 (0.1786)	–	0.2512 (0.0664)	0.8058 (0.0838)	12.544
$\ln IV_{t+1}$	−0.9142 (0.1587)	0.8133 (0.2350)	–	−0.1240 (0.2540)	16.012
Panel B: Variables in std. dev. form					
Dep. var.	Constant	$RV_{t+1}^{1/2}$	$RV_t^{1/2}$	$IV_t^{1/2}$	BG
$RV_{t+1}^{1/2}$	−0.0316 (0.0084)	–	−0.0729 (0.0487)	1.1630 (0.0617)	21.007
$IV_{t+1}^{1/2}$	0.0127 (0.0583)	−1.7417 (1.6790)	–	2.5170 (1.8520)	11.691
Panel C: Variables in variance form					
Dep. var.	Constant	RV_{t+1}	RV_t	IV_t	BG
RV_{t+1}	−0.0154 (0.0020)	–	−0.1718 (0.0260)	1.5532 (0.0407)	25.873*
IV_{t+1}	0.0162 (0.0062)	−0.3524 (0.2838)	–	0.8558 (0.4189)	2.6751

Note: The table shows FIML estimation results for the simultaneous system (28) and the corresponding standard deviation and log-volatility systems. Standard errors are in parentheses and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 10: LR tests in VAR models using realized volatility

Panel A: Variables in logarithmic form			
Hypothesis	Test statistic	d.f.	p-value
$H_1 : \beta = 1$	5.3859	1	0.0203
$H_2 : \beta = 1, \alpha_1 = 0$	7.5704	2	0.0227
$H_3 : \beta = 1, \alpha_1 = 0, A_{11} = 0$	74.125	3	0.0000
Panel B: Variables in std. dev. form			
Hypothesis	Test statistic	d.f.	p-value
$H_1 : \beta = 1$	6.9592	1	0.0083
$H_2 : \beta = 1, \alpha_1 = 0$	13.839	2	0.0010
$H_3 : \beta = 1, \alpha_1 = 0, A_{11} = 0$	38.100	3	0.0000
Panel C: Variables in variance form			
Hypothesis	Test statistic	d.f.	p-value
$H_1 : \beta = 1$	131.83	1	0.0000
$H_2 : \beta = 1, \alpha_1 = 0$	131.95	2	0.0000
$H_3 : \beta = 1, \alpha_1 = 0, A_{11} = 0$	140.46	3	0.0000

Note: The table shows LR test results for the simultaneous system (28) and the corresponding standard deviation and log-volatility systems.

Table 11: VAR models using jump and continuous components

Panel A: Variables in logarithmic form							
Dep. var.	Constant	$\ln C_{t+1}$	$\ln C_t$	$\ln \tilde{J}_{t+1}$	$\ln \tilde{J}_t$	$\ln IV_t$	BG
$\ln C_{t+1}$	−0.1123 (0.1892)	–	0.2388 (0.0654)	–	−0.0086 (0.0242)	0.8202 (0.0804)	13.639
$\ln \tilde{J}_{t+1}$	−2.5833 (0.5608)	–	0.2826 (0.1938)	–	0.1801 (0.0718)	0.5126 (0.2382)	6.3450
$\ln IV_{t+1}$	−0.9275 (0.4365)	1.0382 (0.3361)	–	−0.0501 (0.1302)	–	−0.3010 (0.3083)	16.301
Panel B: Variables in std. dev. form							
Dep. var.	Constant	$C_{t+1}^{1/2}$	$C_t^{1/2}$	$J_{t+1}^{1/2}$	$J_t^{1/2}$	$IV_t^{1/2}$	BG
$C_{t+1}^{1/2}$	−0.0272 (0.0075)	–	−0.0881 (0.0513)	–	−0.0324 (0.0899)	1.1099 (0.0541)	17.896
$J_{t+1}^{1/2}$	−0.0097 (0.0064)	–	−0.0905 (0.0435)	–	0.3426 (0.0762)	0.2995 (0.0459)	34.033**
$IV_{t+1}^{1/2}$	0.0186 (0.0430)	−1.5362 (1.2490)	–	−0.4906 (0.7048)	–	2.3475 (1.3520)	11.278
Panel C: Variables in variance form							
Dep. var.	Constant	C_{t+1}	C_t	J_{t+1}	J_t	IV_t	BG
C_{t+1}	−0.0148 (0.0016)	–	−0.1561 (0.0264)	–	−0.3027 (0.1552)	1.4352 (0.0337)	19.816
J_{t+1}	−0.0005 (0.0008)	–	−0.0324 (0.0125)	–	0.2966 (0.0733)	0.1132 (0.0160)	110.07**
IV_{t+1}	0.0161 (0.0062)	−0.3513 (0.2888)	–	−0.4454 (1.1830)	–	0.8654 (0.4242)	2.7798

Note: The table shows FIML estimation results for the simultaneous system (29) and the corresponding standard deviation and log-volatility systems. Standard errors are in parentheses and BG is the Breusch-Godfrey statistic (with 12 lags) for the residuals. One and two asterisks denote rejection of the null of no serial autocorrelation at 5% and 1% significance level, respectively.

Table 12: LR tests in VAR models using jump and continuous components

Panel A: Variables in logarithmic form			
Hypothesis	Test statistic	d.f.	p-value
$\mathrm{H}_1 : \beta_1 = 1$	5.0420	1	0.0247
$\mathrm{H}_2 : \beta_1 = 1, \alpha_1 = 0$	5.1091	2	0.0777
$\mathrm{H}_3 : \beta_1 = 1, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$	117.30	4	0.0000
$\mathrm{H}_4 : \beta_1 = 1, \beta_2 = 0$	11.911	2	0.0026
$\mathrm{H}_5 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0$	11.978	3	0.0075
$\mathrm{H}_6 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$	123.67	5	0.0000
Panel B: Variables in std. dev. form			
Hypothesis	Test statistic	d.f.	p-value
$\mathrm{H}_1 : \beta_1 = 1$	4.1633	1	0.0413
$\mathrm{H}_2 : \beta_1 = 1, \alpha_1 = 0$	12.998	2	0.0015
$\mathrm{H}_3 : \beta_1 = 1, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$	76.931	4	0.0000
$\mathrm{H}_4 : \beta_1 = 1, \beta_2 = 0$	39.608	2	0.0000
$\mathrm{H}_5 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0$	48.443	3	0.0000
$\mathrm{H}_6 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$	113.78	5	0.0000
Panel C: Variables in variance form			
Hypothesis	Test statistic	d.f.	p-value
$\mathrm{H}_1 : \beta_1 = 1$	122.43	1	0.0000
$\mathrm{H}_2 : \beta_1 = 1, \alpha_1 = 0$	125.70	2	0.0000
$\mathrm{H}_3 : \beta_1 = 1, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$	143.97	4	0.0000
$\mathrm{H}_4 : \beta_1 = 1, \beta_2 = 0$	129.81	2	0.0000
$\mathrm{H}_5 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0$	133.07	3	0.0000
$\mathrm{H}_6 : \beta_1 = 1, \beta_2 = 0, \alpha_1 = 0, A_{11} = 0, A_{12} = 0$	151.34	5	0.0000

Note: The table shows LR test results for the simultaneous system (29) and the corresponding standard deviation and log-volatility systems.

Figure 1: Time series plots of volatility measures