ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Assessing the Impact of Market Microstructure Noise and Random Jumps on the Relative Forecasting Performance of Option-Implied and Returns-Based Volatility

Gael M. Martin, Andrew Reidy and Jill Wright

May 2006

Working Paper 10/06

Assessing the Impact of Market Microstructure Noise and Random Jumps on the Relative Forecasting Performance of Option-Implied and Returns-Based Volatility

Gael M. Martin, Andrew Reidy and Jill Wright*

May 26, 2006

Abstract

This paper presents a comprehensive empirical evaluation of option-implied and returns-based forecasts of volatility, in which new developments related to the impact on measured volatility of market microstructure noise and random jumps are explicitly taken into account. The option-based component of the analysis also accommodates the concept of model-free implied volatility, such that the forecasting performance of the options market is separated from the issue of misspecification of the option pricing model. The forecasting assessment is conducted using an extensive set of observations on equity and option trades for News Corporation for the 1992 to 2001 period, yielding certain clear results. According to several different criteria, the model-free implied volatility is the best performing forecast, overall, of future volatility, with this result being robust to the way in which alternative measures of future volatility accommodate microstructure noise and jumps. Of the volatility measures considered, the one which is, in turn, best forecast by the option-implied volatility is that measure which adjusts for microstructure noise, but which retains some information about random jumps.

Keywords: Volatility Forecasts; Quadratic Variation; Intraday Volatility Measures; Model-free Implied Volatility.

JEL Classifications: C10, C53, G12.

*Department of Econometrics and Business Statistics, Monash University. Corresponding author: Gael Martin, Department of Econometrics and Business Statistics, P.O. Box, 11E, Monash University, Victoria, 3800, Australia. *(Email: Gael.Martin@Buseco.monash.edu.au).* This research has been supported by Australian Research Council Discovery Grant No. DP0664121.

1 Introduction

In recent years, many studies have investigated the relative performance of option-implied and returns-based forecasts of the future volatility of an asset. Since the advent of the realized volatility literature (e.g. Barndorff-Neilsen and Shephard, 2002, Andersen *et al.,* 2003), the measurable proxy used for the unobserved asset volatility has almost exclusively been constructed from high-frequency intraday returns. The most common such measure has been based on the sum of squared returns over small, regular intervals, such as 5 or 30 minutes (e.g. Poteshman, 2000, Blair *et al.,* 2001, Koopman *et al.,* 2003, Martens and Zein, 2003, Pong *et al.,* 2004, and Jiang and Tian, 2005), with such time intervals deemed to be sufficiently small to provide an accurate estimate of volatility over the time period of interest (a day, say), whilst, at the same time, avoiding much of the bias induced by the microstructure noise present in transactions data.[1] Studies that have adopted the realized volatility proxy have produced more definitive results, overall, than earlier work which used squared (or absolute) daily returns as the volatility measure (e.g. Day and Lewis, 1995). Nevertheless, conclusions have still been mixed, with the information content of option prices sometimes deemed to be superior to (or to subsume) that of historical returns (e.g. Blair *et al.,* 2001, Jiang and Tian, 2005) and sometimes not (e.g. Martens and Zein, 2003).

The primary aim of this paper is to reassess the relative importance of option and spot prices in the prediction of future volatility by exploiting very recent developments related to the measurement of volatility in the presence of the empirical regularities of microstructure noise, including price discreteness, and random jumps. The forecasting assessments are performed using a range of measures of future volatility that are alternatives to the conventional estimator based on squared returns sampled at an arbitrarily chosen regular interval. The first three such measures are designed explicitly to cater for microstructure noise, namely: the two scales realized volatility estimator of Zhang *et al.* (2005); the realized kernel estimator of Barndorff-Neilsen *et al.* (2005); and the optimal sampling frequency estimator of Bandi and Russell (2006). As a fourth alternative, we follow the approach of Anderson and Vahid (2005), by measuring only the continuous path component of future volatility, via the bi-power variation estimator of Barndorff-Neilsen and Shepherd (2004). The bi-power calculations are robustified to microstructure noise using the approach proposed in Andersen, Bollerslev and Diebold, (2005). Finally, we pursue the method of Large (2005), whereby a consistent estimator of quadratic variation, constructed from a scaled function of the discrete price movements from transaction to transaction, is based on the assumption that prices follow a pure jump process.[2]

[1] Jiang and Tian (2005) make some adjustment to the conventional realized variance measure to accommodate autocorrelation in intraday returns; see also Andersen *et al.* (2003).

[2] In quantifying the impact on the ranking of volatility models of different proxies of the true unobservable

A secondary aim of our paper is to better model the efficiency of the options market by using the 'model free' (MF) estimate of implied volatility of Britten-Jones and Neuberger (2000) and Jiang and Tian (2005), as an alternative to the Black-Scholes (Black and Scholes, 1973) implied volatility which underlies most previous forecasting evaluations. The advantages of such an approach are two-fold. Firstly, the eschewing of a specific option price model enables a direct test of the informational content of the options market to be conducted, rather than a joint test of market efficiency *and* the validity of the option price model. In particular, avoidance of the empirically misspecified Black-Scholes (BS) model is likely to allow for a clearer assessment of the forecasting ability of the options market[3]. Secondly, as demonstrated in Jiang and Tian (2005), the particular MF volatility to be used here is an estimate of quadratic variation in the continuous and jump component of returns. Hence, in the presence of jumps in the underlying asset price process, the MF implied volatility may produce a better prediction of any measure of future volatility that itself incorporates jump information. This component of our work serves to extend the empirical analysis in Jiang and Tian (2005), in which the MF volatility measure is assessed as a predictor of one particular measure of realized volatility as constructed from regularly spaced intraday returns.

To assess the relative performance of returns- and options-based forecasts of volatility, we use both univariate and encompassing forecast regressions, in the spirit of Mincer and Zarnowitz (1969), with there being a different set of such regressions for alternative volatility proxies. All assessment is of *out-of-sample* forecasting performance, with forecasts evaluated using R^2 measures and various regression-based tests. Returns-based forecasts are produced both *directly*, via time series models for the volatility proxy itself, and *indirectly,* via generalized autoregressive conditional heteroscedastic (GARCH)-type models for daily returns. In the spirit of much of the recent literature, and as tallies with the features of our empirical data, we focus on long memory autoregressive fractionally integrated moving average (ARFIMA) models for the volatility proxy, with short memory ARMA specifications included for comparative purposes. We also consider both short memory and long memory fractionally integrated GARCH (FIGARCH) models for daily returns, as well as certain asymmetric specifications. The forecasting assessment is conducted using a comprehensive set of intraday spot and option price data for News Corporation over the ten year period from 1992 to 2001.[4]

volatility, we expand upon the theme in Hansen and Lunde (2006a). In the latter work, the conventional realized volatility estimator, as proxy, is compared with squared daily returns, with the more accurate former measure found to produce a more reliable ranking of models in simulation experiments; see also Blair *et al.* (2001). Our work is also related to that of Andersen, Bollerslev, and Meddahi (2005), in which the R^2 of regression-based evaluations of alternative forecasting models are adjusted (upwards) to cater for the error-in-variables problem associated with proxying the unobserved forecast variable with a realized volatility measure that is biased in the presence of microstructure noise.

[3] The BS model assumes that returns on the underlying asset are normal with constant variance; assumptions that conflict with virtually all empirical evidence on financial returns.

[4] The empirical work is conducted using Time Series Modelling 4.17 (www.timeseriesmodelling.com.) and Ox.

An outline of the remainder of the paper is as follows. In Section 2, we present the continuous time jump diffusion model for asset prices that underlies our analysis. Within the context of that model we present the conventional measure of realized volatility, based on regularly spaced returns. In Section 3, we then present the five alternative volatility proxies to be considered. The issues associated with forecasting (measured) volatility are addressed in Section 4, with the method for producing the MF implied volatility described. In Section 5, all aspects of the numerical application are outlined, including the empirical properties of the daily returns data, the option price data, the intraday data and the alternative volatility measures. Issues to do with re-scaling the latter to represent 24 hour measures are also addressed here. The results of the forecast evaluation are then presented and commented upon. Overall, the results provide quite strong evidence of the effectiveness of the MF implied volatility as a forecast of future volatility, and of the fact that the latter does best at forecasting a measure of volatility that is adjusted for microstructure noise, but which incorporates some jump information. Section 6 concludes.

2 Theoretical Model

Denoting by $p(t)$ the logarithm of the asset price $P(t)$ at time t, we assume a continuous time jump diffusion process,

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

where $\mu(t)$ is a continuous (locally bounded) function, $\sigma(t)$ is a strictly positive volatility process, $W(t)$ is standard Brownian motion, and $\kappa(t)dq(t)$ is a random jump process that allows for occasional jumps in $p(t)$ of size $\kappa(t)$. The *quadratic variation* (QV) for the return

$$r_{t+1} = p(t+1) - p(t) \tag{2}$$

is then given by

$$QV_{t+1} = \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s\leq t+1} \kappa^2(s). \tag{3}$$

That is, QV_{t+1} is equal to the sum of the integrated volatility of the continuous sample path component ($\int_t^{t+1} \sigma(s)ds$) and the sum of the $q(t)$ squared jumps that occur between time periods t and $t+1$. As demonstrated in Barndorff-Neilsen and Shepherd (2002) and Andersen *et al.* (2003), a consistent estimator of QV_{t+1} is provided by the sum of squared discretely sampled Δ-period returns, $r_{t,\Delta} = p(t) - p(t-\Delta)$,

$$RV_{t+1}(\Delta) = \sum_{j=1}^{1/\Delta} r_{t+j\Delta,\Delta}^2, \tag{4}$$

(www.nuff.ox.ac.uk/Users/Doornik).

where $RV_{t+1}(\Delta)$ is referred to as *realized volatility*[5]. That is, as $\Delta \to 0$,

$$RV_{t+1}(\Delta) \to \int_t^{t+1} \sigma^2(s)ds + \sum_{t<s\leq t+1} \kappa^2(s). \tag{5}$$

Three comments can be made about (5):

1. The result in (5) is contingent upon *observed* price data adhering to the model in (1). In practice, observed prices should be viewed as reflecting both the process in (1) and a process that results from market microstructure noise.

2. The sample quantity $RV_{t+1}(\Delta)$ will reflect both the continuous and jump components of the asset price process. In particular, only in the absence of jumps ($\kappa(t) = 0$) will realized volatility estimate integrated volatility alone.

3. In practice, prices are not continuous random variables, but move in discrete numbers of ticks. This discreteness can be viewed as one component of the microstructure noise referred to in Point 1.

We take up these points in Sections 3.1, 3.2 and 3.3 respectively .

3 Alternative Approaches to Realized Volatility Calculation

3.1 Realized Volatility Calculation in the Presence of Microstructure Noise

With regard to Point 1, as highlighted in Barndorff-Neilsen *et al.* (2005), Zhang *et al.* (2005) and Bandi and Russell (2006), amongst others, observed transactions data do not adhere to (1), due to a range of factors collectively referred to as *market microstructure.* That is, the true price is distorted by effects that include price discreteness, separate trading prices for buyers and sellers (the bid-ask spread), the information asymmetry of market participants, and the risk aversion of market makers. Due to the presence of such factors, the 'true' latent logarithmic price process, $p^*(t)$, may be assumed to follow (1), but is observed with error. Hence, a suitable model for the observed logarithmic price process, $p(t)$, is

$$p(t) = p^*(t) + \varepsilon(t), \tag{6}$$

where $\varepsilon(t)$ is assumed (at least initially) to be an *i.i.d.* white noise component, with variance σ_ε^2, and with $\varepsilon(t)$ independent of $p^*(t)$. Viewed in terms of the discretely sampled Δ-period returns,

[5] As is quite common in the literature, we use the term 'volatility' to refer to either a variance or a standard deviation quantity. Exactly which type of quantity is being referenced in any particular instance will be made clear by both the context and the notation.

$r_{t,\Delta}$, we have

$$\begin{aligned} r_{t,\Delta} &= p(t) - p(t-\Delta) \\ &= p^*(t) - p^*(t-\Delta) + \varepsilon(t) - \varepsilon(t-\Delta) \\ &= r^*_{t,\Delta} + \eta_{t,\Delta}. \end{aligned} \tag{7}$$

That is, observed returns ($r_{t,\Delta}$) are equal to latent returns ($r^*_{t,\Delta}$) plus a first order moving average (MA) process, $\eta_{t,\Delta}$.

In practice of course, transaction data does not occur at regular time intervals, i.e. Δ between successive transactions is not constant. To reflect this fact we introduce the notation

$$G = \{t_1, t_2, \ldots, t_i, t_{i+1}, \ldots, t_n\}, \tag{8}$$

to denote the full grid of times at which each transaction, $i = 1, 2, \ldots, n$ occurs, where n is the number of transactions in the relevant time period $[t, t+1]$ (i.e. on day t say). With an obvious use of notation, the logarithmic price associated with transaction t_i is given by

$$p(t_i) = p^*(t_i) + \varepsilon(t_i), \tag{9}$$

and the observed transaction-to-transaction return expressed as

$$\begin{aligned} r_{t_{i+1}} &= p(t_{i+1}) - p(t_i) \\ &= r^*_{t_{i+1}} + \eta_{t_{i+1}}. \end{aligned} \tag{10}$$

Realized volatility constructed from all transactions is then defined by

$$RV_{t+1}(i) = \sum_{t_i, t_{i+1} \in [t,t+1]} r^2_{t_{i+1}}. \tag{11}$$

Beginning with the expression in (9), it is straightforward to show (see Zhang *et al.* 2005) that the expectation of the realized variance estimator constructed using all n of the transaction prices observed over $[t, t+1]$, conditional on the true latent price process, is

$$E\left(RV_{t+1}(i) | \, p^*(t_i)\right) = \sum_{t_i, t_{i+1} \in [t,t+1]} r^{*2}_{t_{i+1}} + 2n\sigma^2_\varepsilon. \tag{12}$$

Remembering that $p^*(t)$ is the object assumed to follow (1), it is the quantity $\sum_{t_i, t_{i+1} \in [t,t+1]} r^{*2}_{t_{i+1}}$ that estimates the quadratic variation for $p^*(t)$, as per (4). Hence, as is clear from (12), realized volatility constructed from the *observed* returns is a biased representation of $\sum_{t_i, t_{i+1} \in [t,t+1]} r^{*2}_{t_{i+1}}$ and, hence, a biased estimator of quadratic variation. Moreover, the bias is $O(n)$, meaning that bias is proportional to the number of transactions used to construct the realized volatility measure. Defining

$$\widehat{\sigma^2_\varepsilon} = \frac{1}{2n} RV_{t+1}(i), \tag{13}$$

Zhang *et al.* (2005) also demonstrate that as $n \to \infty$,

$$n^{1/2}(\widehat{\sigma_\varepsilon^2} - \sigma_\varepsilon^2) \to N(0, E(\varepsilon^4)).$$

That is, (scaled) realized volatility constructed from observed transactions data is a consistent estimator of the variance of the microstructure noise, σ_ε^2!

3.1.1 The Two-Scale Realized Volatility (TSRV) Estimator

Given the clear deficiency of the realized volatility estimator based on all observed data, Zhang *et al.* (2005) suggest a range of modifications. Certain of these modifications bear some relationship with estimators presented in independent work of Barndorrf-Neilsen *et al.* (2005) and Bandi and Nelson (2006). These estimators are considered respectively in Section 3.1.2 and 3.1.3 below. We focus in this section only on the 'first-best' option of Zhang *et al.* (2005), which is based on a weighted difference between two estimators: 1) an average of realized volatilities calculated essentially as per (11), but over moving windows of subgrids defined on a 'slow' time scale (only observations several transactions apart are used); and 2) realized volatility calculated on a 'fast' time scale, as per (11) with all transactions used. More specifically, the full grid of observational points, G in (8) is partitioned into K nonoverlapping subgrids $G^{(k)}$, $k = 2, 3, \ldots, K$, where

$$G^{(k)} = \{t_{k-1}, t_{k-1+K}, t_{k-1+2K}, \ldots, t_{k-1+n_k K}\},$$

for some integer n_k. Realized volatility is then constructed from returns over successive time points in $G^{(k)}$, denoted by t_i and $t_{i,+}$ respectively,

$$RV_{t+1}^{(k)}(i) = \sum_{t_i, t_{i,+} \in G^{(k)}} r_{t_{i,+}}^2. \tag{14}$$

The resultant two scales realized volatility (TSRV) estimator is then defined as

$$TSRV_{t+1} = \left(\frac{n}{(K-1)\overline{n}_K}\right)\left(RV_{t+1}^{(K)}(i) - \frac{\overline{n}_K}{n} RV_{t+1}(i)\right), \tag{15}$$

where

$$RV_{t+1}^{(K)}(i) = \frac{1}{K}\sum_{k=2}^{K} RV_{t+1}^{(k)}(i), \tag{16}$$

$RV_{t+1}(i)$ is as defined in (11), $\overline{n}_K = \frac{1}{K}\sum_{k=2}^{K} n_k$ and the scale factor, $\left(\frac{n}{(K-1)\overline{n}_K}\right)$, is used to improve the performance of the estimator when K is large. Clearly $\overline{n}_K = n_k$ if we choose a constant $n_k = \dim(G^{(k)})$.

The TSRV measure is shown to be a consistent estimator of quadratic variation, in the presence of microstructure noise.[6] As can be deduced from the discussion in Ait-Sahalia *et al.*

[6] In Ait-Sahalia *et al.* (2005) various modifications are made to the estimator in (15) to render it robust to the presence of autocorrelated noise in (6). Given that we found little difference between these modified estimators and the estimator in (15), for the empirical data under study here, we use only the latter estimator in the forecasting evaluation.

(2005) regarding the robustness of the TSRV estimator to the deletion of outliers in the data, this estimator would be expected to eliminate some of the jump information in the data. That is, large returns impact to some extent on both the slow and fast time scale components of (15), and thereby cancel in the construction of the estimator. This is despite the fact that, theoretically, the estimator still converges to the sum of the continuous and discrete jump components of quadratic variation, as per (5).

Zhang *et al.* (2005) derive the optimal value for K for the period t to $t+1$, as

$$K = cn^{2/3}, \tag{17}$$

where $c = \left(16\sigma_\varepsilon^4 / TE\left(\eta^2\right)\right)^{1/3}$ and $\eta^2 = \frac{4}{3}\int_t^{t+1}\sigma^4(s)ds$. The term σ_ε^4 is square of the variance of the noise, while $\int_t^{t+1}\sigma^4(s)ds$ is the integrated quarticity. σ_ε^2 is estimated as in (13) and $E(\eta^2)$ estimated as $\widehat{E(\eta^2)} = \frac{4}{3}\left[RV_{t+1}(\Delta)\right]^2$ for some reasonably large Δ; see Barndorff-Neilsen *et al.* (2006). In the empirical exercise we use $\Delta \approx 30$ minutes.

3.1.2 The Realized Kernel (RKERN) Estimator

Barndorff-Neilsen *et al.* (2005) develop kernel estimators of the quadratic variation, with the weights used in constructing the kernel chosen to ensure that the resultant estimator is consistent in the presence of microstructure noise[7]. Estimators that assume both regularly and irregularly spaced data are derived. We focus here on the latter type of estimator, with returns measured in transaction time, rather than calendar time, as is consistent with the returns underlying the TSRV estimator of Zhang *et al.* (2005).

Consistent with the definition of $RV_{t+1}^{(k)}(i)$ in (14), (although with a slight abuse of notation), we define

$$RCV_{t+1}^{(k)}(i,h) = \sum_{t_i, t_{i,+}, t_{i+h}, t_{i+h,+} \in G^{(k)}} r_{t_{i,+}} r_{t_{i+h,+}}, \quad h = -H, \ldots, -1, 0, 1, 2, \ldots H,$$

as the realized covariance function constructed from returns observed over pairs of successive time points in $G^{(k)}$, $k = 2, 3, \ldots, K$, with the returns being h time points apart.[8] When $h = 0$, we regain the variance quantity, $RV_{t+1}^{(k)}(i)$. The averaged (over k) version of $RCV_{t+1}^{(k)}(i,h)$ is then given by

$$RCV_{t+1}^{(K)}(i,h) = \frac{1}{K}\sum_{k=2}^{K} RCV_{t+1}^{(k)}(i,h),$$

analogously with the averaged version of $RV_{t+1}^{(k)}(i)$, $RV_{t+1}^{(K)}(i)$, in (16).

[7] Although the kernel estimator is introduced within the context of general semimartingales, the properties of the estimator are demonstrated under the assumption of a model without random jumps (i.e. with $\kappa(t) = 0$ in (6)).

[8] The notation $r_{t_{i+h,+}}$ denotes the return over successive time-points in the sub-grid $G^{(k)}$, where that return is h time points distant from $r_{t_{i,+}}$ according to the sub-grid $G^{(k)}$, not the full grid G in (8).

A symmetric version of the realized kernel (RKERN) estimator is given by

$$\begin{aligned} RKERN_{t+1} &= \sum_{h=-H}^{H} w_h RCV_{t+1}^{(K)}(i,h) \\ &= w_0 RV_{t+1}^{(K)}(i) \\ &\quad + \sum_{h=1}^{H} w_h \left\{ RCV_{t+1}^{(K)}(i,h) + RCV_{t+1}^{(K)}(i,-h) \right\}, \end{aligned} \tag{18}$$

with weights

$$w_0 = w_1 = 1 \tag{19}$$

$$w_h = \frac{(H+2h)(H-h+1)(H-h+2)}{H(H+1)(H+2)}, \quad h = 2,3,\ldots,H. \tag{20}$$

and H is the closest integer to

$$H = 3.6867\sqrt{\frac{\widehat{\sigma_\varepsilon^2}}{\widehat{\int_t^{t+1}\sigma^2(s)ds}}n}. \tag{21}$$

The weights in (19) ensure that the kernel is asymptotically unbiased, with inclusion of the additional terms in the kernel ($h = 2,3,\ldots,H$) serving to reduce the variance. The value of H in (21) (approximately) minimizes the asymptotic variance of the estimator. The estimates of the noise variance (σ_ε^2) and integrated volatility used in the construction of H are respectively $\widehat{\sigma_\varepsilon^2}$ as defined in (13) and $\widehat{\int_t^{t+1}\sigma^2(s)ds} = RV_{t+1}(\Delta)$, with $RV_{t+1}(\Delta)$ as defined in (4), for fairly large Δ ($\Delta \approx 30$ minutes in the empirical application).

3.1.3 The Optimally Sampled Realized Volatility (OSRV) Estimator

Motivated by the relative computational simplicity of the conventional realized volatility estimator in (4), Bandi and Russell (2006) propose an estimator,

$$OSRV_{t+1} = \textstyle\sum_{j=1}^{M_{t+1}} r_{t+j\delta,\delta}^2, \tag{22}$$

based on M_{t+1} discretely sampled δ-period returns, $r_{t,\delta} = p(t) - p(t-\delta)$, where the sampling frequency, $\delta_{t+1} = 1/M_{t+1}$, is chosen to minimize the mean squared error (MSE). We refer to (22) as the optimally sampled realized volatility (OSRV) estimator. Under certain conditions[9], the MSE is shown to be a function of M_{t+1}, the second and fourth moments of the noise process, the integrated variance, $\int_t^{t+1}\sigma^2(s)ds$, and the integrated quarticity, $\int_t^{t+1}\sigma^4(s)ds$. Given sample estimates of all population moments, M_{t+1} is chosen so as to minimize MSE where, as indicated by the notation, M_{t+1} (and, hence, δ_{t+1}) varies with t. When the optimal sampling frequency is high, the following approximation can be used for the optimal value of M_{t+1},

$$M_{t+1}^* \sim \left(\frac{\widehat{Q}_{t+1}}{\widehat{\sigma}_\varepsilon^4}\right)^{1/3}, \tag{23}$$

[9] In particular, with reference to (1), it is assumed that $\mu(t) = \kappa(t) = 0$.

where $\widehat{Q}_{t+1} = \frac{1}{3\Delta}\sum_{j=1}^{1/\Delta} r_{t+j\Delta,\Delta}^4$, is an estimate of the integrated quarticity based upon some relatively large time interval ($\Delta \approx 30$ minutes in the empirical example). The term in the denominator of (23) is given by $\widehat{\sigma}_\varepsilon^4 = \left(\frac{1}{NM^*}\sum_{t=1}^{N}\sum_{j=1}^{M^*} r_{t+j\delta^*,\delta^*}^2\right)^2$, and is an estimate of the squared second moment of the noise in (6), with $\delta^* = 1/M^*$ equal to the highest optimal sampling frequency over the N time periods (days say) in the sample. From (23) it is clear that returns on day t are to be sampled less frequently (M_{t+1}^* is smaller), the larger is the squared variance of the noise in the data relative to the quarticity of the underlying efficient price process.[10]

3.2 Realized Bi-Power Variation

With regard to Point 2 in Section 2, Barndorff-Neilsen and Shepherd (2004) focus on the separate identification and estimation of integrated volatility, exclusive of jumps. Defining *realized bi-power variation* as

$$BPV_{t+1}(\Delta) = \frac{\pi}{2}\sum_{j=2}^{1/\Delta} |r_{t+j\Delta,\Delta}|\left|r_{t+(j-1)\Delta,\Delta}\right|, \tag{24}$$

they show that as $\Delta \rightarrow 0$,

$$BPV_{t+1}(\Delta) \xrightarrow{p} \int_t^{t+1}\sigma(s)ds, \tag{25}$$

i.e. that realized bi-power variation consistently estimates the integrated variance of the continuous sample path component of the price process in (1). Using (5) and (25), it also follows that

$$RV_{t+1}(\Delta) - BPV_{t+1}(\Delta) \xrightarrow{p} \sum_{t<s\leq t+1}\kappa^2(s), \tag{26}$$

with tests for jumps being based on various standardized statistics constructed from (26); see also Andersen, Bollerslev and Deibold (2005) and Huang and Tauchen (2005).

Analogous to the realized volatility estimator in (4), for very small Δ the statistic in (25) is adversely affected by the presence of microstructure noise. Moreover, given the assumption of independent noise, the implied MA(1) structure for $\eta_{t,\Delta}$ in (7) means that the two adjacent observed returns in (24) will be autocorrelated, resulting, in turn, in a source of bias in addition to that present in realized variance. To offset this bias, Andersen, Bollerslev and Deibold (2005) and Huang and Tauchen (2005) propose a modification of (24), whereby the sum of absolute adjacent returns is replaced with the sum of the corresponding one-period staggered returns, as follows

$$BV_{t+1}(\Delta) = \frac{\pi}{2}(1-2\Delta)^{-1}\sum_{j=3}^{1/\Delta} |r_{t+j\Delta,\Delta}|\left|r_{t+(j-2)\Delta,\Delta}\right|, \tag{27}$$

where the additional term in front of the sum reflects the loss of two observations due to the staggering.[11] In the empirical section we implement an averaged version of (27), based on transaction sampling,

[10] For related work, based on the assumption of a pure jump process for the asset price, see Oomen (2006).

[11] We reserve the acronym BV for the noise-adjusted measure of bi-power variation that we use in the empirical analysis.

$$BV_{t+1}^{(K)}(i) = \frac{1}{K}\sum_{k=2}^{K} BV_{t+1}^{(k)}(i), \tag{28}$$

where $BV_{t+1}^{(k)}(i) = \frac{\pi n}{2n-4k} \sum_{t_i, t_{i,+}, t_{i+2}, t_{i+2,+} \in G^{(k)}} \left|r_{t_{i,+}}\right| \left|r_{t_{i+2,+}}\right|$, K is determined as per (17) above, and all subscript notation is consistent with that defined in earlier sections.

A-priori one would anticipate that direct forecasts of this measure, using historical observations on it, would be more accurate than corresponding forecasts of the various realized volatility measures, which are, to some extent, influenced by the random jump component. This is the rationale underlying the forecasting exercise in Anderson and Vahid (2005). On the other hand, indirect forecasts of bi-power variation, to the extent that such forecasts themselves incorporate jump information, may be less accurate than the corresponding indirect forecasts of the realized volatility measures. These issues are investigated in Section 5.

3.3 Realized Volatility for Discrete Prices

In the spirit of Point 3 in Section 2, Large (2005) proposes an estimator of quadratic variation that focusses on the number and direction of price changes during the day, rather than the magnitude of such changes, as measured by intraday returns. The estimator, which we refer to as the 'alternation' estimator, is given by

$$ALT_{t+1} = n^{(ch)} tick^2 \frac{C}{A}, \tag{29}$$

where $n^{(ch)} \in \mathbb{N}$ is the number of price changes in a day and *tick* is the price tick (i.e. the minimum amount by which the price can change on the relevant exchange). Defining an alternation as a price change that occurs in the opposite direction to the previous price change, and a continuation as a price change in the same direction, A then denotes the number of alternations and C the number of continuations, with $A + C = n^{(ch)}$.[12]

Without the presence of microstructure noise, the estimator $n^{(ch)} tick^2$ is a consistent estimator of quadratic variation, whilst in the presence of noise the value of $n^{(ch)} tick^2$ is asymptotically biased. Given that the presence of noise implies an excess of alternations, multiplication by the fraction C/A produces a consistent estimator in the presence of noise.

The modified version of the alternation estimator that we apply in the empirical investigation (see also Barndorff-Neilsen and Shephard, 2005), and which we denote by the acronym ALTM, is given by

$$ALTM_{t+1} = RV_{t+1}^{(K)}(i) \frac{C}{A}, \tag{30}$$

which is simply the (averaged) realized volatility measure in (16) multiplied by C/A in order to correct for the upward bias induced by the noise.

[12] The first price of the day is defined as an alternation.

4 Forecasting Volatility

4.1 Overview

Forecasts of future asset price volatility are required for various financial decisions, such as portfolio allocation, option pricing and value at risk calculation. Prior to the advent of the recent realized volatility literature, such forecasts would be produced via simple historical standard deviations ('historical' volatility), times series models constructed from daily returns data (e.g. GARCH-type models; exponentially weighted deterministic models such as the 'RiskMetrics' model; stochastic volatility models) or via implied volatilities, usually produced via the BS option pricing model. The relative worth of each competing forecast would then be measured in terms of the accuracy with which it predicted some measurable proxy for future unobserved volatility, based on low frequency (e.g. daily, weekly) returns data; see Day and Lewis (1995), amongst others.

Since the advent of the realized volatility literature, not only has the measurable proxy used for volatility changed, now being based on intraday day returns, but the focus has also shifted to the construction of standard time series models for such proxies, and the production of forecasts *directly* from these models. In particular, the stylized empirical properties of the (logarithmic) realized volatility measures are such that long-memory Gaussian ARFIMA models for this (transformation of) realized volatility have become the mainstay of empirical work. As such, the interest is now in the merit of these *direct* forecasts of some proxy of future volatility, compared with *indirect* forecasts based on low-frequency (usually daily) returns, in particular returns produced via the ubiquitous GARCH-type specifications. Such returns-based specifications are then compared with forecasts from the options market, with the informational efficiency of the latter thereby assessed.

In this paper, four new key questions are addressed regarding the relative performance of different volatility forecasts:

1. Which forecasting method (direct, indirect, option-implied) performs best overall when the assessment takes into account the alternative measures of volatility outlined above?

2. Is the ranking of the different forecasting methods robust to the measure of volatility used in the analysis, both as forecast variable and as the basis for the direct forecast? In other words, is the ranking robust to the different ways in which the alternative volatility measures accommodate microstructure noise and jumps?

3. Is the performance of the MF implied volatility forecast superior to that of the BS implied volatility forecast, again in the context of alternative approaches to volatility measurement?

4. What light do the results shed, if any, on the accuracy of option-based forecasts in the presence of microstructure noise, and on the way in which the options market factor in jumps?

The precise specifications of the ARFIMA and GARCH-type specifications used to produce the returns-based forecasts of volatility are determined by the properties of the data used in the empirical analysis, with discussion of those models deferred to Section 5 as a result. The way in which the option pricing model of Britten-Jones and Neuberger (2000) and Jiang and Tian (2005) model is used to produce MF implied volatility forecasts is outlined in the following section.

4.2 Model-Free Implied Volatility

A European call option is an asset that gives the owner the right to buy the underlying asset (or 'exercise the option') at a future point in time, T, at a pre-specified exercise or strike price K. The price of the option is thus dependent on the expected future price of the underlying asset which, in turn, is dependent on the assumed generating process for that underlying asset price. The BS option price model assumes that the asset price, $P(t)$, follows a geometric Brownian motion process with constant diffusion parameter σ. In this case, the expectation that defines the BS option price has the solution,

$$BS(\sigma) = P_t\Phi(d_1) - K^{-i_t\tau}\Phi(d_2), \tag{31}$$

where $d_1 = \left(\ln(P_t/K) + \left(i_t + 0.5\sigma^2\right)\tau\right)/\sigma\sqrt{\tau}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, $P_t =$ the (dividend discounted) spot price at time t, $K =$ the strike price, $i_t =$the (annualized) risk free rate of return at time t, $\tau = T-t =$ the time to maturity (expressed as a proportion of a year) and $\Phi(.) =$ the cumulative normal distribution. An observed market option price at time t for a call option with maturity T and strike K, $C(T,K)$, can be used to produce an estimate of σ *implied* by $C(T,K)$, by equating $C(T,K)$ to the right-hand-side of (31) and solving for σ. Given the joint assumptions that the BS model is valid and that the option market is efficient, the option-implied volatility estimate should subsume all information in historical volatility in terms of predicting future volatility.

As is now standard knowledge in the empirical finance literature, neither asset returns, nor market option prices adhere to the BS specifications, with stylized 'smile' and 'skew' patterns in implied volatilities across the strike price (or 'moneyness') spectrum being viewed as one manifestation of the misspecification of the model. Despite the large amount of attention devoted to producing alternative option price formulae that cater for the standard empirical features of asset returns (e.g. Heston, 1993, Bakshi *et al.,* 1997, Corrado and Su, 1997, Bates, 2000, Heston and Nandi, 2000, Lim *et al.*, 2005), it is the BS option formula that still underlies the implied volatilities used in many assessments of the relative performance of option-implied and returns-based volatility forecasts. As a consequence, it could be argued that these results are subject to model misspecification errors, and do not necessarily give a clear indication of the quality of information present in the options market.

With this misspecification issue in mind, we adopt the MF approach to implied volatility calculation of Britten-Jones and Neubeger (2000) and Jiang and Tian (2005). As demonstrated by these authors, under the assumption of a diffusion process for the spot price, $P(t)$, a (risk-neutral) forecast of integrated variance for the period t to T can be determined from observed call option prices with maturity T as follows

$$E_t\left[\int_t^T \sigma^2(s)ds\right] = 2\int_0^\infty \frac{C_t(T,K)e^{i_t\tau} - \max\left[0, P_t e^{i_t\tau} - K\right]}{K^2}dK, \quad (32)$$

where all notation is as defined above. The calculation in (32) invokes no specific assumptions about the spot price process. Given a finite number of strike prices, with maximum and minimum values $K_{\max}$ and $K_{\min}$ respectively, (32) is estimated as

$$\begin{aligned} E_t\left[\int_t^T \sigma^2(s)ds\right] &\approx 2\int_{K_{\min}}^{K_{\max}} \frac{C_t(T,K)e^{i_t(T-t)} - \max\left[0, P_t e^{i_t(T-t)} - K\right]}{K^2}dK \\ &\approx \sum_{j=1}^{M}\left[g(T,K_j) + g(T,K_{j-1})\right]\Delta K, \end{aligned} \quad (33)$$

where $\Delta K = (K_{\max} - K_{\min})/M$, $K_j = K_{\min} + j\Delta K$ for $0 \leq j \leq M$ and $g(T,K_j) = (C_t(T,K_j)e^{i_t(T-t)}$ $- \max\left[0, P_t e^{i_t(T-t)} - K_j\right])/K_j^2$. Given the limited number of strikes (and hence option prices) that occur in any empirical setting, a curve-fitting method is used to interpolate between the observed strikes. The procedure adopted follows that of Jiang and Tian (2005), with steps as follows: 1) Use observed call option prices (for available strikes) to produce implied BS volatilities, via (31); 2) Fit a smooth function to the implied volatilities and use this function to extract implied volatilities at grid points K_j; 3) Use the BS model in (31) to translate the K_j into 'observed' prices $C_t(T,K_j)$; 4) Use the full set of M K_j and $C_t(T,K_j)$ values to estimate integrated volatility as in (33).[13] The implied volatility extracted from option prices observed at time t, $C_t(T,K_j)$, represents the market's estimate of volatility over the maturity period t to T.[14] In the forecasting exercise, in order to avoid to so-called 'telescoping' problem (see Christensen *et al.*, 2001) we artificially construct options that always have an expiry of approximately 22 trading

[13] As pointed out by Jiang and Tian (2005), the BS model is simply being used as a mechanism to produce (artificially) a larger range of strike prices than is available in practice. The curve fitting procedure followed here does not require the BS model to be the 'true' model underlying the observed prices.

[14] Following Jiang and Tian (2005) we do not use options which expire within one week. We also only use observed prices of options with a moneyness between 0.9 and 1.1. The moneyness of an option is a function of the difference between the strike price and the spot price at time t. Broadly, options are said to be out-of-the-money if $P_t < K$, in-the-money if $P_t > K$ and at-(or near-) the-money if $P_t \approx K$.

days ahead.[15]

As demonstrated in Jiang and Tian (2005), the result in (32) can be extended to jump-diffusion processes, in which case the method produces a forecast of quadratic variation. That is, in the case where the true latent price follows the model in (1), the implied variance is an estimate of (3), rather than an estimate of integrated volatility only.

5 Empirical Analysis Using Australian Stock Market Data

5.1 Introduction

The numerical analysis is performed using data on equity and option trades for the Australian listed company, News Corporation (Newscorp) over the ten year period from 2 January, 1992 to 28 December 2001.[16] Rolling one step ahead forecasts are produced for the period 9 January, 1997 to 28 December 2001, meaning that the one step ahead forecast regressions are estimated using 1249 observations. Ten and 22 steps ahead forecasts are produced over the same period, with the associated forecast regressions based on 1239 and 1227 observations respectively. Each returns-based forecast is produced using both daily and intraday observations from $N = 1000$ days. The first year of observations (2 January to 30 December, 1992) is used to set pre-sample values in the estimation of the long-memory models. Each option-implied forecast is based on option prices observed on the day immediately prior to the forecast day (or period).

In Section 5.2 we present a brief descriptive analysis of all relevant empirical features of the data (both equity and option data), followed by a documentation of the key empirical features of the alternative volatility measures in Section 5.3. Section 5.4 then reports all results regarding the evaluation of those alternative forecasts, with the key research questions outlined earlier addressed.

[15] To do this we take the set of options that are the closest and second closest to expiry. For example the closest expiry date may be $t+18$, while the second closest may be $t+40$. In step 1) above, the BS implied volatilites are calculated for all options expiring on $t+18$ (with moneyness between 0.9 and 1.1). A parabola is then fitted to these BS implied volatilities as per step 2). The same two steps are performed for the set of options expiring on $t+40$. These two BS implied volatility curves are then linearly interpolated between to obtain a implied volatility curve for day $t+22$. Outside of the moneyness interval, 0.9 to 1.1, we construct extrapolated call prices using the BS model with an implied volatility set equal to the implied volatility at moneyness of 0.9 or 1.1 depending on which is the closest.

[16] Newscorp is one of the world's largest media conglomerates with USD $23.859 billion in revenue in 2005. In Australia, Newscorp owns 17 newspapers, however most of its assets are located overseas with the most well known being the Fox Broadcasting Corporation. In 2004, Newscorp was re-incorporated in the US. It now trades on the New York Stock Exchange, the London Stock Exchange and the ASX.

5.2 Empirical Features

5.2.1 Daily Returns Data

Panel A of Figure 1 shows the daily (close to close) returns over the 1992 to 2001 period, adjusted for share splits and dividends, and filtered as described in Appendix A.[17] The kurtosis, skewness and Jarque-Bera statistics are respectively 11.224, 0.467 and 7150.730, indicating significant excess kurtosis and (positive) skewness, using 5% asymptotic critical values. Panel B of Figure 1 is the autocorrelation function (ACF) of squared daily returns, which is a proxy for daily volatility. The dotted line in the figure is the asymptotic upper 95% confidence bound. As is typical for squared returns, the ACF decays relatively slowly, with significant correlation remaining after 80 lags.

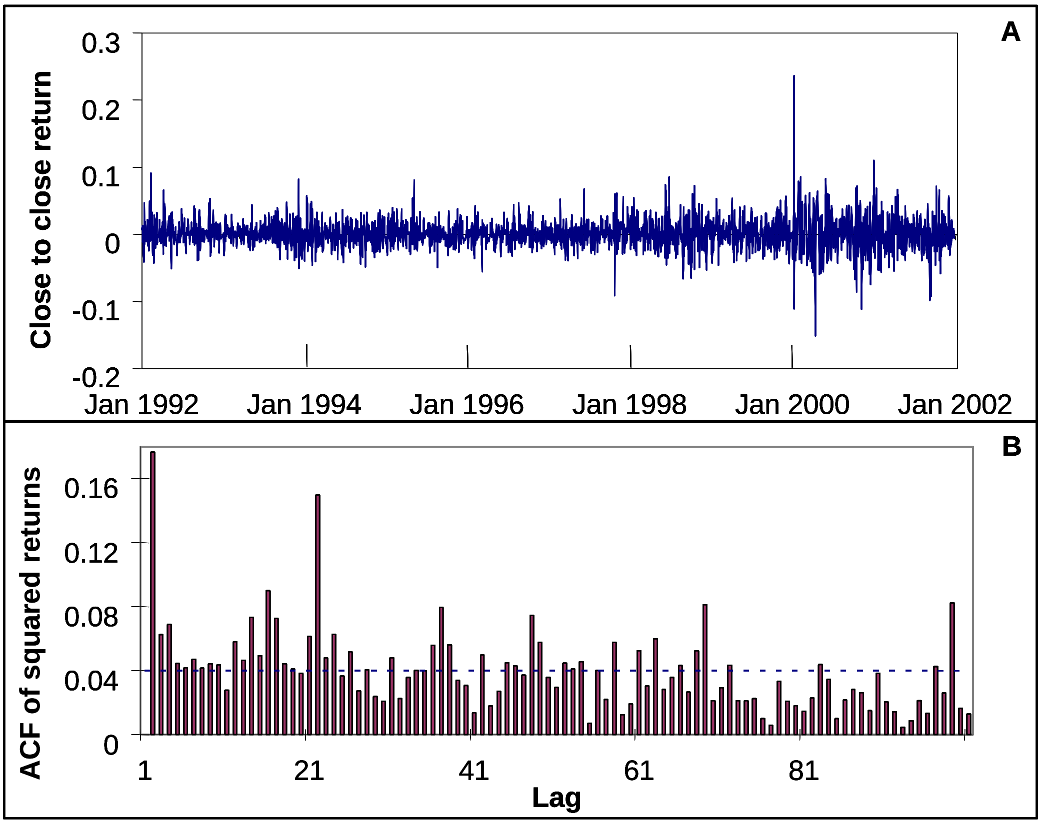

Figure 1: Panel A: close to close daily returns for Newscorp over the period 2 January 1992 to 28 December 2001. Panel B: autocorrelation function (ACF) of squared daily returns over this period. The dotted line is the upper 95% confidence bound.

In order to cater for the empirical features of the daily returns data, preliminary analysis

[17] The largest one day increase in the share price occurs on 11 January 2000, at the time of the merger of America Online and Time Warner (Collins, 2000a). The largest one day decrease occurs on 17 April 2000, associated with a large decrease on the NASDAQ market (Collins 2000b).

focussed on a range of GARCH-type specifications with leptokurtic conditional distributions. Given $r_{t+1} = \mu + \varepsilon_{t+1} = \mu + \sigma_{t+1} e_{t+1}$, where r_{t+1} denotes the daily return in (2), μ the mean daily return, σ_{t+1}^2 the daily variance and $e_{t+1} \sim Student\ t(0, 1, \nu)$, model selection criteria and significance tests were used to choose the following GARCH, threshold GARCH (TGARCH), asymmetric power ARCH (APARCH)) and fractionally integrated GARCH (FIGARCH)) models for use in the forecasting exercise:

$$\begin{aligned} GARCH(1,1) &: \quad \sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \\ TGARCH(1,1) &: \quad \sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \alpha\gamma s_{t+1}\varepsilon_t^2 + \beta\sigma_t^2 \\ APARCH(1,1) &: \quad \sigma_{t+1}^\delta = \omega + \alpha\left|\varepsilon_t\right|^\delta - \alpha\gamma^* s_{t+1}\left|\varepsilon_t\right|^\delta + \beta\sigma_t^\delta \\ FIGARCH(1,d,1) &: \quad (1-L)^d(1-\alpha L)\varepsilon_{t+1}^2 = \omega + (1-\beta L)\sigma_{t+1}^2. \end{aligned} \tag{34}$$

The notation L is used to denote the lag operator, $d > -1$ is the fractional parameter, $(1-L)^d = \sum_{j=0}^{\infty} b_j L^j$, with $b_0 = 1$ and $b_j = \frac{-d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)}$, and the remaining parameters satisfy the usual restrictions. In the asymmetric models (TGARCH and APARCH) $s_{t+1} = 1$ if $\varepsilon_t < 0$ and 0 otherwise, with the APARCH model nesting the TGARCH model when $\gamma^* = -\gamma$ and $\delta = 2$.

All models are estimated using conditional maximum likelihood, with the infinite lag structure in the FIGARCH model truncated at the lag determined by the number of sample observations plus the number of pre-sample observations. For the rolling samples, the persistence of the GARCH model $(\alpha + \beta)$ varies from 0.820 and 0.997 and the degrees of freedom for the conditional Student t distribution (ν) varies from 4.9 to 12, with similar results for the other models. In the FIGARCH model the long memory parameter (d) varies from 0.09 to 0.39 over the rolling period. For the asymmetric models (TGARCH and APARCH), the estimate of the asymmetry parameters (γ and γ^*) are found to be insignificant at the 5% level during the model selection period (1993-1996), but significant for some of the rolling samples used to produce the forecasts, ranging from about 0.1 to 2 (in magnitude).

5.2.2 Option Price Data

Options transaction data for Newscorp were obtained from the ASX for the period 8 January 1997 to 27 December 2001. The forecasting analysis is based only on transactions that occur within the last hour of trading (3pm to 4pm) for each trading day during this period. In this way, the implied volatility estimates produced from the option price data can be viewed as a forecast of volatility over the next day(s). The average number of option prices used in the calculation of implied volatility estimates on each day is 38.

As the spot market for Newscorp stock is very liquid, for each option trade it is possible to obtain a virtually simultaneous equity price: usually recorded within a few seconds of the option trade. When several equity trades are recorded at exactly the same time, a weighted average is taken, with the weights determined by the trading volume. With reference to the option price formulae in (31) and (33), i_t is equated to the three month bond rate on day t.[18] The dividends paid on Newscorp shares average about 0.1% of share value, and are paid six-monthly. The impact of dividends on share prices is therefore so small that they have only been taken into account as a constant continuous discount factor, with P_t in (31) and (33) replaced by $P_t e^{-d(T-t)}$, where d is the average dividend rate over the sample period.[19]

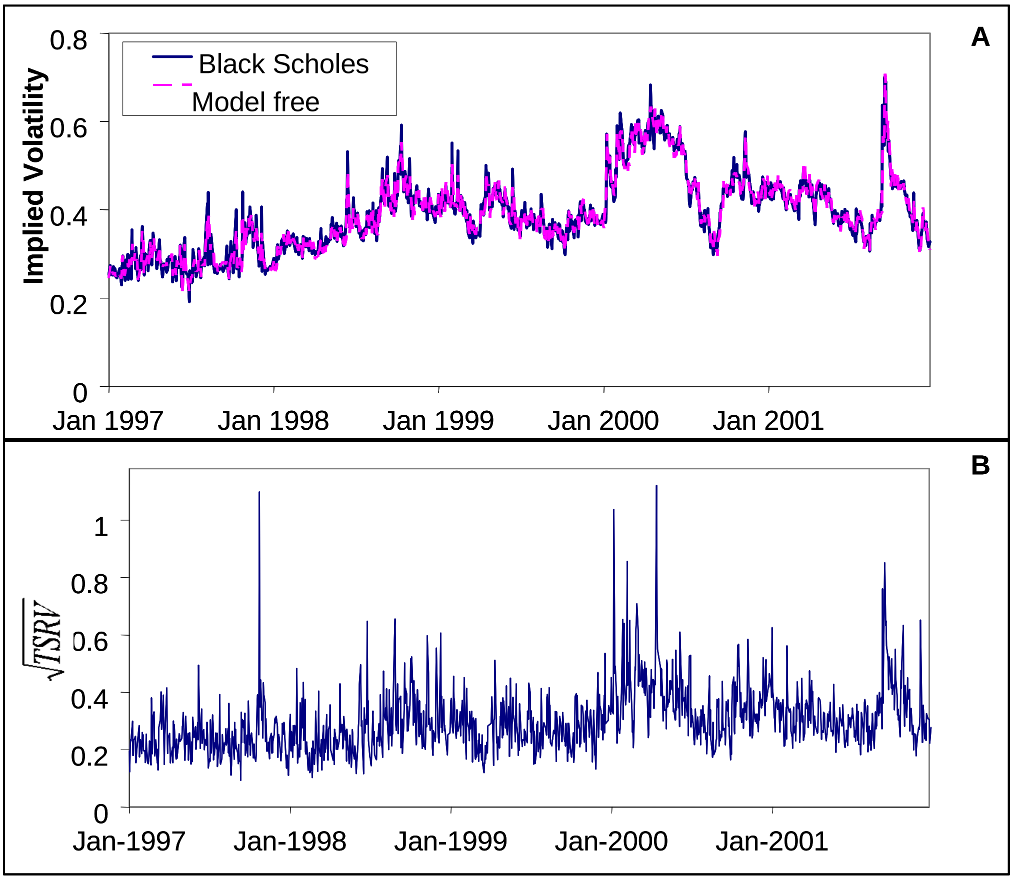

Figure 2: Panel A: one step ahead (annualized) implied volatility forecasts for the period 8 January 1997 to 27 December 2001 calculated using the BS and MF specifications. Panel B: next day (annualized) $\sqrt{TSRV}$ calculated over this period.

Implied volatility estimates are produced via both the MF approach described in Section 4.2 and the BS model, (31), for the purpose of comparison. We adopt an approach to BS implied volatility calculation that is fairly representative of that adopted by others. That is, four close-

[18] The interest rate data has been obtained from the Reserve Bank of Australia website: www.rba.gov.au.

[19] As are most stock options traded on the ASX, Newscorp (NCP) options are American options. However, as noted, dividends paid on this particular stock are infrequent and negligible in magnitude. In this case, the European formulae used to produce the implied volatility estimates, with the current spot price discounted as described in the text, are appropriate; see Hull (2000).

to-the-money options (one put and one call above and below the money) with maturity as close as possible to the forecast horizon (but with at least one week to maturity), are selected from the last hour of trading for each day.[20] The implied volatility is then calculated by minimizing the sum of squared percentage deviations of the observed prices from the BS option price in (31).[21]

Panel A of Figure 2 shows the one step ahead (annualized) implied volatility forecasts calculated over 8 January 1997 to 27 December 2001 using both option specifications. Both series track each other fairly closely, but with the MF volatility series being slightly smoother than the BS series. This extra smoothness is perhaps to be expected, given the extra degree of averaging (across both strikes and maturities) that occurs in the computation of the MF volatility. Pre-empting the forecasting assessment in Section 5.4 we show in Panel B the next-day volatility, calculated using the TSRV measure in (15). The measure is re-weighted in the manner to be described in Section 5.3.4, then presented in annualized standard deviation form. It can be seen that both implied volatility predictions track the $\sqrt{TSRV}$ series closely, with the implied volatility series much smoother because they represent the market's forecasts of volatility over longer time horizons than one day. Note also that both implied volatility series are biased upward as forecasts of the next-day $\sqrt{TSRV}$ measure, which is expected due to the implicit risk premium factored into option prices; see, for example, Guo (1998).

5.2.3 Intraday Data

The realized volatility and bi-power measures are constructed using the prices of all intraday transactions, including successive transactions for which there is no price change.[22] Panel A of Figure 3 plots the transaction-to-transaction returns for Newscorp for 1 August 2001 from 10am to 4pm. According to the model in (10), the observed return is a sum of the return in the latent price process and the noise process, where the latter has variance estimated consistently by the estimator in (13). It can be seen that there are many small returns which would be noise induced by the bid-ask bounce alone. Panel B of Figure 3 is a histogram showing the estimate of the standard deviation of the noise, as calculated using (13), for the period 2 January 1992 to 28 December 2001. The magnitude of these numbers indicates that the transaction data contains a substantial amount of noise that needs to be accounted for when estimating the volatility of the efficient price. Over the sample period, the noise to signal ratio, estimated as

[20] When more than one option satifies the criteria for a particular option category, the option which is both traded closest to 4pm and closest to the money is selected.

[21] As noted in Section 4.2, the MF implied volatilities do not suffer fom the so-called 'telescoping' problem, whereby the maturity with which the implied volatility forecast is associated changes with t. The BS forecasts will suffer from this problem to some extent.

[22] The only exception to this is in the calculation of the noise variance estimates, in which case we use only only non-zero transactions (referred to as 'tick' sampling by Griffin and Oomen, 2006).

$\sqrt{\widehat{\sigma_\varepsilon^2}}/\sqrt{TSRV}$, decays consistently from approximately 0.15 in 1992 to approximately 0.04 in 2001[23]. These figures correspond respectively to the 'large noise' and 'moderate noise' cases in Barndorff-Neilsen *et al.* (2005). The mean value of $\sqrt{\widehat{\sigma_\varepsilon^2}}$, 0.0008, is also very similar to the mean value of the corresponding noise estimate for the S&P100 over one month in 2002, 0.0007, as reported in Bandi and Russel (2006).

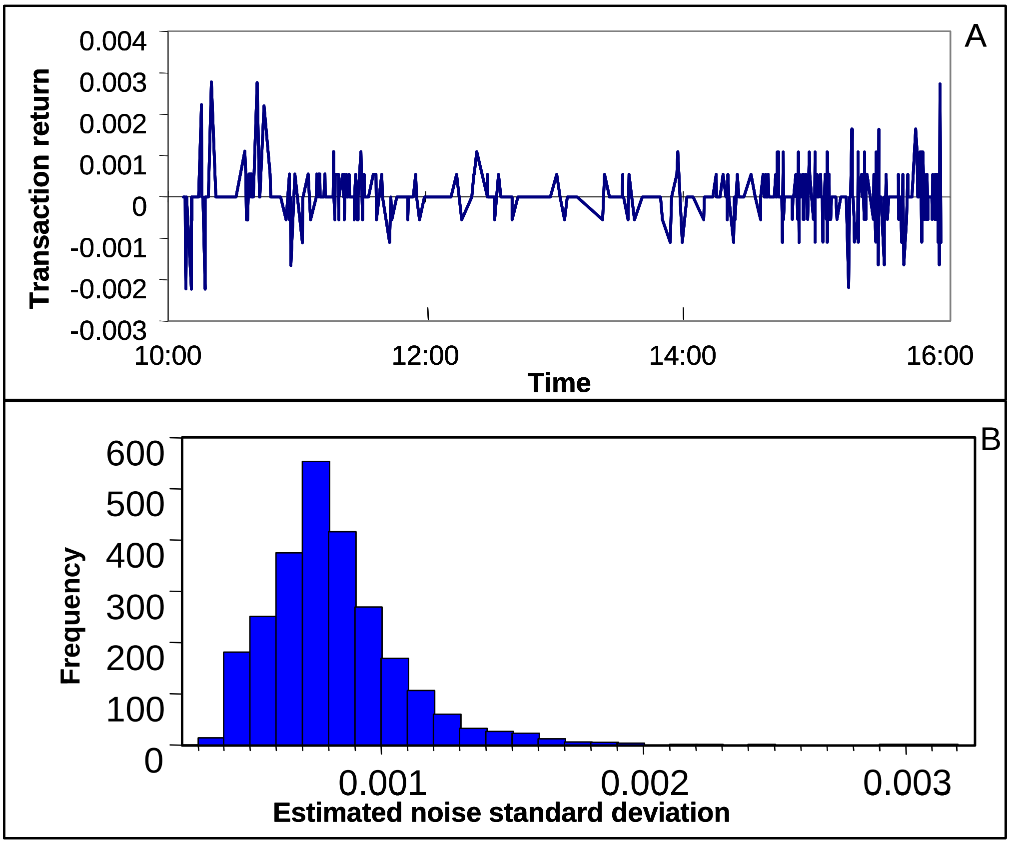

Figure 3: Panel A: intraday transaction to transaction return for 1 August 2001 from 10am to 4pm. Panel B: histogram of the estimated standard deviation of the noise $\left(\sqrt{\widehat{\sigma_\varepsilon^2}}\right)$ for each day over the period 2 January 1992 to 28 December 2001.

5.3 Alternative Volatility Measures

5.3.1 Signature Plots

Signature plots graph realized volatility and bi-power measures across a range of different sampling frequencies; see, for example, Andersen *et al.* (2000). At each sampling frequency, the variance measure is an average of the measure calculated at that frequency, on each day in the

[23]This decline is in part due to the much higher liquidity in the later part of the sample, which causes, in turn, a reduction in the bid-ask spread.

sample. Signature plots provide a way of visualizing bias problems with realized volatility and bi-power measures. An estimator that is robust to microstructure noise should produce an approximately flat signature plot. Although the focus of our analysis in on predictive performance, robustness according to the signature plots remains an important consideration.

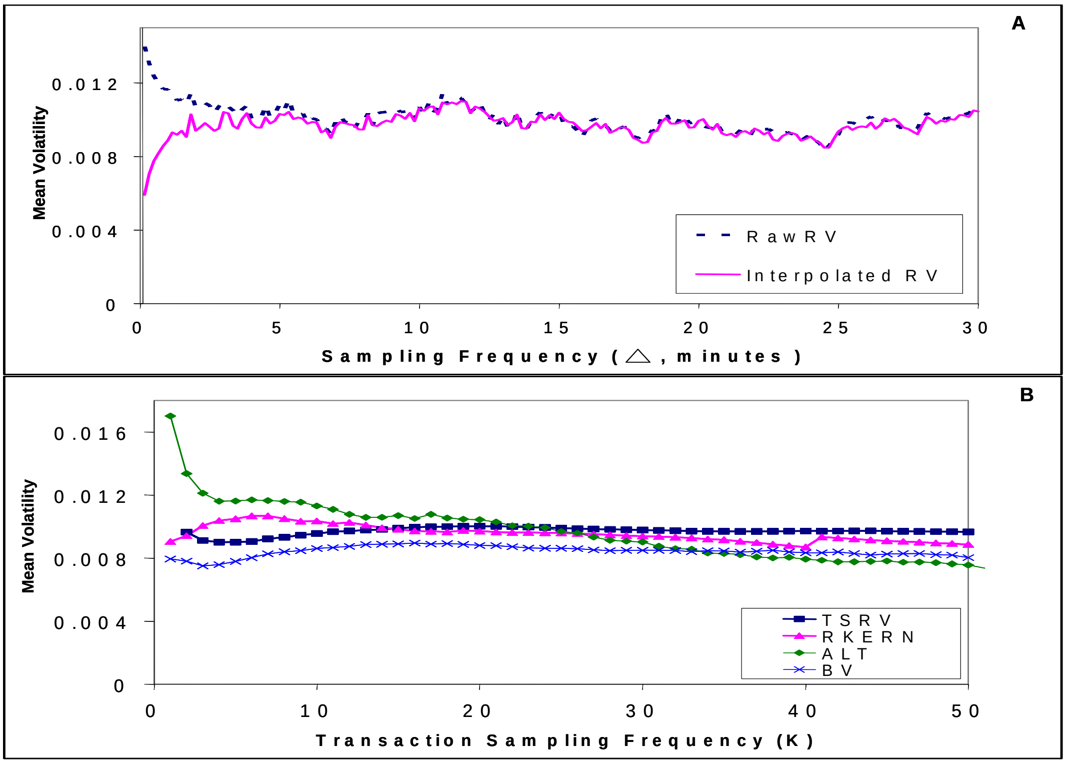

Figure 4: Panel A: Signature plots of Raw RV(Δ) and Interpolated RV(Δ) against Δ. The series shown are averaged over the trading days in August 2001. The average optimal sampling frequency over this period, calculated via (23), is 5.36 minutes. Panel B: signature plots of TSRV, RKERN, ALTM and BV against K. The series shown are averaged over the trading days in August 2001. The average optimal sampling frequency over this period, calculated as (17) is $K = 3$.

Panel A in Figure 4 plots two versions of RV(Δ) in (4) over different time periods Δ. Given that the observed data is irregularly spaced, two main choices are available for construction of RV(Δ) at regularly spaced points in time determined by Δ. With the so-called "raw" sampling method, if a transaction price is not observed at a given time point, then the previous transaction price is used as the sampled price. With the "linear interpolation" method, if a price is not observed at a given time point, then the sampled price is obtained by linearly interpolating between the price of the previous and next transaction. As $\Delta \to 0$, raw sampling produces an estimator that mimics the behaviour of the transaction based RV estimator in (11) as $n \to \infty$; hence, the asymptoting behaviour observed for the Raw RV measure as $\Delta \to 0$ in Panel A of

Figure 4. As noted by Hansen and Lunde (2006b), the linear interpolation method can produce downward bias at high frequencies, as is also evident in Panel A. Clearly there is little to choose between the two methods for $\Delta = 5$ minutes onwards. In the empirical application we use the linear interpolation method in constructing RV(5), RV(15) and OSRV, with the optimal frequency used in the construction of OSRV varying between approximately 5 and 30 minutes across the sample period.

In Panel B of Figure 4 we plot the four transactions-based estimators: TSRV in (15), RKERN in (18), ALTM in (30) and BV in (28), against sampling frequency K.[24] Four features are worthy of note. Firstly, the ALTM performance is the worst, with the value of the estimator changing substantially as K declines.[25] Secondly, the other three estimators have very similar, and very flat signature plots, meaning that the nature of their construction has served to render them quite robust to the microstructure noise that contaminates the RV(Δ) measures at high frequencies. Thirdly, all four signature plots are smoother than those in Panel A due, in the main, to the averaging across k ($k = 2, 3, \ldots, K$) used in the construction of the measures. Finally, the TSRV estimator has the flattest signature plot of all those considered.

5.3.2 Disributional and Memory Properties

Panels A and B of Figure 5 display histograms of the natural logarithms of the TSRV and BV volatility measures, and panels C and D the ACF's for the two measures. The corresponding results for the other measures are qualitatively similar and, hence, not reported. The normal densities imposed on the empirical densities demonstrate that the log of the variance measures are well approximated by a normal distribution, but with a small amount of excess kurtosis evident. The slow decline in the ACF's is typical of the memory behaviour exhibited throughout the literature for intraday data-based volatility measures; see, for example Andersen *et al.* (2003).

To cater for these empirical features, we produce direct forecasts using the following ARFIMA(p,d,q) model with Student t innovations (where the generic notation y_t refers to the logarithm of any of the volatility measures described in Section 3, and α its mean):

$$\phi(L)(1-L)^d(\ln y_t - \alpha) = \theta(L)u_t \; ; \qquad u_t \sim \textit{Student } t \; (0, \sigma^2 \frac{\nu}{\nu - 2}, \nu).$$

The autoregressive and moving average polynomials $\phi(L)$ and $\theta(L)$ are of lag length p and q respectively and $(1-L)^d$ is as defined earlier. The ARFIMA (p,d,q) models are estimated using conditional maximum likelihood, with the infinite lag structure induced by $(1-L)^d$ truncated at

[24] Note that in the empirical exercise, the optimal value of K in (17) varies between approximately 2 and 6 across the sample period.

[25] ALTM is based on the assumption that quoted prices change by the minimum tick value ($0.01 in our case). While this assumption may be valid for very actively traded stocks on large exchanges, we find that for our data the price change is commonly ten times the minimum tick amount. This may be a reason why ALTM does not perform as well as the other measures according to the signature plots.

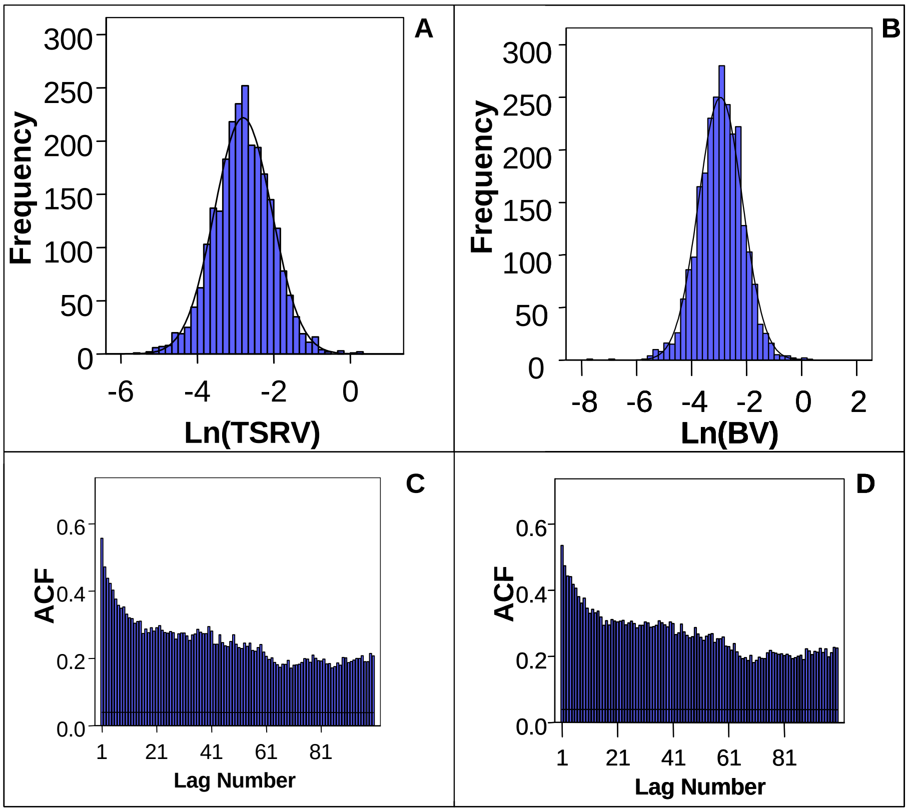

Figure 5: Panel A and B: histogram, with normal density superimposed, of ln(TSRV) and ln(BV) respectively (24 hour weighted and annualized) for the period 2 January 1992 to 28 December 2001. Panel C and D: autocorrelation function (ACF) for each measure over the same period.

the lag determined by the number of sample observations plus the number of pre-sample observations. Preliminary analysis has led to the use of ARFIMA (0,d,1) specifications in producing all forecasts. The rolling estimates of d range from 0.27 to 0.36, values which are typical for this type of data; see, for example, Andersen *et al.* (2003). For comparative purposes we also produce forecasts via short memory ARMA (2,0,1) models, with the lag lengths again determined via preliminary analysis.[26]

5.3.3 The Impact of Jumps

Of all alternatives considered, the two measures expected to display less volatile behaviour, due to their handling of jumps, are the BV and TSRV measures. As described in Section 3.2, the former measure is explicitly designed to estimate only the continuous component of quadratic variation. The latter measure, on the other hand, is expected to incorporate less jump

[26]Note that the application of standard model selection criteria to both ARFIMA and ARMA models for the different measures lead to slightly different choices of p and q. It was decided to use just one specification (of each model) for all measures.

information due to the nature of its construction, despite the fact that it formally converges to the sum of the continuous and jump components of quadratic variation, as per the results in Zhang *et al.* (2005).

In Figure 6, Panels A, B, and C respectively, we plot the time series of RV(15), TSRV, and BV over the time period used in the forecasting evaluation in Section 5.4. RV(15) is included as an estimate of quadratic variation, BV as estimate of the continuous component of quadratic variation, and TSRV as an estimator of quadratic variation in which the impact of jump variation is nonetheless reduced. In Figure 7, Panels A and B we then plot the difference between RV(15) and each of the two alternative estimators. As expected, BV displays less volatile behaviour than TSRV. The difference between RV(15) and BV in Panel A is formally an estimate of the jump component of quadratic variation. The graph in Panel B is an indication of the extent of jump behaviour that remains after the construction of the TSRV estimator via the difference in (15). As would be expected, the latter does not eliminate jump variation to the same extent as does BV, a fact which accords with the plots in Panels B and C of Figure 6. *A priori,* we would expect the relative smoothness of the alternative volatility series to have an impact on their forecasting performance, and on their ability to be forecast. This expectation is indeed vindicated by the empirical results in Section 5.4.

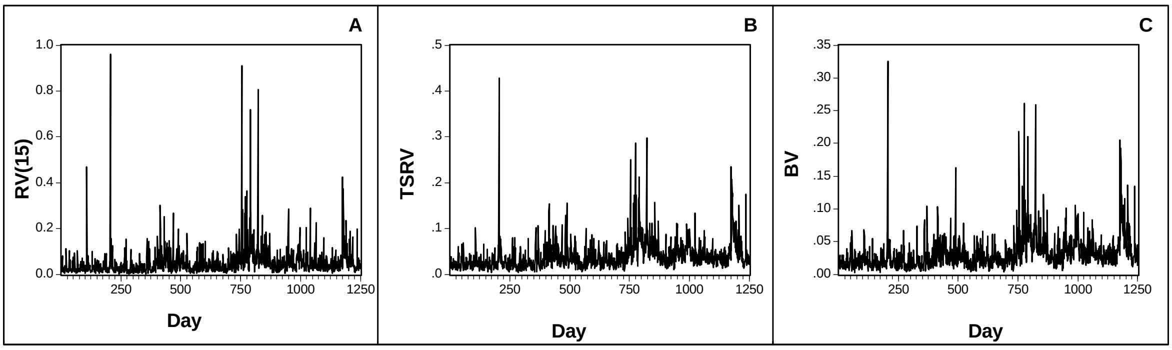

Figure 6: Panels A, B and C plot respectively the RV(15), TSRV and BV series for the January, 1997 to December 2001 period over which the forecast evaluation is performed.

5.3.4 Variance Measures for the Whole Day

Due to the fact that the ASX is open only for several hours during the trading day, the realized volatility and bi-power measures do not contain that portion of daily (i.e. 24 hour) volatility that is associated with the overnight return. Given that the overnight return, in this case, reflects price activity in the major U.S, U.K. and European exchanges, volatility in this component of the daily return may well be a significant component of the overall volatility for the 24 hour period.

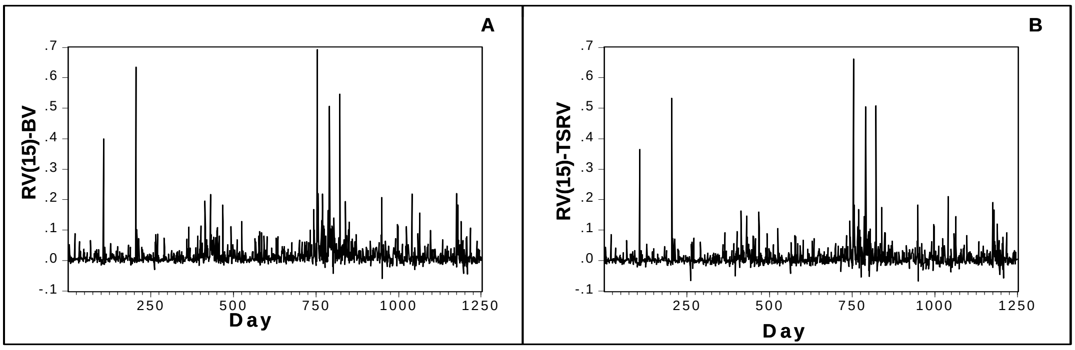

Figure 7: Panels A and B plot respectively the RV(15)-BV and RV(15)-TSRV series for the January, 1997 to December 2001 period over which the forecast evaluation is performed.

Following Hansen and Lunde (2005) we adjust the within-day volatility measures by taking a weighted average of the within-day measure and the squared overnight (close-to-open) return, where the weights are determined empirically using a mean squared error (MSE) criterion.[27] Defining $r_{on,t+1}$ as the overnight return, and using the TSRV measure in (15) for the purpose of illustration, the all-day measure is

$$TSRV_{t+1,weight} = w_1 r_{on,t+1}^2 + w_2 TSRV_{t+1}, \tag{35}$$

where the weights w_1 and w_2 are the solution to minimizing $var(TSRV_{t+1,weight})$ subject to the restriction that $E(TSRV_{t+1,weight}) = w_1 E(r_{on,t+1}^2) + w_2 E(TSRV_{t+1})$.[28] As shown in Hansen and Lunde, this criterion leads to

$$w_1 = (1-\phi)\frac{E(TSRV_{t+1,weight})}{E(r_{on,t+1}^2)}; w_2 = \phi\frac{E(TSRV_{t+1,weight})}{E(TSRV_{t+1})}, \tag{36}$$

where ϕ is referred to as a relative importance factor, and is defined by $\phi = \frac{a}{b}$, with

$$\begin{aligned} a &= [E(TSRV_{t+1})]^2 \, var(r_{on,t+1}^2) - E(r_{on,t+1}^2)E(TSRV_{t+1})cov(r_{on,t+1}^2, TSRV_t) \\ b &= [E(TSRV_{t+1})]^2 \, var(r_{on,t+1}^2) + \left[E(r_{on,t+1}^2)\right]^2 var(TSRV_{t+1}) \\ &\quad -2E(r_{on,t+1}^2)E(TSRV_{t+1})cov(r_{on,t+1}^2, TSRV_{t+1}). \end{aligned}$$

When $cov(r_{on,t+1}^2, TSRV_{t+1}) = 0$, it follows that $w_1/w_2 = \left[E(r_{on,t+1}^2)/E(TSRV_{t+1})\right] \times [var(TSRV_{t+1})/var(r_{on,t+1}^2)]$, in which case more weight is given to the within-day measure the

[27] We also produced some preliminary results using another method advocated by Hansen and Lunde, whereby the measure is rescaled according the average proportion of total volatility (of the open-to-open return) that the within-day volatility represents. As the results produced using both methods were quite similar, we decided to focus only on the weighted average method.

[28] This restriction essentially requires the variance estimator to be an asymptotically (as $n \to \infty$) unbiased estimator of integrated volatility (or quadratic variation in the case of the more general model in (1)). This restriction is essentially ignored in the case of certain of our estimators, e.g. RV($\Delta = 5$ minutes), in which some bias may remain.

larger is the average value of that volatility (relative to that of the overnight volatility), and less weight the larger is the relative variance of the within-day measure.[29]

Estimating all population quantities in (36) by the relevant sample counterparts, for all of the measures considered in the empirical exercise, we find that the estimates of w_1 vary between approximately 0.06 and 0.3, whilst the estimates of w_2 vary between approximately 1.8 and 2.7. The relatively high weight given to the within-day measure reflects both the magnitude and the variation in the overnight volatility; a reflection, in turn, of the extent of the activity in the important world markets in the Northern Hemisphere.[30] Importantly, the more noisy within-day measures, such as RV(15), for example, produce, all other things equal, a higher value of w_1. This means that the 24 hour RV(15) measure is even noisier relative to 24 hour versions of less noisy measures, such as TSRV, for example.

5.4 Empirical Results: Forecast Evaluation

As is common in evaluations of competing forecasts of volatility, we use both univariate and encompassing forecast regressions. In the univariate regressions, one of the alternative proxies of volatility is regressed on a single forecast, with the fit of the regression assessed via R^2. The significance and unbiasedness of the forecasts can also be assessed via tests of the appropriate parameter restrictions. In the encompassing regressions, two alternative forecasts are included, with conclusions then drawn about the extent to which one forecast remains significant given the presence of the other and/or subsumes all information contained in the other forecast(s). We begin by using logarithmic variance quantities as the arguments in all regressions, with all quantities expressed in annualized terms (and all intraday measures re-weighted as 24 hour measures prior to being annualized). The logarithmic quantities are used in order to reduce the impact on the results of outliers (see Pagan and Schwert, 1990) and to render our results comparable to most of those reported in the literature. Robustness to outliers is particularly important for the Australian data set under analysis, given the impact on the intraday measures, when re-weighted as 24 hour measures, of activity in the Northern Hemisphere markets. However, in view of the issues raised in Hansen and Lunde (2006a) regarding 'robust' versus 'sensitive' selection criteria, we also produce results for the variance quantities. As indicated by the results reported below, despite some changes, the overall ranking of the models is largely unaffected by the choice of transformation.

Results are produced for one day ahead, 10 days ahead and one month (22 days) ahead

[29] This interpretation of the weights holds approximately when $cov(r^2_{on,t+1}, TSRV_{t+1}) \neq 0$. Note that there is an error in the relevant formula in Hansen and Lunde (2005).

[30] In contrast, typical values of w_1 and w_2 reported in Hansen and Lunde (2005) for 30 stocks in the Dow Jones Industrial Average are about 0.5 and 1.2 respectively. Like Hansen and Lunde, in the empirical analysis we omit approximately 1% of the overnight observations as extreme outliers, in the calculation of w_1 and w_2. Once the weights have been calculated, however, the observations are retained in the subsequent analysis.

forecasts. A multiple steps ahead forecast is the (annualized) average of the individual forecasts for each day in the forecast period. For example, the 10 day ahead forecast is the average of the one step ahead, two step ahead, up to 10 step ahead forecasts. Correspondingly, the variance measure being forecast corresponds to the (annualized) average of the values of the variance measures over the forecast period. Forecast regressions are run for seven alternative proxy measures: RV(5) and RV(15) in (4), OSRV in (22), TSRV in (15), RKERN in (18), BV in (28) and ALTM in (30). The regressions are estimated using ordinary least squares, with robust standard errors used for the multistep ahead regressions, to adjust for the overlapping data problem.

Using generic notation $V_{t+1,h}^{(j)}$ to denote the logarithm of the jth variance proxy measure, $j = 1, 2, \ldots, 7$, over a period of h days from $t+1$ to $t+h$, we denote by $V_{t,h}^{(j)ARMA}$ and $V_{t,h}^{(j)ARFIMA}$ respectively the $h-$day ahead forecasts of $V_{t+1,h}^{(j)}$ based on ARMA and ARFIMA models for $V_t^{(j)}$, estimated using the intraday sample data up to period t. We denote by $GARCH_{t,h}$ the forecast based on a GARCH model estimated from daily returns up to period t. Variations of the GARCH model, including the long memory FIGARCH specification, are denoted by the appropriate acronym. Implied logarithmic variance forecasts extracted from the option price data on day t are denoted by $IV_{t,h}^{MF}$ (MF) and $IV_{t,h}^{BS}$ (BS). All results thus relate to *out-of-sample* forecasts. This needs to be remembered when comparing our results with other results in the literature, many of which relate to *within-sample* evaluations.

To explain the criteria used to evaluate the forecast regressions, we use as an example the univariate regression of a logarithmic variance proxy for $t+1$, $V_{t+1,h}^{(j)}$, on the direct ARFIMA forecast, $V_{t,h}^{(k)ARFIMA}$, $j, k = 1, 2, \ldots, 7$, the univariate regression of $V_{t,h}^{(j)}$ on the indirect TGARCH forecast, the univariate regression of $V_{t,h}^{(j)}$ on the indirect $IV_{t,h}^{MF}$ forecast, plus the encompassing regression of $V_{t+1,h}^{(j)}$ on $V_{t,h}^{(k)ARFIMA}$ and $IV_{t,h}^{MF}$:

$$V_{t+1,h}^{(j)} = \alpha + \beta_1 V_{t,h}^{(k)ARFIMA} + e_t \quad (37)$$

$$V_{t+1,h}^{(j)} = \alpha + \beta_1 TGARCH_{t,h} + e_t \quad (38)$$

$$V_{t+1,h}^{(j)} = \alpha + \beta_1 IV_{t,h}^{MF} + e_t \quad (39)$$

$$V_{t+1,h}^{(j)} = \alpha + \beta_1 V_{t,h}^{(k)ARFIMA} + \beta_2 IV_{t,h}^{MF} + e_t. \quad (40)$$

With regard to (37) we test for the information content of the long memory direct forecast of $V_{t+1,h}^{(j)}$ by testing $H_0 : \beta_1 = 0$. The unbiasedness of the forecast is, in turn, assessed via a test of the joint null $H_0 : \alpha = 0; \beta_1 = 1$. The fit of the regression is measured in the usual way, via the coefficient of determination, R^2. *A priori* we would anticipate a better fit when $j = k$. However, it may be that one particular measure does well at forecasting all other measures, whilst another measure may be the easiest to forecast (see, for e.g. Bandi and Russell, 2006). With regard to (38), we conduct similar tests to above, but with the focus being on the ability of the indirect forecasts based on daily returns data to forecasts the various measures of volatility. In (39), on the other hand, the focus is on the ability of the option market to forecast future volatility. *A priori,* given the longer term focus of option contracts, we would expect the performance of the option forecasts to improve as the time horizon increases. In (40) we test for the informational efficiency of each regressor in the presence of another forecast. For example, if $IV_{t,h}^{MF}$ subsumes the information contained in $V_{t,h}^{(k)ARFIMA}$, then the null $H_0 : \beta_1 = 0$ will not be rejected. If, on the other hand, $V_{t,h}^{(k)ARFIMA}$ contains incremental information over and above information contained in $IV_{t,1}^{MF}$, then the estimate of β_1 will be significantly different from zero, even in the presence of a significant $IV_{t,1}^{MF}$.

Tables 1 contains goodness of fit results for all specifications of the form of (37), while Table 2 contains the results for the corresponding specifications using ARMA(2,1) (rather than ARFIMA(0,d,1)) forecasts. The R^2 associated with the best performing forecast and the measure that is easiest to forecast, on average, are highlighted in bold in both tables.

<<< **Tables 1 and 2 here** >>>

The key results in Tables 1 and 2 are as follows. The bi-power measure, BV, is the *easiest to forecast*, in terms of producing the largest R^2 for the out-of-sample forecast regression. This finding holds no matter which of the seven alternative measures is used to produce the direct forecast of BV, no matter whether a short or long memory model underlies those forecasts, and for all forecast horizons considered. This is a remarkably clear-cut result and somewhat vindicates the focus of Anderson and Vahid (2005) on constructing forecasting models for the continuous component of quadratic variation. Note that BV is the smoothest series of those displayed in Section 5.3.3, and is indeed the least volatile of all seven measures considered,

The TSRV measure is the second-easiest measure to forecast. This result is also uniform across all scenarios considered, and is consistent with the preliminary analysis in Section 5.3.3. The conventional RV(15) measure, being one of the two measures that does not explicitly adjust for microstructure noise (or jumps), is the most difficult measure to forecast, again for all direct forecasts, models and forecast horizons considered.

The results regarding the *best performing direct forecast* are more varied. Out of the six panels of results in Tables 1 and 2, RV(15) is the best performing forecast three times, ALTM twice and TSRV once. It would appear that the relatively large degree of variation in the RV(15) measure (the highest of all seven measures considered) helps reduce the forecast error and, thereby, increases the value of the R^2 in the forecast regressions. Although all comparable results are quite similar for the ARFIMA and ARMA specifications, contrary to expectations, the ARFIMA forecasts perform slightly better one step ahead and the ARMA results slightly better overall 22 steps ahead.[31]

For both the ARFIMA and ARMA specifications in Tables 1 and 2, the R^2 increases as the forecast is changed from one step ahead to (the average over) 10 steps ahead. However, as might be anticipated, once the forecast horizon extends to 22 steps ahead, the inaccuracy of the forecasts for days far out in that horizon counteracts the positive impact of the averaging process, with the R^2 in the bottom panel of all three tables exhibiting only some slight improvements over the R^2 in the corresponding second panel, and some slight reductions.

In light of these results, and in an attempt to render the number of reported results manageable, we conduct the subsequent analysis using only three of the realized variance measures: RV(15), TSRV and BV. The first measure is retained as it produces the most accurate direct forecasts. The second measure, TSRV, produces, after ALTM, the next most accurate forecasts. Given the signature plots in Section 5.3.1, the forecasting performance of TSRV can be viewed as more robust than that of ALTM, with the signature plot for TSRV being the flattest of all considered. Hence, we retain TSRV in the subsequent analysis rather than ALTM. The BV measure is retained as it is the easiest measure to forecast overall. The fact that the TSRV measure is the second easiest measure to forecast is a further reason for its retention.

In Table 3 we report the forecasting regression R^2 associated with the indirect daily returns-based forecasts. Results for all four GARCH models in (34) are included. In summary, the indirect forecasts are inferior, overall, to the direct forecasts summarized in Tables 1 and 2. However, the ranking of the measures, in terms of ease of forecast, remains the same as for the direct forecasting methods, namely 1. BV; 2. TSRV; 3. RV(15). Similarly, the same pattern of behaviour in the R^2 over the forecast horizons as evident in Tables 1 and 2, occurs for the indirect forecasts. There is little to choose between the alternative forecast models, especially at the one step ahead horizon. However, overall, the TGARCH model produces the best performing forecasts.

<<< Table 3 here >>>

In Tables 4 to 7 the returns-based forecasts, both direct and indirect, are compared with the option-implied volatility forecasts (both MF and BS). Table 4 contains the R^2 results for all forecast horizons, whilst Tables 5 to 7 report estimates of the parameters in the regressions of the form of (37) to (40), including the results of the relevant hypothesis tests, for the one step and 22 step ahead forecast horizons only. The direct forecasts used in Tables 5 to 7 are the best

[31] The similarity of the ARFIMA and ARMA results is consistent with that of Pong *et al.* (2004). Regarding the relative performance of the ARFIMA/ARMA forecasts at the longer horizons, it is important to remember that the multiple steps ahead forecasts are averages of forecasts for all days over the forecast horizon. If single forecasts for 10 and 22 days ahead were produced, the long-memory model may well have performed better for the longer-term forecast.

performing ARFIMA forecasts according to the results in Table 4, with the measure from which that forecast has been constructed included in parentheses. Similarly, the IV forecast used in the encompassing regressions reported in Tables 5 to 7 is the best performing of the two IV forecasts according to the results in Table 4.

<<< **Tables 4 to 7 here** >>>

The key results in Table 4 are as follows. The overall average ranking of the measures, in terms of ease of forecast, remains the same 1. BV; 2. TSRV; 3. RV(15), even when the IV measures are included in the assessment. However, when the IV measures are considered on their own, the BV measure is the easiest to forecast only at the 22 day horizon. This is an interesting result given that IV^{MF}, at least, explicitly estimates quadratic variation in both the continuous and jump component, with the latter being omitted in the BV measure. For the shorter horizons, the TSRV measure is the easiest for both IV^{MF} and IV^{BS} to forecast. The MF IV forecast performs better than the BS IV forecast, in eight out of the nine cases considered.

One day ahead, the direct ARFIMA forecast based on TSRV is the best performing forecast, no matter what measure is being forecast; with the IV forecasts the worst performing. By 22 steps ahead however, the situation changes, with IV^{MF} being the best performing forecast, on average, and TSRV one of the lesser performers. The TGARCH model provides the best forecast in only one of the nine cases considered.

The results in Tables 5 to 7 shed further light on the relative performance of the alternative forecasts, with all comments referring to significance at the 1% level. Considering the results for all univariate regressions, the $p-$values for testing $H_0 : \beta_1 = 0$ indicate that all four forecasts considered, the (best performing) direct ARFIMA forecast, the TGARCH forecast and the two IV forecasts, are all individually significant for any of the three volatility measures. This is the case at both the one and 22 day horizons.[32] For the direct ARFIMA forecasts, in five of the six cases considered, individual tests of $H_0 : \alpha = 0$ and $H_0 : \beta_1 = 1$ fail to reject the null hypotheses. That is, the ARFIMA model can be viewed as producing unbiased forecasts of volatility in virtually all scenarios considered.[33] Using this same criterion, the IV^{MF} is the only other forecast to be unbiased (for RV(15) one day ahead). TGARCH and IV^{BS} are never unbiased according to this criterion. Focussing on the slope coefficient only in the univariate

[32] Following Andersen *et al.* (2005), our reported p-values are calculated using the Bartlett/Newey-West heteroskedasticity consistent covariance matrix estimator with a bandwidth of five used for the one step ahead estimates and 44 used for the 22 step ahead estimates.

[33] Strictly speaking, unbiasedness should be expressed as the joint null: $H_0 : \alpha = 0; \beta_1 = 1$. At the 5% level, this joint null was rejected for all specifications based on logarithmic variance and variance level quantities. Similarly joint tests of efficiency and unbiasedness in the encompassing regressions were rejected (at the 5% level) in all cases and, hence, not reported. In virtually all cases the joint null was rejected at the 1% level also.

regressions, IV^{MF} has a coefficient insignificantly different from one in five of the six cases. In all cases but one, its slope coefficient is closer to one than is the coefficient associated with IV^{BS}.

Considering the results for the encompassing regressions, the most notable result is that IV^{MF} provides incremental information in forecasting all three measures considered, over both time horizons. In the case of the 22 day ahead forecasts of RV(15), IV^{MF} is not only a significant forecast, it also subsumes the information in the ARFIMA and TGARCH forecasts. It subsumes the information in the latter forecast in the case of the 22 day ahead forecasts of TSRV also. For the longer time horizon, the TGARCH and ARFIMA forecasts, when combined, provide individually insignificant forecasts of future volatility in all but one case (whereby the RV(15)-based ARFIMA forecast of BV is significant).

<<< **Tables 8 to 11 here** >>>

When comparable results to those reported in Tables 4 to 7 are produced using all quantities in variance rather than logarithmic variance form, the basic flavour of most the key results still obtains. In particular, with reference to Tables 9 to 11, IV^{MF} is still a significant forecast in all forecasting regression, both univariate and encompassing.[34] However, it now provides incremental information over the returns-based forecasts, for all three measures considered, and over both time horizons. Crucially, at the 22 day horizon, IV^{MF} now subsumes information in the ARFIMA and TGARCH forecasts of all three measures. In the univariate regressions for both IV measures, the negative intercepts are consistent with a risk premium having been factored into the observed market prices, rendering the option-implied volatility estimates larger, on average, than the returns-based estimates. When combined with ARFIMA forecasts the TGARCH forecasts are now insignificant not just at the 22 day time horizon (as before), but also one day ahead.

Interestingly, given the higher weighting given to the (potentially noisy) overnight returns in the case of RV(15) (compared with the weighting given in the case of TSRV), the TSRV-based ARFIMA forecasts of RV(15) do not go close to satisfying the unbiasedness condition in the univariate regression, and are insignificant at both time horizons when combined with either the TGARCH or IV^{MF} forecasts. In contrast, the TSRV-based one-step ahead ARFIMA forecasts of both BV and TSRV itself still do reasonably well. When used to forecast TSRV and BV 22 steps ahead, the RV(15) measure, when combined with the TGARCH forecast, is more significant than in the log variance case, with a coefficient closer to one.

The R^2 results in Table 8, a 'robust criterion function' when based on variance quantities, according to the results in Hansen and Lunde (2006a), still give the highest ranking to the IV^{MF}

[34]Note that in order for the variance-based results to be comparable with those based on the log variance quantities, we have included IV^{MF} , rather than IV^{BS}, in all encompassing specifications, despite the fact that IV^{BS} performs slightly better than IV^{MF} , one-step-ahead, according to the results in Table 9.

measure at the longest time horizon. However, in contrast with the corresponding 'sensitive' R^2 results based on the log variance quantities, as reported in Table 4, the IV measures now also win at both of the smaller time horizons. As anticipated, all R^2 values are noticeably lower, in particular for the one-step ahead regressions for the RV(15) measure, which are likely to be most influenced by extreme overnight returns.

5.5 Summary of Key Results

With reference to the key research questions outlined in the Introduction, and further delineated in Section 4.1, the following statements can be made. Firstly, and most importantly, according to all criteria considered, the option-implied IV^{MF} forecast is superior overall to the direct and indirect returns-based forecasts, and to the alternative option-implied forecasts, IV^{BS}, in particular at the longer-term horizons. The direct forecasts tend to perform second-best, followed by indirect TGARCH forecasts, with there being no clear-cut ranking of long-and short memory returns-based specifications. The significance (and incremental informativeness) of the IV^{MF} forecast is robust to: the measure of volatility being forecast and the measure of volatility used to produce the competing direct forecast. As such, the superior performance of the option-based forecasts, as based on the criterion of statistical significance, is robust to the treatment of microstructure noise and jumps in the measurement of latent volatility.[35] The superiority of the option-based forecasts is also invariant to the time horizon and the transformation of volatility used in the forecasting assessment. The goodness of fit criterion ranks the IV^{MF} forecasts first at the 22 day horizon, whilst the variance-based R^2 ranks both option-implied forecasts as a clear first and second, at all time horizons, with the IV^{MF} forecast being the best performing forecast for the 10 and 22 day horizons.

The measures that are easiest to forecast overall are two of the measures that expressly cater for the microstructure noise, namely the (noise-adjusted) bi-power measure, BV, and the TSRV measure. The most difficult measure to forecast is the conventional RV(15) measure, in which no explicit adjustment is made for microstructure noise. BV, the only measure in which in which the quadratic variation in the random jump component of returns is formally omitted from the calculation, is consistently ranked as the easiest measure to forecast. However, when the option-implied forecasts only are considered, only in two cases of the twelve considered in Tables 4 and 8 is BV the easiest measure to forecast (in terms of having the highest R^2). From this one can conclude that the option prices have factored in some jump information and, hence, do better at predicting measures of volatility in which some of that information is also implicitly incorporated. With respect to IV^{MF} in particular, this result is consistent with the fact that

[35] Remembering that at this stage of the assessment, only three representative measures of volatility are retained, based on the preliminary assessment of all seven measures initially considered.

IV^{MF} is theoretically an estimate of quadratic variation, including that of the jump component. In all ten other cases, TSRV is the measure that the option-implied volatilities forecast best, of the three measures, BV, TSRV and RV(15), considered at this point in the assessment. In other words, the option market does best at forecasting a measure of volatility that makes adjustment for microstructure noise and only partial (but not complete) adjustment for random jumps.

6 Conclusions

This paper presents the first empirical evaluation of option-implied versus returns-based volatility forecasts that takes into account all of the important recent developments regarding market microstructure noise and random jumps. The option-based component of the analysis also accommodates the concept of model-free implied volatility, in an attempt to separate the forecasting performance of the options market from the issue of misspecification of the option pricing model.

The numerical results suggest that the particular option market considered in the analysis provides the best forecasts, overall, of future volatility, no matter how the latter is measured. That said, the model-free option implied volatility tallies with its theoretical underpinnings in producing its best forecasts when the measure of volatility to be forecast is adjusted for microstructure noise, but retains jump information. These results give further weight to the growing consensus regarding the informational efficiency of options market, but with more light now shed on the precise nature of that information.

References

[1] Ait-Sahalia, Y, Mykland, P.A. and Zhang, L. 2005. Ultra High Frequency Volatility Estimation with Dependent Noise, *National Bureau of Economic Research Working Paper 11380.*

[2] Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. 2000. Great realizations, *Risk,* 13: 105–108.

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. 2003. Modelling and Forecasting Realized Volatility, *Econometrica,* 71: 579-625.

[4] Andersen, T.G., Bollerslev, T., Diebold, F.X. 2005. Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, *National Bureau of Economic Research Working paper 11775.*

[5] Andersen, T.G., Bollerslev, T. and Meddahi, N. 2005. Correcting the Errors: Volatility Forecast Evaluation using High Frequency Data and Realized Volatilities, *Econometrica,* 73: 279-296.

[6] Anderson, H.M. and Vahid, F. 2005. Forecasting the Volatility of Australian Stock Returns: do Common Factors Help?, *Australian National University Draft Paper.*

[7] Bakshi, G., Cao, C. and Chen, Z. 1997. Empirical Performance of Alternative Option Pricing Models, *Journal of Finance,* 52: 2003-2049.

[8] Bandi, F.M. and Russell, J.R. 2006. Separating Microstructure Noise from Volatility, *Journal of Financial Economics,* 79: 655-692.

[9] Barndorff-Neilsen, O.E. and Shepherd, N. 2002. Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society B,* 64: 253-280.

[10] Barndorff-Neilsen, O.E. and Shepherd, N. 2004. Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics,* 2: 1-37.

[11] Barndorff-Neilsen, O.E. and Shephard, N. 2005. Variations, Jumps, Market Frictions and High Frequency Data in Financial Econometrics, Forthcoming in *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress,* (eds. R. Blundell, P. Torsten and W. Newey), Econometric Society Monographs, Cambridge University Press.

[12] Barndorff-Neilsen, O.E., Lunde, A., Hansen, P.R. and Shephard, N. 2005. Realized Kernels can Consistently Estimate Integrated Variance: Correcting Realized Variance for the Effect of Market Frictions, *Draft Paper.*

[13] Barndorff-Neilsen, O.E., Lunde, A., Hansen, P.R. and Shephard, N. 2006. Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise, *Draft Paper.*

[14] Bates, D.S. 2000. Post-87 Crash Fears in the S&P 500 Futures Option Market, *Journal of Econometrics,* 94: 181-238.

[15] Black, F. and Scholes, M., 1973 The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81: 637-659.

[16] Blair, B.J, Poon, S-H. and Taylor, S.J. 2001. Forecasting S&P100 Volatility: the Incremental Information Content of Implied Volatilities and High Frequency Index Returns, *Journal of Econometrics*, 105, 5-26.

[17] Britten-Jones, M. and Neuberger, A. 2000. Option Prices, Implied Price Processes and Stochastic Volatility. *The Journal of Finance,* LV: 839-866.

[18] Brownlees, C.T. and Gallo, G.M. 2005. Ultra High-Frequency Data Management, in *Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione* (ed. C. Provasi), Atti del Convegno S. Co., CLUEP, Padova.

[19] Chenoweth, N. 1994. Prefs Listing Seals Murdoch Control, *Australian Financial Review, Companies and Markets,* 4th of November.

[20] Christensen, B.J. and Prabhala, N.R. 1998. The Relation Between Implied and Realized Volatility. *Journal of Financial Economics,* 50: 125-150.

[21] Christensen, B.J., Hansen, C.S. and Prabhala, N.R. 2001. The Telescoping Overlap problem in Options Data, *Draft Paper.*

[22] Collins, L. 2000a. Suddenly, Murdoch's Become A Millstone, *Australian Financial Review, News,* 12th of January 2000.

[23] Collins, L. 2000b. $10.9bn Wiped Off News Corp, *Australian Financial Review, Investment,* 18th of November.

[24] Corrado, C.J. and Su, T. 1997. Implied Volatility Skews and Stock Index Skewness and Kurtosis implied by S&P 500 index Option Prices, *Journal of Derivatives*, Summer: 8-19.

[25] Dacorogna, M.M., Gençay, R., Müller, U., Olsen, R.B. and Pictet, O.V. 2001. *An Introduction to High-Frequency Finance,* Academic Press, San Diego.

[26] Day, T.E. and Lewis, C.M. 1995. Stock Market Volatility and the Information Content of Stock Index Options, in *ARCH, Selected Readings* (ed. R. Engle), Oxford University Press, Oxford.

[27] Griffin, J.E. and Oomen, R.C.A. 2006. Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time?, Forthcoming in *Econometric Reviews.*

[28] Guo, D. 1998. The Risk Premium of Volatility Implicit in Currency Options, *Journal of Business and Economic Statistics*, 16: 498-507.

[29] Hansen, P.R. and Lunde, A. 2005. A Realized Variance for the Whole day Based on Intermittent High-Frequency Data, *Journal of Financial Econometrics,* 3: 525–554.

[30] Hansen, P.R and Lunde, A. 2006a. Consistent Ranking of Volatility Models, *Journal of Econometrics,* 131: 97-121.

[31] Hansen, P.R and Lunde, A. 2006b. Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics*, 24: 127-218.

[32] Heston, S.L. 1993. A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *The Review of Financial Studies*, 6: 327-343.

[33] Heston, S. and Nandi, S. 2000. A Closed-Form GARCH Option Valuation Model. *Review of Financial Studies*, 13: 585-625.

[34] Huang, X. and Tauchen, G. 2005. The Relative Contribution of Jumps to Total Price Variation, *Journal of Financial Econometrics*, 3: 456-499.

[35] Hull, J.C. (2000). *Options, Futures, and Other Derivative Securities*, 4rd ed., Prentice Hall, New Jersey.

[36] Koopman, S. J., Jungbacker, B. and Hol, E. 2004. Forecasting Daily Variability of the S&P100 Stock Index using Historical, Realized and Implied Volatility Measurements, *Department of Econometrics, Free University Amsterdam Working Paper.*

[37] Jiang, G.J. and Tian, Y.S. 2005. The Model-Free Implied Volatility and its Information Content, *The Review of Financial Studies*, 18: 1305-1342.

[38] Large, J. 2005. Estimating Quadratic Variation when Quoted Prices Jump by a Constant Increment, *Nuffield College, Oxford Draft Paper.*

[39] Lim, G.C., Martin, G.M. and Martin, V.L. 2005. Parametric Pricing of Higher Order Moments in S&P500 Options, *Journal of Applied Econometrics,* 20: 377-404.

[40] Martens, M. and Zein, J. 2003. Predicting Financial Volatility: High-Frequency Time Series Forecasts Vis-a-Vis Implied Volatility, *Journal of Futures Markets,* 24: 1005-1028.

[41] Mincer, J. and Zarnowitz, V. 1969. The Evaluation of Economic Forecasts and Expectations, in *Economic Forecasts and Expectations,* (ed. J. Mincer), *National Bureau of Economic Research.*

[42] Oomen, R.C.A. 2006. Properties of Realized Variance under Alternative Sampling Schemes, *Journal of Business and Economic Statistics*, 24: 219-237.

[43] Pong, S., Shackleton, M.B. and Taylor, S.J. 2004. Forecasting Currency Volatility: a Comparison of Implied Volatilities and AR(FI)MA Models, *Journal of Banking and Finance,* 28: 2541-2563.

[44] Poteshman, A.M. 2000. Forecasting Future Volatility from Option Prices, *Department of Finance, University of Illinois Working Paper.*

[45] Zhang, L., Mykland, P.A. and Ait-Sahalia, Y. 2005. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data, *Journal of the American Statistical Association,* 100: 1394-1411.

Appendix: Data Cleaning Details

In order to calculate realized volatility measures the intraday data needs to be cleaned to remove certain extreme outliers. Recently there has been much discussion about the amount of data cleaning that should be performed prior to the calculation of realized volatility measures; see e.g. See Ait-Sahalia *et al.* (2005) and Hansen and Lunde (2006). The argument in favour of performing a large degree of data cleaning is that the removal of especially unusual transactions reduces the noise and hence improves the accuracy of realized volatility measures. The counter argument is that a large amount of data cleaning can constitute significant "surgery" on the data and somewhat defeats the purpose of designing realized volatility measures to be robust to noisy data. For further discussion of data cleaning methods see Dacorogna et al. (2001) and Brownlees and Gallo (2005).

The pre-filtering of the Newscorp equity data has involved: 1) removal of all days with less than three hours of active trading; 2) removal of all trades outside of the period from 10am to 4pm; 3) removal of all off market transactions; 4) aggregation by volume transactions with the same time stamp; and 5) removal of all "instantaneous price reversals" whereby the price changes by more than 0.5% to a level that has not occurred within the previous half hour and does not occur within the subsequent half an hour of trading, and then instantaneously reverses to its previous level on the next transaction. The latter criterion leads to less than 2% of observations being deleted. The 3rd of November 1994 has also been removed due to the issuing of "non-voting preference shares" which caused the share price to fall from \$8.14 to \$5.85 and resulted in exceptionally high turnover and intraday volatility (Chenoweth, 1994).

The pre-filtering of the option data has involved: 1) removal of all days with less than three hours of active trading; 2) removal of all off market transactions; and 3) removal of arbitrage trades performed by market makers.

Table 1:

R^2 for the univariate regressions using **direct long memory** out-of-sample forecasts, as in (37). All variables are expressed as logarithmic variance quantities and ARFIMA(0,d,1) models are used to produce all forecasts.

$V_{t,h}^{ARFIMA}$:	$RV(5)$	$RV(15)$	$OSRV$	$TSRV$	$RKERN$	BV	$ALTM$	
$V_{t+1,1}$			R^2 from 1 step ahead forecast regression					Average
$RV(5)$	0.344	0.349	0.335	0.355	0.349	0.336	0.359	0.347
$RV(15)$	0.292	0.297	0.287	0.301	0.298	0.281	0.303	0.294
$OSRV$	0.323	0.330	0.323	0.333	0.331	0.317	0.334	0.327
$TSRV$	0.371	0.381	0.367	0.388	0.377	0.368	0.386	0.377
$RKERN$	0.350	0.359	0.346	0.364	0.361	0.339	0.369	0.356
BV	0.383	0.386	0.377	0.399	0.381	0.390	0.388	**0.386**
$ALTM$	0.347	0.355	0.341	0.355	0.351	0.333	0.365	0.350
Average	0.344	0.351	0.340	0.356	0.350	0.338	**0.358**	
$V_{t+1,10}$			R^2 from 10 steps ahead forecast regression					Average
$RV(5)$	0.412	0.417	0.395	0.414	0.412	0.384	0.425	0.409
$RV(15)$	0.354	0.357	0.335	0.355	0.356	0.320	0.366	0.349
$OSRV$	0.399	0.406	0.385	0.402	0.404	0.368	0.412	0.397
$TSRV$	0.447	0.455	0.430	0.449	0.449	0.419	0.455	0.444
$RKERN$	0.421	0.429	0.403	0.423	0.428	0.383	0.435	0.418
BV	0.478	0.481	0.464	0.481	0.472	0.465	0.479	**0.474**
ALT	0.396	0.404	0.376	0.392	0.395	0.356	0.409	0.390
Average	0.415	0.421	0.398	0.417	0.417	0.385	**0.426**	
$V_{t+1,22}$			R^2 from 22 steps ahead forecast regression					Average
$RV(5)$	0.413	0.426	0.400	0.409	0.410	0.376	0.426	0.409
$RV(15)$	0.335	0.347	0.323	0.332	0.335	0.296	0.351	0.331
$OSRV$	0.393	0.408	0.383	0.391	0.394	0.354	0.407	0.390
$TSRV$	0.436	0.452	0.426	0.435	0.437	0.402	0.446	0.434
$RKERN$	0.412	0.426	0.400	0.409	0.418	0.366	0.429	0.409
BV	0.476	0.489	0.468	0.474	0.468	0.455	0.477	**0.472**
$ALTM$	0.380	0.394	0.365	0.374	0.378	0.335	0.396	0.375
Average	0.406	**0.420**	0.395	0.404	0.406	0.369	0.419	

Table 2:

R^2 for the univariate regressions using **direct short memory** out-of-sample forecasts (i.e. (37) but with $V_{t,h}^{ARMA}$ replacing $V_{t,h}^{ARFIMA}$ as the forecast). All variables are expressed as logarithmic variance quantities and ARMA(2,1) models are used to produce all forecasts.

$V_{t,h}^{ARMA}$:	$RV(5)$	$RV(15)$	$OSRV$	$TSRV$	$RKERN$	BV	$ALTM$	
$V_{t+1,1}$	R^2 from 1 step ahead forecast regression							Average
$RV(5)$	0.339	0.342	0.335	0.353	0.345	0.336	0.354	0.344
$RV(15)$	0.286	0.288	0.284	0.299	0.293	0.281	0.298	0.290
$OSRV$	0.314	0.320	0.317	0.327	0.324	0.313	0.326	0.320
$TSRV$	0.364	0.372	0.366	0.384	0.373	0.367	0.380	0.372
$RKERN$	0.342	0.350	0.344	0.359	0.355	0.337	0.362	0.350
BV	0.376	0.378	0.375	0.395	0.377	0.388	0.383	**0.382**
$ALTM$	0.342	0.349	0.340	0.354	0.348	0.333	0.361	0.347
Average	0.338	0.343	0.337	**0.353**	0.345	0.336	0.352	
$V_{t+1,10}$	R^2 from 10 steps ahead forecast regression							Average
$RV(5)$	0.399	0.422	0.410	0.418	0.417	0.396	0.418	0.411
$RV(15)$	0.338	0.357	0.347	0.356	0.358	0.330	0.357	0.349
$OSRV$	0.380	0.401	0.392	0.399	0.401	0.375	0.400	0.393
$TSRV$	0.427	0.454	0.443	0.448	0.450	0.429	0.444	0.442
$RKERN$	0.400	0.429	0.417	0.422	0.429	0.393	0.424	0.416
BV	0.460	0.477	0.471	0.480	0.472	0.474	0.468	**0.472**
$ALTM$	0.384	0.413	0.398	0.399	0.405	0.372	0.405	0.397
Average	0.398	**0.422**	0.411	0.417	0.419	0.395	0.417	
$V_{t+1,22}$	R^2 from 22 steps ahead forecast regression							Average
$RV(5)$	0.407	0.444	0.431	0.425	0.430	0.396	0.427	0.423
$RV(15)$	0.328	0.362	0.351	0.345	0.353	0.314	0.349	0.343
$OSRV$	0.380	0.416	0.405	0.397	0.404	0.367	0.401	0.396
$TSRV$	0.419	0.461	0.450	0.442	0.450	0.416	0.438	0.439
$RKERN$	0.395	0.438	0.425	0.416	0.430	0.379	0.420	0.415
BV	0.460	0.491	0.483	0.481	0.478	0.468	0.469	**0.476**
$ALTM$	0.379	0.422	0.407	0.396	0.407	0.360	0.403	0.397
Average	0.396	**0.434**	0.422	0.414	0.422	0.386	0.415	

Table 3:

R^2 for the univariate regressions using **indirect daily returns-based** out-of-sample forecasts (of which (38) is an example). All variables are expressed as logarithmic variance quantities and $h-$step ahead forecasts are produced used GARCH-type specifications (e.g. $GARCH_{t,h}$) for all measures. $p = q = 1$ in all forecasting models.

$Forecast\ Model:$	$GARCH$	$TGARCH$	$APARCH$	$FIGARCH$	
$V_{t+1,1}$	R^2 from 1 step ahead forecast regression				Average
$RV(15)$	0.288	0.287	0.268	0.264	0.277
$TSRV$	0.351	0.352	0.331	0.330	0.341
BV	0.352	0.352	0.332	0.335	**0.343**
Average	**0.330**	**0.330**	0.311	0.310	
$V_{t+1,10}$	R^2 from 10 steps ahead forecast regression				Average
$RV(15)$	0.363	0.368	0.354	0.347	0.358
$TSRV$	0.438	0.441	0.414	0.422	0.429
BV	0.455	0.456	0.432	0.444	**0.447**
Average	0.419	**0.422**	0.400	0.404	
$V_{t+1,22}$	R^2 from 22 steps ahead forecast regression				Average
$RV(15)$	0.361	0.370	0.324	0.338	0.348
$TSRV$	0.439	0.447	0.350	0.423	0.415
BV	0.465	0.471	0.379	0.450	**0.441**
Average	0.422	**0.429**	0.351	0.404	

Table 4:

R^2 for the univariate regressions using **direct** and **indirect** out-of-sample forecasts, including the option-implied indirect forecasts as exemplified by (39). The measures for which the direct ARFIMA(0,d,1) forecasting models are constructed are noted in brackets in the second line of the table. All variables are expressed as logarithmic variance quantities.

$Model:$	$ARFIMA$ $(RV(15))$	$ARFIMA$ $(TSRV)$	$ARFIMA$ (BV)	$TGARCH$	IV^{MF}	IV^{BS}	
$V_{t+1,1}$		R^2 from 1 step ahead forecast regression					Average
$RV(15)$	0.297	0.301	0.281	0.287	0.265	0.266	0.283
$TSRV$	0.381	0.388	0.368	0.352	0.326	0.318	0.356
BV	0.386	0.399	0.390	0.352	0.316	0.303	**0.358**
Average	0.355	**0.363**	0.346	0.330	0.302	0.296	
$V_{t+1,10}$		R^2 from 10 steps ahead forecast regression					Average
$RV(15)$	0.357	0.355	0.320	0.368	0.358	0.344	0.350
$TSRV$	0.455	0.449	0.419	0.441	0.435	0.410	0.435
BV	0.481	0.481	0.465	0.456	0.423	0.395	**0.450**
Average	**0.431**	0.428	0.401	0.422	0.405	0.383	
$V_{t+1,22}$		R^2 from 22 steps ahead forecast regression					Average
$RV(15)$	0.347	0.332	0.296	0.370	0.383	0.365	0.349
$TSRV$	0.452	0.435	0.402	0.447	0.458	0.432	0.438
BV	0.489	0.474	0.455	0.471	0.465	0.436	**0.465**
Average	0.429	0.414	0.384	0.429	**0.435**	0.411	

Table 5:

Results for regressions using direct and indirect one step and 22 steps ahead out-of-sample forecasts of RV(15), including the encompassing regressions as exemplified by (40). The measure from which the direct ARFIMA(0,d,1) forecast is constructed is indicated in parentheses. All variables are expressed as logarithmic variance quantities. The p-values reported in parentheses beneath each parameter estimate are for the two-sided test of the null hypothesis that the parameter equals zero. The null hypothesis associated with the $p-$values in the last column is detailed at the top of the table.

	α	β_1	β_2	R^2	$H_0 : \beta_1 = 1$
$V_{t+1,1}= RV(15)$					
					p-value
$ARFIMA\ (TSRV)$	0.486 (0.001)	1.151 (0.000)		0.301	0.008
$TGARCH$	-0.899 (0.000)	0.751 (0.000)	-	0.287	0.000
IV^{MF}	-0.354 (0.016)	1.115 (0.000)	-	0.265	0.130
IV^{BS}	-0.427 (0.003)	1.060 (0.000)	-	0.266	0.408
$ARFIMA + TGARCH$	0.106 (0.588)	0.727 (0.000)	0.339 (0.000)	0.304	-
$ARFIMA + IV^{MF}$	0.423 (0.006)	0.736 (0.000)	0.529 (0.000)	0.310	-
$TGARCH + IV^{MF}$	-0.417 (0.003)	0.433 (0.000)	0.605 (0.000)	0.300	-
$V_{t+1,22}= RV(15)$					
					p-value
$ARFIMA\ (RV(15))$	-0.058 (0.882)	0.814 (0.000)	-	0.347	0.184
$TGARCH$	-0.959 (0.000)	0.584 (0.000)	-	0.370	0.000
IV^{MF}	-0.632 (0.008)	0.822 (0.000)	-	0.383	0.111
IV^{BS}	-0.723 (0.003)	0.762 (0.000)	-	0.365	0.033
$ARFIMA + TGARCH$	-0.549 (0.241)	0.320 (0.226)	0.381 (0.018)	0.378	-
$ARFIMA + IV^{MF}$	0.319 (0.334)	0.273 (0.128)	0.610 (0.000)	0.394	-
$TGARCH + IV^{MF}$	-0.621 (0.008)	0.249 (0.042)	0.551 (0.002)	0.404	-

Table 6:

Results for regressions using direct and indirect one step and 22 steps ahead out-of-sample forecasts of TSRV, including the encompassing regressions as exemplified by (40). The measure from which the direct ARFIMA(0,d,1) forecast is constructed is indicated in parentheses. All variables are expressed as logarithmic variance quantities. The p-values reported in parentheses beneath each parameter estimate are for the two-sided test of the null hypothesis that the parameter equals zero. The null hypothesis associated with the $p-$values in the last column is detailed at the top of the table.

	α	β_1	β_2	R^2	$H_0 : \beta_1 = 1$
$V_{t+1,1}= TSRV$					
					p-value
$ARFIMA\ (TSRV)$	0.047 (0.654)	0.995 (0.000)	-	0.388	0.902
$TGARCH$	-1.184 (0.000)	0.633 (0.000)	-	0.352	0.000
IV^{MF}	-0.739 (0.000)	0.933 (0.000)	-	0.326	0.289
IV^{BS}	-0.826 (0.000)	0.873 (0.000)	-	0.318	0.031
$ARFIMA + TGARCH$	-0.220 (0.094)	0.698 (0.000)	0.238 (0.000)	0.392	-
$ARFIMA + IV^{MF}$	0.002 (0.985)	0.702 (0.000)	0.374 (0.000)	0.398	-
$TGARCH + IV^{MF}$	-0.793 (0.000)	0.374 (0.000)	0.492 (0.000)	0.372	-
$V_{t+1,22}= TSRV$					
					p-value
$ARFIMA\ (RV(15))$	-0.436 (0.126)	0.739 (0.000)	-	0.452	0.010
$TGARCH$	-1.295 (0.000)	0.510 (0.000)	-	0.447	0.000
IV^{MF}	-1.025 (0.000)	0.710 (0.000)	-	0.458	0.000
IV^{BS}	-1.110 (0.000)	0.654 (0.000)	-	0.432	0.000
$ARFIMA + TGARCH$	-0.737 (0.026)	0.435 (0.015)	0.234 (0.034)	0.456	-
$ARFIMA + IV^{MF}$	-0.623 (0.012)	0.351 (0.007)	0.437 (0.000)	0.489	-
$TGARCH + IV^{MF}$	-1.010 (0.000)	0.232 (0.012)	0.457 (0.000)	0.489	-

Table 7:

Results for regressions using direct and indirect one step and 22 steps ahead out-of-sample forecasts of BV, including the encompassing regressions as exemplified by (40). The measure from which the direct ARFIMA(0,d,1) forecast is constructed is indicated in parentheses. All variables are expressed as logarithmic variance quantities. The p-values reported in parentheses beneath each parameter estimate are for the two-sided test of the null hypothesis that the parameter equals zero. The null hypothesis associated with the $p-$values in the last column is detailed at the top of the table.

	α	β_1	β_2	R^2	$H_0 : \beta_1 = 1$
$V_{t+1,1}= BV$					
					p-value
$ARFIMA\ (TSRV)$	0.111	1.059	-	0.399	0.160
	(0.315)	(0.000)			
$TGARCH$	−1.221	0.663	-	0.352	0.000
	(0.000)	(0.000)			
IV^{MF}	−0.788	0.960	-	0.316	0.561
	(0.000)	(0.000)			
IV^{BS}	−0.892	0.890	-	0.303	0.089
	(0.000)	(0.000)			
$ARFIMA + TGARCH$	−0.133	0.787	0.217	0.405	-
	(0.329)	(0.000)	(0.000)		
$ARFIMA + IV^{MF}$	0.074	0.816	0.310	0.406	-
	(0.532)	(0.000)	(0.000)		
$TGARCH + IV^{MF}$	−0.848	0.417	0.469	0.370	-
	(0.000)	(0.000)	(0.000)		
$V_{t+1,22}= BV$					
					p-value
$ARFIMA\ (RV(15))$	−0.180	0.855	-	0.489	0.176
	(0.550)	(0.000)			
$TGARCH$	−1.180	0.589	-	0.471	0.000
	(0.000)	(0.000)			
IV^{MF}	−0.915	0.793	-	0.465	0.021
	(0.000)	(0.000)			
IV^{BS}	−1.014	0.729	-	0.436	0.002
	(0.000)	(0.000)			
$ARFIMA + TGARCH$	−0.510	0.523	0.256	0.492	-
	(0.145)	(0.005)	(0.020)		
$ARFIMA + IV^{MF}$	−0.358	0.486	0.416	0.513	-
	(0.196)	(0.000)	(0.000)		
$TGARCH + IV^{MF}$	−0.902	0.312	0.454	0.508	-
	(0.000)	(0.001)	(0.001)		

Table 8:

R^2 for the univariate regressions using **direct** and **indirect** out-of-sample forecasts, including the option-implied indirect forecasts as exemplified by (39). The measures for which the direct ARFIMA(0,d,1) forecasting models are constructed are noted in brackets in the second line of the table. All variables are expressed as **variance** quantities.

$Model:$	$ARFIMA$ $(RV(15))$	$ARFIMA$ $(TSRV)$	$ARFIMA$ (BV)	$TGARCH$	IV^{MF}	IV^{BS}	
$V_{t+1,1}$	R^2 from 1 step ahead forecast regression						Average
$RV(15)$	0.083	0.092	0.080	0.108	0.146	0.160	0.112
$TSRV$	0.083	0.243	0.224	0.227	0.287	0.294	0.226
BV	0.225	0.246	0.239	0.220	0.269	0.280	**0.247**
Average	0.130	0.194	0.181	0.185	0.234	**0.245**	
$V_{t+1,10}$	R^2 from 10 steps ahead forecast regression						Average
$RV(15)$	0.198	0.199	0.165	0.194	0.285	0.275	0.308
$TSRV$	0.339	0.346	0.311	0.306	0.426	0.404	**0.338**
BV	0.342	0.359	0.345	0.300	0.386	0.368	0.329
Average	0.293	0.301	0.274	0.267	**0.366**	0.349	
$V_{t+1,22}$	R^2 from 22 steps ahead forecast regression						Average
$RV(15)$	0.237	0.219	0.172	0.214	0.333	0.319	0.249
$TSRV$	0.359	0.343	0.296	0.301	0.439	0.414	0.359
BV	0.373	0.359	0.322	0.304	0.414	0.390	**0.360**
Average	0.323	0.307	0.263	0.273	**0.395**	0.374	

Table 9:

Results for regressions using direct and indirect one step and 22 steps ahead out-of-sample forecasts of RV(15), including the encompassing regressions as exemplified by (40). The measure from which the direct ARFIMA(0,d,1) forecast is constructed is indicated in parentheses. All variables are expressed as **variance** quantities. The *p*-values reported in parentheses beneath each parameter estimate are for the two-sided test of the null hypothesis that the parameter equals zero. The null hypothesis associated with the $p-$values in the last column is detailed at the top of the table.

	α	β_1	β_2	R^2	$H_0 : \beta_1 = 1$
$V_{t+1,1}= RV(15)$					
					p-value
$ARFIMA\ (TSRV)$	-0.023 (0.151)	1.858 (0.000)	-	0.092	0.000
$TGARCH$	0.027 (0.207)	0.693 (0.000)	-	0.108	0.050
IV^{MF}	-0.078 (0.024)	1.288 (0.000)	-	0.146	0.237
IV^{BS}	-0.072 (0.044)	1.275 (0.000)	-	0.160	0.280
$ARFIMA + TGARCH$	-0.005 (0.735)	0.751 (0.216)	0.492 (0.095)	0.114	-
$ARFIMA + IV^{MF}$	-0.077 (0.013)	-0.076 (0.870)	1.322 (0.002)	0.146	-
$TGARCH + IV^{MF}$	-0.070 (0.037)	0.154 (0.528)	1.093 (0.002)	0.148	-
$V_{t+1,22}= RV(15)$					
					p-value
$ARFIMA\ (RV(15))$	0.027 (0.202)	1.339 (0.000)	-	0.237	0.284
$TGARCH$	0.075 (0.000)	0.396 (0.000)	-	0.214	0.000
IV^{MF}	0.012 (0.526)	0.731 (0.000)	-	0.333	0.042
IV^{BS}	0.023 (0.202)	0.677 (0.000)	-	0.319	0.012
$ARFIMA + TGARCH$	0.037 (0.104)	0.914 (0.067)	0.158 (0.164)	0.247	-
$ARFIMA + IV^{MF}$	0.012 (0.539)	0.027 (0.933)	0.721 (0.000)	0.333	-
$TGARCH + IV^{MF}$	0.013 (0.552)	0.009 (0.929)	0.721 (0.001)	0.333	-

Table 10:

Results for regressions using direct and indirect one step and 22 steps ahead out-of-sample forecasts of TSRV, including the encompassing regressions as exemplified by (40). The measure from which the direct ARFIMA(0,d,1) forecast is constructed is indicated in parentheses. All variables are expressed as **variance** quantities. The p-values reported in parentheses beneath each parameter estimate are for the two-sided test of the null hypothesis that the parameter equals zero. The null hypothesis associated with the $p-$values in the last column is detailed at the top of the table.

	α	β_1	β_2	R^2	$H_0:\beta_1=1$
$V_{t+1,1}=TSRV$					
					p-value
$ARFIMA\ (TSRV)$	-0.002 (0.827)	1.253 (0.000)	-	0.243	0.009
$TGARCH$	0.040 (0.000)	0.418 (0.000)	-	0.227	0.000
IV^{MF}	-0.019 (0.135)	0.752 (0.000)	-	0.287	0.006
IV^{BS}	-0.012 (0.358)	0.719 (0.000)	-	0.294	0.002
$ARFIMA+TGARCH$	0.006 (0.419)	0.786 (0.000)	0.207 (0.017)	0.265	-
$ARFIMA+IV^{MF}$	-0.024 (0.035)	0.439 (0.009)	0.556 (0.000)	0.298	-
$TGARCH+IV^{MF}$	-0.013 (0.335)	0.125 (0.109)	0.594 (0.000)	0.295	-
$V_{t+1,22}=TSRV$					
					p-value
$ARFIMA\ (RV(15))$	0.028 (0.005)	0.958 (0.000)	-	0.359	0.794
$TGARCH$	0.064 (0.000)	0.273 (0.000)	-	0.301	0.000
IV^{MF}	0.023 (0.012)	0.488 (0.000)	-	0.439	0.000
IV^{BS}	0.031 (0.000)	0.449 (0.000)	-	0.414	0.000
$ARFIMA+TGARCH$	0.033 (0.003)	0.748 (0.003)	0.078 (0.193)	0.366	-
$ARFIMA+IV^{MF}$	0.020 (0.037)	0.235 (0.183)	0.397 (0.000)	0.445	-
$TGARCH+IV^{MF}$	0.025 (0.014)	0.029 (0.595)	0.454 (0.000)	0.440	-

Table 11:

Results for regressions using direct and indirect one step and 22 steps ahead out-of-sample forecasts of BV, including the encompassing regressions as exemplified by (40). The measure from which the direct ARFIMA(0,d,1) forecast is constructed is indicated in parentheses. All variables are expressed as **variance** quantities. The p-values reported in parentheses beneath each parameter estimate are for the two-sided test of the null hypothesis that the parameter equals zero. The null hypothesis associated with the $p-$values in the last column is detailed at the top of the table.

	α	β_1	β_2	R^2	$H_0 : \beta_1 = 1$
$V_{t+1,1}$= BV					
					p-value
$ARFIMA\ (TSRV)$	-0.011 (0.166)	1.273 (0.000)	-	0.246	0.014
$TGARCH$	0.033 (0.000)	0.415 (0.000)	-	0.220	0.000
IV^{MF}	-0.024 (0.100)	0.735 (0.000)	-	0.269	0.011
IV^{BS}	-0.017 (0.237)	0.708 (0.000)	-	0.280	0.005
$ARFIMA + TGARCH$	-0.004 (0.647)	0.853 (0.000)	0.187 (0.042)	0.264	-
$ARFIMA + IV^{MF}$	-0.031 (0.020)	0.569 (0.003)	0.482 (0.002)	0.286	-
$TGARCH + IV^{MF}$	-0.017 (0.263)	0.140 (0.082)	0.559 (0.000)	0.279	-
$V_{t+1,22}$= BV					
					p-value
$ARFIMA(RV(15))$	0.016 (0.109)	1.071 (0.000)	-	0.373	0.666
$TGARCH$	0.057 (0.000)	0.301 (0.000)	-	0.304	0.000
IV^{MF}	0.016 (0.128)	0.520 (0.000)	-	0.414	0.000
IV^{BS}	0.024 (0.013)	0.478 (0.000)	-	0.390	0.000
$ARFIMA + TGARCH$	0.021 (0.062)	0.874 (0.001)	0.073 (0.247)	0.379	-
$ARFIMA + IV^{MF}$	0.009 (0.386)	0.423 (0.030)	0.356 (0.001)	0.431	-
$TGARCH + IV^{MF}$	0.019 (0.100)	0.059 (0.382)	0.450 (0.000)	0.418	-