Long Memory in Volatility and Trading Volume[*]

Jeff Fleming[a], Chris Kirby[b]

[a]*Jesse H. Jones Graduate School of Management, Rice University, Houston, TX 77005*
[b]*John E. Walker Department of Economics, Clemson University, Clemson, SC 29634*

Abstract

We use fractionally-integrated time-series models to investigate the joint dynamics of equity trading volume and volatility. Bollerslev and Jubinski (1999) show that volume and volatility have a similar degree of fractional integration, and they argue that this evidence supports a long-run view of the mixture-of-distributions hypothesis. We examine this issue using more precise volatility estimates obtained using high-frequency returns (i.e., realized volatilities). Our results indicate that volume and volatility both display long memory, but we can reject the hypothesis that the two series share a common order of fractional integration for a fifth of the firms in our sample. Moreover, we find a strong correlation between the innovations to volume and volatility, which suggests that trading volume can be used to obtain more precise estimates of daily volatility for cases in which high-frequency returns are unavailable.

Key words: realized volatility, fractional integration, strongly autocorrelated, bivariate mixture models, long-range dependence

July 18, 2006

Long Memory in Volatility and Trading Volume

ABSTRACT

We use fractionally-integrated time-series models to investigate the joint dynamics of equity trading volume and volatility. Bollerslev and Jubinski (1999) show that volume and volatility have a similar degree of fractional integration, and they argue that this evidence supports a long-run view of the mixture-of-distributions hypothesis. We examine this issue using more precise volatility estimates obtained using high-frequency returns (i.e., realized volatilities). Our results indicate that volume and volatility both display long memory, but we can reject the hypothesis that the two series share a common order of fractional integration for a fifth of the firms in our sample. Moreover, we find a strong correlation between the innovations to volume and volatility, which suggests that trading volume can be used to obtain more precise estimates of daily volatility for cases in which high-frequency returns are unavailable.

Long Memory in Volatility and Trading Volume

There is a consensus among financial econometricians that volatility is characterized by long memory. The consensus began to take shape with reports of hyperbolic decay in the autocorrelations of stock index and currency absolute returns (Taylor, 1986; Ding et al., 1993; Dacorogna et al., 1993). It gained momentum as fractionally-integrated GARCH models made inroads into the volatility modeling literature (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996). More recently, studies of realized volatility (i.e., cumulative squared intraday returns), such as Andersen, Bollerslev, Diebold, and Ebens (henceforth ABDE) (2001), Andersen, Bollerslev, Diebold, and Labys (henceforth ABDL) (2001), and Andersen et al. (2003) have produced compelling evidence of long memory for both equities and currencies. Realized volatility appears to display all the hallmarks of a fractionally-integrated process with a degree of fractional integration in the 0.3 to 0.5 range.

Despite this finding there is an ongoing debate about whether it is appropriate to model volatility as fractionally integrated. Other mechanisms such as structural breaks or regime-switching in volatility can also give rise to long-range dependence in returns (Diebold and Inoue, 2001; Granger and Hyung, 2004), and this has led some researchers to argue that structural change can produce spurious evidence of long memory. Diebold and Inoue (2001), however, caution against this argument and show why it is potentially naive. They discuss a setting in which structural change and long memory are essentially just different labels for the same phenomenon, and conclude that calling one "real" and the other "spurious" makes little sense. Even if the "real" process includes structural change, fractionally-integrated models provide a useful description of volatility dynamics because they effectively allow the unconditional variance to slowly change over time. This view is reinforced by Hyung et al. (2006) who point out the difficulty of forecasting volatility breaks and find that, in the absence of a way to identify the breaks before they occur, fractionally-integrated models provide the best volatility forecasts.

Bollerslev and Jubinski (1999) build on the evidence of long memory in volatility to investigate the extent to which volume and volatility share common long-run dependencies. A theoretical result called the mixture-of-distributions hypotheses (MDH) predicts that returns and trading volume are jointly subordinate to an unobserved directing variable that measures the rate at which information arrives to the market (Tauchen and Pitts, 1983). Although early tests of the MDH were generally supportive, subsequent

work using more sophisticated methods revealed that a single latent directing variable cannot account for the observed short-run dynamics of volume and volatility. In light of this finding, and the evidence of long memory in volume as well as volatility, Bollerslev and Jubinski (1999) argue that the focus on short-run dynamics may be misplaced. They propose that the MDH might be better viewed as long-run proposition.

To test this supposition, they use a semiparametric frequency-domain approach to estimate the fractional order of integration of the absolute return and trading volume series for each firm in the S&P 100 index. Their results suggest that volume and volatility have remarkably similar memory characteristics. The estimates of the integration order for the absolute returns have a mean value of 0.41, while the estimates for detrended trading volume have a mean value of 0.40. Moreover, only eight of the 100 firms in the S&P index produce evidence against the hypothesis that volume and volatility share a common order of fractional integration. This finding differs sharply with the findings of studies that focus on the short-run dynamics of volume and volatility, leading Bollerslev and Jubinski (1999) to conclude that "allowing for differing short-lived news impacts, while imposing a common long-memory component, may provide a better characterization of the joint volume-volatility relationship in U.S. equity markets."

Our paper investigates the empirical performance of a model that is structured along the lines suggested by Bollerslev and Jubinski (1999). Specifically, we consider a trend-stationary fractionally-integrated model for volume and volatility that allows for VAR(p) dynamics after fractional differencing. The idea is to strike a middle ground between the semiparametric estimation approach of Bollerslev and Jubinski (1999) and the type of fully parametric approach that would be needed to incorporate all the structural and distributional assumptions of the MDH. Although we assume enough parametric structure to allow the long- and short-memory components to compete on an equal footing, we do not impose a common order of fractional integration, nor do we specify the joint distribution of the errors. This makes it easy to assess the relative importance of each component in explaining the joint dynamics of volume and volatility, while reducing the potential for model misspecification.

The dataset for our empirical analysis consists of daily observations of realized volatility and trading volume for the 20 firms in the Major Market Index (MMI). Using

realized volatilities rather than absolute returns should lead to more precise inferences.[1] We construct the realized volatilities for the full day, as suggested by Hansen and Lunde (2005), by choosing the linear combination of the trading-day realized variance and the overnight squared return that is the most efficient estimator of the integrated variance. This accounts for the possibility that the overnight return contains relevant information and the trading-day realized volatility is a biased estimator of the full-day integrated volatility. We construct the trading-day realized variance using Newey and West (1987) weights to account for serial correlation in the intraday returns induced by microstructure effects, and we determine the optimal combination of the sampling frequency for the intraday returns and the lag truncation parameter for the weights by evaluating the effects of these choices on the bias and efficiency of the resulting estimator.

When we fit our model to the log volatility and log volume series for each firm, the results clearly support the view that volatility and volume display long memory: the average estimated order of fractional integration is about 0.35 in both cases. Nonetheless, the evidence on whether the two series share a common fractional order of integration is mixed. We can reject this hypothesis for four of the 20 firms at the 5 percent significance level. Interestingly, fractional differencing removes most of the predictability in the data. The estimated first-order autocorrelation of the fractionally-differenced series ranges from -0.05 to -0.12 for the volatility series and -0.01 to 0.22 for the volume series. It appears, therefore, that the short-memory components play a relatively minor role in explaining the dynamics of realized volatility and trading volume.

We also find that the innovations in the two series are highly correlated. This suggests that volume shocks are informative about contemporaneous volatility shocks, raising the possibility that volume can be used in conjunction with daily returns to obtain more precise volatility estimates for cases in which the high-frequency return data needed to construct realized volatilities are unavailable. Our analysis indicates that the increase in precision could be substantial. For example, if we regress the log realized variance on estimates of the variance that condition on lagged information (i.e., past realized volatility and volume observations) plus contemporaneous volume, the mean-squared-error is 30 percent lower on average than the MSE obtained using variance estimates that condition on lagged information alone.

---

[1] Luu and Martens (2003) use realized volatilities to estimate an econometric model similar to Bollerslev and Jubinski (1999).

The remainder of the paper is organized as follows. Section 1 introduces our long-memory specification and discusses time-domain methods of estimation and inference. Section 2 describes the dataset and details of our construction of the realized volatilities. Section 3 presents the model fitting results. Section 4 investigates the information content of volume innovations for volatility forecasting. Section 5 offers some concluding remarks.

1 Model and Econometric Methodology

Let δ denote the memory parameter of a fractionally-integrated process. It is often convenient to estimate δ using frequency-domain methods. Perhaps the most widely used approach is the log periodogram regression first proposed by Geweke and Porter-Hudak (1983). Indeed, Bollerslev and Jubinski (1999) use a bivariate log periodogram regression in their study of long memory in volume and volatility. They motivate this approach by noting that the log-periodogram estimator is much simpler to implement than the exact maximum likelihood estimator and that it imposes minimal structure beyond that required to ensure that the spectral density is well behaved. In our case, however, the drawbacks of the estimator outweigh its advantages.

First, the log-periodogram estimator of δ converges at less than a $\sqrt{T}$ rate. This raises concerns about efficiency because the high-frequency data needed to construct the realized volatilities are unavailable before 1993. The efficiency of the log-periodogram estimator varies somewhat depending on the data generating process and the choice of truncation parameter that determines the Fourier frequencies used to construct the estimator, but we know that it is considerably less efficient than parametric alternatives. For example, Bollerslev and Jubinski (1999) use truncation parameters that are proportional to $\sqrt{T}$. As a result, their estimator achieves only a $T^{1/4}$ rate of convergence under the assumption of Gaussian observations.

Second, the asymptotic theory for the log-periodogram estimator is not well established for cases in which the observations used to compute the periodogram are non-Gaussian. This is important because we want to compare the results of estimating δ using the realized volatilities to the results obtained using absolute returns. Deo and Hurvich (2001) study the asymptotics of the log-periodogram estimator based on the first m Fourier frequencies for the case in which returns are generated by a simple univariate stochastic volatility model. They find that for consistency and asymptotic normality to obtain, the

4

value of m must increase at a rate less than $T^{4\delta/(1+4\delta)}$. Thus, unlike in the Gaussian case, the choice of an appropriate m depends on the (unknown) value of δ. This suggests that robustness is a concern if the data exhibit clear departures from normality.

Finally, the log-periodogram estimator focuses exclusively on long-run frequencies. Bollerslev and Jubinski (1999) argue that this is an advantage because it allows them to estimate δ without making any assumptions about the short-run dynamics of the process. However, our objective is to provide evidence on both the short- and long-run relations between volume and volatility. In order to accomplish this, we need to use a more flexible approach that allows the short- and long-memory components of the process to compete on an equal footing. The methodology we adopt represents a middle ground between the semiparametric framework of the log-periodogram estimator and the type of fully parametric approach that would be needed to incorporate all the structural and distributional assumptions of the MDH. We turn now to the details.

1.1 Multivariate linear regression with fractionally-integrated errors

Suppose for generality that we observe an $N \times 1$ vector time series $\{y_t\}_{t=1}^T$ that is described by the multivariate linear regression model

$$y_t = \beta x_t + \varepsilon_t, \tag{1}$$

where x_t is a $K \times 1$ vector of deterministic components, β is an unknown $N \times K$ matrix, and ε_t is an unobserved $N \times 1$ error vector. Let η_t denote an $N \times 1$ vector of white noise innovations with $\mathrm{E}[\eta_t \eta_t'] = \Omega$. We focus on specifications in which the regression errors are generated by a process of the form

$$\Phi(L)\Delta(L)\varepsilon_t = \eta_t 1(t > 0), \tag{2}$$

where $\Phi(L)$ is a $N \times N$ matrix polynomial of order p in the lag operator L with all roots outside the unit circle, $\Delta(L)$ is an $N \times N$ diagonal matrix with nth diagonal element

$$(1 - L)^{\delta_n} = \sum_{j=0}^{\infty} \binom{\delta_n}{j} (-L)^j, \tag{3}$$

and $1(\cdot)$ is the indicator function. In other words, after fractional differencing, the errors are described by a VAR(p) process with white noise innovations. We assume that $0 \leq \delta_n < 0.5$ for all n. Under this restriction, the error process is asymptotically stationary and displays long memory unless all elements of $\delta = (\delta_1, \ldots, \delta_N)'$ are zero.

The regression model in equation (1) allows for deterministic trends that are independent of the long-memory characteristics of the errors. This conforms with both predictions of market microstructure theories and empirical evidence. Trading volume often displays strong time trends that are qualitatively consistent with the theoretical consequences of growth in the number of traders in the market (see, e.g., Tauchen and Pitts (1983)), and a number of recent studies suggest that the volatility of individual stock returns has increased over time (e.g., Campbell et al. (2001)). Since most of the models studied in the volume and volatility literature assume an absence of time trends, it is common to detrend the data before conducting the econometric analysis (see, e.g., Gallant et al. (1992), Andersen (1996), and Bollerslev and Jubinski (1999)). We explicitly consider the potential impact of this detrending on the procedures used for estimation and inference.

1.2 Model fitting and hypothesis testing

First consider the case in which $\beta = 0$. Let $\theta = (\delta', \phi')'$ where $\phi = \text{vec}(\Phi_1 \Phi_2 \cdots \Phi_N)$ is an $Np \times 1$ vector that contains the VAR(p) parameters. After concentrating out Ω, we can express the log likelihood function for a model with Gaussian η_t as

$$\mathcal{L}_T(\theta) = -\frac{TN}{2} - \frac{TN}{2}\log(2\pi) - \frac{T}{2}\log|\Omega(\theta)|, \tag{4}$$

where

$$\Omega(\theta) = \frac{1}{T}\sum_{t=1}^{T}[\Phi(L)\Delta(L)\varepsilon_t][\Phi(L)\Delta(L)\varepsilon_t]'. \tag{5}$$

If η_t is non-Gaussian, then $\mathcal{L}_T(\theta)$ is a quasi-log likelihood function. Intuitively, quasi-maximum likelihood should be a reasonable estimation strategy provided that suitable regularity conditions are satisfied.

Nielsen (2004b), for example, analyzes the properties of the QML estimator under the assumption that η_t is i.i.d. with positive-definite covariance matrix and finite fourth moments. He shows that there exists a local maximizer $\hat{\theta} = (\hat{\delta}', \hat{\varphi}')'$ of $\mathcal{L}(\theta)$ that is a $\sqrt{T}$-consistent and asymptotical normal estimator of θ. Specifically, we have

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma), \tag{6}$$

with

$$\Sigma = \begin{bmatrix} \frac{\pi^2}{6}\Omega \odot \Omega^{-1} & J'(\Lambda' \otimes I) \\ (\Lambda \otimes I)J & \Gamma \otimes \Omega^{-1} \end{bmatrix}^{-1}, \tag{7}$$

where Γ is the covariance matrix of $(\varepsilon_t', \ldots, \varepsilon_{t-p+1}')'$, $J = (\text{vec}\, e_1 e_1', \ldots, \text{vec}\, e_N e_N')$ with e_i denoting the ith unit N-vector, and $\Lambda = (\Lambda_1', \ldots, \Lambda_p')'$ with $\Lambda_i = \sum_{j=i}^{\infty} j^{-1}\Psi_{j-i}$ and Ψ_i denoting the coefficient matrix on ε_{t-i} in the Wold representation of $\{\varepsilon_t\}$.

In light of these results, testing whether δ is equal to a known vector is straightforward. Consider a null hypothesis of the form $H_0 : \delta = \delta_0$ where δ_0 is a prespecified $N \times 1$ vector and let $\tilde{\phi}$ denote the estimate of ϕ obtained by imposing H_0. The quasi-likelihood ratio (LR), Wald (W), and Lagrange multiplier (LM) statistics are

$$LR = 2(\mathcal{L}_T(\hat{\delta}, \hat{\phi}) - \mathcal{L}_T(\delta_0, \tilde{\phi})), \tag{8}$$

$$W = T(\hat{\delta} - \delta_0)'(\tfrac{\pi^2}{6}\Omega \odot \Omega^{-1} - (\Omega\Lambda'\Gamma^{-1}\Lambda\Omega) \odot \Omega^{-1})(\hat{\delta} - \delta_0) \tag{9}$$

$$LM = S_T(\delta_0, \tilde{\phi})'(\tfrac{\pi^2}{6}\Omega \odot \Omega^{-1} - (\Omega\Lambda'\Gamma^{-1}\Lambda\Omega) \odot \Omega^{-1})S_T(\delta_0, \tilde{\phi}) \tag{10}$$

where $S_T(\delta_0, \tilde{\phi}) = \sqrt{T}\sum_{j=1}^{T-1} j^{-1}J'(I \otimes \Omega(\delta_0, \tilde{\phi})^{-1})(\text{vec}\, \tilde{C}(j))$ with $\tilde{C}(j) = (1/T)\sum_{t=j+1}^{T}\tilde{\eta}_t \tilde{\eta}_{t-j}'$ denotes the score vector evaluated under H_0. Under a local alternative of the form $\delta = \delta_0 + d/\sqrt{T}$ with d a fixed N-vector, Nielsen (2004b) shows that the statistics in equations (8) – (10) are asymptotically equivalent and distributed as $\chi_N^2(\zeta)$ with $\zeta = d'(\tfrac{\pi^2}{6}\Omega \odot \Omega^{-1} - (\Omega\Lambda'\Gamma^{-1}\Lambda\Omega) \odot \Omega^{-1})d$. The LR statistic, which does not require estimation of the asymptotic covariance matrix of $\hat{\delta}$, is particularly convenient for applications in which there is no desire to compute standard errors.

We can also apply these results for the case in which $\beta \neq 0$ provided that we estimate β in a way that does not influence our asymptotic inference on δ. To illustrate, consider a univariate model with $H_0 : \delta = \delta_0$. Nielsen (2004a) shows that if we estimate β by

$$\hat{\beta} = \left(\sum_{t=1}^{T} y_t^* x_t^{*\prime}\right)\left(\sum_{t=1}^{T} x_t^* x_t^{*\prime}\right)^{-1}, \tag{11}$$

where $x_t^* = (1-L)^{\delta_0} x_t$ and $y_t^* = (1-L)^{\delta_0} y_t$, then we can treat the residual $y_t - \hat{\beta}x_t$, which equals $e_t + (\beta - \hat{\beta})x_t$ under H_0, as observed for the purpose of conducting asymptotic inference on δ. This presumably holds in the multivariate setting as well.

Of course, this strategy works only if δ_0 is prespecified. In our empirical application, we are mainly interested in testing hypotheses of the form H_0: $\delta = \delta_0\iota$ where δ_0 is an unspecified scalar and ι is an $N \times 1$ vector of ones, i.e., our null is that the observed series share a common unknown order of integration. One way to deal with this situation is to

estimate β by

$$\hat{\beta} = \left(\sum_{t=1}^{T} y_t x_t' \right) \left(\sum_{t=1}^{T} x_t x_t' \right)^{-1}, \tag{12}$$

and then estimate θ by maximizing

$$\mathcal{L}_T(\hat{\beta}, \theta) = -\frac{TN}{2} - \frac{TN}{2} \log(2\pi) - \frac{T}{2} \log |\Omega(\hat{\beta}, \theta)|, \tag{13}$$

where $\Omega(\hat{\beta}, \theta)$ is computed as in equation (5) except that we replace ε_t with the vector of residuals $\hat{\varepsilon}_t = y_t - \hat{\beta} x_t$. This is equivalent to extracting a trend via OLS regression prior to fitting the model to the data.

To justify this approach, we need to argue that replacing $\mathcal{L}_T(\beta, \theta)$ with $\mathcal{L}_T(\hat{\beta}, \theta)$ has no effect on the limiting distribution of $\hat{\theta}$. The analysis of Giraitis and Koul (1998) supports this argument. They consider maximum likelihood estimation of a multiple linear regression model in which the errors are a non-decreasing function of a stationary long-memory Gaussian process. For a fairly broad class of deterministic regressors (e.g., βx_t is a polynomial in t), the estimator of the parameter vector that determines the memory characteristics of errors is $\sqrt{T}$-consistent and asymptotically normal with an asymptotic covariance matrix that attains the Cramer-Rao lower bound. Since multiple linear regression with Gaussian long-memory errors is covered by these results, using OLS to detrend the data is asymptotically sound for $\eta_t \sim \text{NID}(0, \Omega)$. To our knowledge, the question of whether this holds for a Gaussian QML approach with i.i.d. innovations has not been addressed in the literature. But it seems likely that an extension is possible because the non-Gaussian log likelihoods studied by the Giraitis and Koul (1998) have a quadratic structure similar to that of the QML objective function.

2   Data

The dataset consists of daily realized volatilities and trading volumes for the 20 firms in the MMI.[2] We construct the dataset from two sources: intradaily transaction records from the Trade and Quote (TAQ) database of the New York Stock Exchange (NYSE), and information on stock splits, stock dividends, and cash dividends from the Center for

---

[2] These firms are American Express (AXP), AT&T (T), ChevronTexaco (CVX), Coca-Cola (KO), Disney (DIS), Dow Chemical (DOW), DuPont (DD), Eastman Kodak (EK), Exxon-Mobil (XOM), General Electric (GE), General Motors (GM), International Business Machines (IBM), International Paper (IP), Johnson & Johnson (JNJ), McDonald's (MCD), Merck (MRK), 3M (MMM), Philip Morris (MO), Procter and Gamble (PG), and Sears (S).

Research in Security Prices (CRSP) daily stock price file. The sample period is January 4, 1993 to December 31, 2003 (2,771 observations).[3]

We exclude records from the TAQ database that have an out-of-sequence time stamp, a zero price, a correction code greater than two (indicating errors and corrections), or a condition code (indicating nonstandard settlement). We also apply two filters to identify obvious price reporting errors. First, we exclude records that have a reported price more than 20 percent different than the previous transaction price. Second, we flag records that imply a price change greater than two percent in magnitude from the previous transaction price and which are followed by a price reversal greater than two percent. We exclude the record if the implied price change is more than two times greater than the next largest price change on that day, or if the reported price falls outside the day's high-low range (ignoring the flagged price) by more than the next largest price change.

We construct the daily trading volume for each firm by aggregating the transaction volumes for all of the remaining TAQ records on a given day. We adjust the daily volumes as necessary to account for stock splits and stock dividends using the information reported in the CRSP file. We construct the daily realized volatility for each firm using the filtered TAQ transaction prices and the procedure explained in Section 2.1. Implementing this procedure also requires the daily overnight return. We compute this return using the last transaction price on a given day, and the first transaction price on the following day, adjusted for cash dividends and stock distributions reported in the CRSP file.

2.1 Constructing the realized volatilities

Merton (1980) introduced the concept of realized volatility. In its simplest form, the realized variance on day t is the sum of the squared intradaily returns, $r_{t_{i,m}}$, $i = 1, ..., m$, over m equally-spaced intervals,

$$RV_t = \sum_{i=1}^{m} r_{t_{i,m}}^2. \tag{14}$$

Under certain conditions, the realized variance should be close to the true variance. For example, if returns are generated by a continuous-time process with instantaneous volatility σ_t, then it is natural to use the integrated variance $IV_t = \int_0^1 \sigma_{t+\tau}^2 d\tau$ as a measure of the daily variance. ABDL (2001) and Barndorff-Nielsen and Shephard (2002) show that

---

[3] Philip Morris has only 2,770 observations. Its stock did not open on May 25, 1994 in advance of a board meeting regarding a proposal to split the firm's food and tobacco businesses.

under weak regularity conditions, $RV_t - IV_t \to 0$ almost surely as $m \to \infty$. Thus, by increasing the sampling frequency of returns, we can construct consistent nonparametric estimates of the integrated volatility that in principle are arbitrarily efficient.

In practice, realized volatilities can be biased by market microstructure effects, and these effects are typically more pronounced as m increases. A number of strategies have been proposed to deal with this issue. Andersen and Bollerslev (1997) argue that using five-minute returns strikes a reasonable balance between the bias and the efficiency gains associated with more frequent sampling. Bandi and Russell (2005) derive the optimal sampling frequency under a mean squared error criterion. ABDE (2001) construct realized volatility using the residuals from an MA(1) model fitted to returns. ABDL (2001) use linear interpolation to estimate the prices used in constructing the returns. Each of these strategies, however, either places an upper bound on m or does not sufficiently eliminate the serial correlation in returns when m is large. We adopt an alternative approach suggested by Hansen and Lunde (2004) which directly accounts for serial correlation in returns when constructing the realized variances. Specifically, we construct the trading-day realized variance for day t using a Newey and West (1987) estimator of the form,

$$RV_{t[o]} = \sum_{i=1}^{m} r_{t_{i,m}}^2 + 2 \sum_{j=1}^{q} \left(1 - \frac{j}{q+1} \right) \sum_{i=1}^{m-j} r_{t_{i,m}} r_{t_{j,m}}, \tag{15}$$

where q denotes the window length for the autocovariance terms. This estimator guarantees nonnegativity and is consistent in the presence of serial correlation, which allows us to sample returns at very high frequencies.

The trading-day realized variance may be a biased estimator of the daily integrated variance because it ignores returns during non-trading periods overnight and on weekends. We construct the full-day realized variance by combining $RV_{t[o]}$ with the squared nontrading-period return, $r_{t[c]}^2$ using the weighting scheme proposed by Hansen and Lunde (2005). They consider the class of conditionally unbiased estimators that are linear in $RV_{t[o]}$ and $r_{t[c]}^2$ and show that the following weights deliver the lowest mean squared error

$$RV_t = \varphi \frac{\mu}{\mu_o} RV_{t[o]} + (1 - \varphi) \frac{\mu}{\mu_c} r_{t[c]}^2, \tag{16}$$

where

$$\varphi = \frac{\mu_o^2 \varsigma_c^2 - \mu_o \mu_c \varsigma_{oc}}{\mu_c^2 \varsigma_o^2 + \mu_o^2 \varsigma_c^2 - 2\mu_o \mu_c \varsigma_{oc}}, \tag{17}$$

and $\mu = \mathrm{E}(r_t^2)$, $\mu_o = \mathrm{E}(RV_{t[o]})$, $\mu_c = \mathrm{E}(r_{t[c]}^2)$, $\varsigma_o^2 = \mathrm{var}(RV_{t[o]})$, $\varsigma_c^2 = \mathrm{var}(r_{t[c]}^2)$, and $\varsigma_{oc} =$

$\text{cov}(RV_{t[o]}, r_{t[c]}^2)$.[4] The ratios μ/μ_o and μ/μ_c scale the variance estimates to match the unconditional mean of the squared close-to-close returns and φ determines the relative weights on the trading- and nontrading-period variance estimates. In general, φ should be close to one because $r_{t[c]}^2$ is a relatively imprecise estimator of the nontrading-period variance and this variance is typically lower than the variance during the trading day. These effects can easily be seen in equation (17) for the case in which $\varsigma_{oc} = 0$.

To implement equation (15), we need to specify the sampling frequency (m) and the window length for the autocovariance terms (q). Since the optimal values are not known a priori, we consider a range of candidate values and evaluate their effect on the properties of the realized variances. The candidate values for m correspond to returns sampled as finely as every 30 seconds, and the values for q correspond to four different window lengths: 0, 15, 30, and 60 minutes. We construct the returns used to implement equation (15) by applying the Andersen and Bollerslev (1997) linear interpolation scheme to the filtered intraday transaction prices from the TAQ database.

It is common practice to construct realized variances using five-minute returns and with $q = 0$. We find that realized variances constructed using this combination of parameter values are severely biased, with the average realized variance 13 percent greater than the average squared open-to-close return. The magnitude of bias increases rapidly with the sampling frequency, but increasing the window length counteracts the bias. Using a 30-minute window produces realized variances that, for any choice of sampling frequency, are within two percent of the average squared open-to-close return. Increasing the window length further (e.g., 60 minutes) substantially increases the standard deviation of the realized variances. Therefore, we conclude that the best combination of parameter values is a 30-second sampling frequency and a 30-minute window length.

We construct the full-day realized volatilities by substituting the sample analogs of μ, μ_o, μ_c, ς_o^2, ς_c^2, and ς_{oc} into equations (16) and (17). Hansen and Lunde (2005) suggest removing outliers from the estimation to avoid obtaining a negative weight on $r_{t[c]}^2$. Accordingly, we exclude days in which either $RV_{t[o]}$ or $r_{t[c]}^2$ is among the largest 0.5 percent

---

[4] Strictly speaking, this estimator is not optimal under a MSE criterion if we construct $RV_{t[o]}$ using Newey-West weights since this does not produce a conditionally unbiased estimator. As a practical matter, however, the magnitude of the bias appears to be small. Hansen and Lunde (2005) favor a different weighting scheme which leads to conditionally unbiased estimators, but it does so by admitting the possibility of negative estimates. Using their weighting scheme produces an average of over 40 negative variance estimates per stock over our sample period.

of the observations for a given stock. The average φ estimate for the 20 stocks is 0.92. By comparison, the ratio of the average squared close-to-close return to the average squared close-to-open return indicates that 20 percent of the daily variance occurs during the nontrading period. The φ estimate gives less weight than this to the nontrading-period variance estimate because the trading-period variance estimate is much more precise.

3 Model Fitting Results

We fit our long-memory model using three different proxies for the volatility of daily returns. First, to provide a set of baseline results, we fit the model with $y_t = (\log RV_t, \log V_t)'$ where RV_t is constructed with $m = 78$ and $q = 0$, the parameter choices most commonly employed in prior research (i.e., five-minute sampling frequency, without the Newey-West correction). We refer to this measure as the standard estimator of realized volatility. Next, we fit the model replacing RV_t with the estimator obtained using a 30-second sampling frequency and a 30-minute window length. We refer to this measure as the higher-frequency, biased-corrected (HFBC) estimator of realized volatility. Finally, we fit the model using daily squared returns in place of RV_t to investigate how the precision of the volatility estimates impacts our inference.

3.1 Results using the standard estimator

Table 1 reports the model fitting results obtained using the standard estimator of realized volatility. In parameterizing the model, we assume that a VAR(1) process is sufficient to capture the short-run dynamics of the fractionally-differenced series. Panel A of the table reports the memory parameter estimates and their asymptotic standard errors, the VAR coefficient estimates and their asymptotic t-ratios, the value of the quasi-log likelihood, and the likelihood ratio statistic for the hypothesis $\delta_1 = \delta_2$. Panel B reports the trend parameter estimates, the innovation covariance matrix estimates, and selected model diagnostics.

As expected, the estimates of δ_1 point to long memory in the volatility process. The mean of the estimates is 0.43 and the range is relatively narrow (0.39 for Eastman Kodak to 0.46 for IBM and Exxon-Mobil). This finding is broadly consistent with the evidence of Bollerslev and Jubinski (1999), who report that the absolute returns for the S&P 100 firms have a mean estimated order of fractional integration of 0.41. Of course, our memory parameter estimates should be much more precise. All of the standard errors in Table 1 are

12

equal to 0.02, but the standard error of the log-periodogram-based estimator is three times as large using the maximum sample size considered by Bollerslev and Jubinski (1999). [5] Thus, the QML estimator of δ_1 appears to have a substantial efficiency advantage.

The estimates of δ_2 also point to long memory in trading volume. However, the mean of the estimates (0.34) is lower than the mean for the volatility process and the range is wider as well (0.25 for Coca Cola to 0.39 for American Express, IBM, and AT&T). This finding represents a significant departure from Bollerslev and Jubinski (1999). They find that detrended trading volume has a mean estimated order of fractional integration of 0.40, while 13 of our estimates of δ_2 are more than two standard errors below 0.40. Moreover, they report only 8 rejections of the hypothesis $\delta_1 = \delta_2$ for the S&P 100 firms, while we reject this hypothesis for 18 of the 20 MMI firms. There are many factors, such as different estimators, sample periods, volatility measures, and firms, that could contribute to the disparity in findings. We provide more evidence in this regard shortly.

The estimates of Φ suggest that long memory is the main source of time variation in the conditional means of $\log RV_t$ and $\log V_t$, i.e., fractional differencing eliminates most of the serial correlation in the data. Our estimate of the first-order autocorrelation in $(1 - L)^{\delta_1} \log RV_t$ ranges from -0.05 (Procter & Gamble) to -0.15 (Disney), while the range for $(1 - L)^{\delta_2} \log V_t$ is 0.01 (American Express) to 0.22 (Coca Cola). Although most of the autocorrelation estimates are statistically significant at the 5 percent level, the level of predictability implied by the VAR(1) specification is small. Apparently the short-memory component of the model plays only a minor role in explaining the dynamics of realized volatility and trading volume.

Panel B of the table shows that most firms display some evidence of trends of one form or another. This follows by comparing the estimated means of $\log RV_t$ and $\log V_t$ at the start of the sample (i.e., $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$) to the estimated means at the end of the sample (i.e., $\sum_{i=1}^{3} \hat{\beta}_{1i}$ and $\sum_{i=1}^{3} \hat{\beta}_{2i}$). With the exception of IBM, the estimated volume trend is positive, which is not particularly surprising given the growth in equity market participation over time. However, there are also indications of positive volatility trends for 14 of the 20 firms. This could be a symptom of increasing idiosyncratic volatility in recent years (see, e.g., Campbell et al., 2001).

---

[5] In particular, the asymptotic standard error is $(\pi/\sqrt{24})T^{-1/4}$ and the maximum value of T is 8,440 daily observations.

The estimates of Ω imply that the log volatility and log volume innovations typically have a correlation between 40 and 60 percent. It follows, therefore, that volume shocks are quite useful for predicting contemporaneous volatility shocks, which is consistent with the evidence reported by Fleming, Kirby, and Ostdiek (2006). They use linear state-space methods to fit a factor model to daily trading volume and daily squared returns for the MMI firms, and find that the nonpersistent component of volume is highly correlated with the contemporaneous nonpersistent component of volatility. Specifically, this component of volume typically captures 60 to 80 percent of the variation in nonpersistent volatility.

The diagnostics for the model suggest that it is reasonably well specified. The innovations to $\log RV_t$ tend to have small estimated coefficients of skewness and excess kurtosis, with most of the skewness values less than 0.5, and most of the excess kurtosis values less than 2. There is some evidence of statistically significant serial correlation in the innovations. The Leung-Box statistic based on the first five autocorrelations produces four rejections at the 5 percent level, and the statistic based on the first ten autocorrelations produces seven rejections at the 5 percent level. Nonetheless, treating the log volatility innovations as Gaussian white noise is probably a reasonable approximation.

Normality is somewhat less plausible for the log volume innovations. The coefficient of skewness ranges from 0.13 (American Express) to 1.10 (Eastman Kodak), while the coefficient of excess kurtosis ranges from 0.64 (International Paper) to 6.44 (Procter and Gamble). Nonetheless, the Leung-Box statistics produce about the same number of rejections as for realized volatility, which suggests we could reasonably treat the volume innovations as non-Gaussian white noise. Using the first five autocorrelations produces two rejections at the 5 percent level, and using the first ten autocorrelations produces five rejections. We could reduce the number of rejections by fitting a higher-order VAR specification, but standard criteria for selecting the dimension of a model, such as the Schwarz (1978) criterion, favor the more parsimonious VAR(1) parameterization.

3.2   Results using the HFBC estimator

Table 2 reports the model fitting results obtained using the HFBC estimator of realized volatility. These results shed additional insights into the results reported in Table 1. In general, our previous results bolster the case for using long memory specifications to model the dynamics of volatility and volatility. However, the one finding at odds with prior research is the lack of empirical support for the hypothesis that the two series share

a common long-memory parameter. We now assess whether these results, and this last finding in particular, are robust to our choice of realized volatility measure.

The estimates of δ_1 reported in Table 2, like our earlier results, point to long memory in the volatility process. However, a key difference is that the estimates are now much lower; the mean of the estimates is 0.36, and the range is 0.33 to 0.40. Since these estimates are much closer to the estimates of δ_2 and the standard errors are essentially unchanged, the number of rejections of the hypothesis $\delta_1 = \delta_2$ falls from 18 in Table 1 to only four in Table 2. It seems that using a more precise realized volatility estimator has a substantial impact on our inferences regarding the empirical plausibility of a common long-memory parameter. In all other respects the results in Panel A of Table 2 are similar to those reported in Panel A of Table 1.

The diagnostics reported in Panel B of Table 2 suggest a modest improvement in the specification of the model. In particular, the volatility innovations tend to have smaller estimated coefficients of skewness and excess kurtosis than in Table 1. Most of skewness values are less than 0.4, and most of the excess kurtosis values are less than 1. Although there is still evidence of statistically significant serial correlation in the innovations, the number of rejections based on the first five autocorrelations falls from four to three, and the number of rejections based on the first ten autocorrelations falls from seven to five. These changes, while relatively minor, suggest that there are gains to using an improved realized volatility measure to fit the long memory model.

3.3 Results using daily squared returns

Although we find more support for the proposition that volume and volatility share a common order of fractional integration using our preferred realized volatility estimator, we still find less support than Bollerslev and Jubinski (1999). We reject the hypothesis for 20 percent of the MMI firms, while they reject for only 8 percent of the S&P 100 firms. To investigate the extent to which this finding may be due to our use of a more precise proxy for the volatility of daily returns, we fit the model a third time using daily squared returns in place of realized volatilities.

Table 3 reports the model fitting results using daily squared returns. As in both of the previous tables, the estimates of δ_1 and δ_2 support the hypothesis that volume and volatility display long memory. However, the δ_1 estimates now range from 0.07 (3M) to 0.17 (American Express), which is substantially lower than the estimates reported in

Tables 1 and 2, while the magnitude of the δ_2 estimates are relatively unchanged. As a result, the results in Table 3 provide strong evidence against the hypothesis of a common order of fractional integration. Using the likelihood ratio test, we reject the hypothesis for all of the firms at the 5 percent level.

The diagnostics in Panel B of Table 3 suggest a possible explanation for the disparity between the different sets of δ estimates. The estimated variance of the innovation to $\log r_t^2$ is typically an order of magnitude larger than the estimated variance of the innovations to $\log RV_t$ and $\log V_t$. Since this implies a much more difficult signal-extraction problem, we should expect to obtain less precise information about the dynamics of the conditional variance from the squared returns than from the realized variances. Moreover, the innovation to $\log r_t^2$ displays more serious departures from normality than the innovations to $\log RV_t$ and $\log V_t$. It seems likely that the performance of the QML estimator would deteriorate under such circumstances.

Another concern is that the standard errors in Table 3 may be unreliable if the limiting distribution of the QML estimator is a poor approximation to its finite-sample distribution. Nielsen (2004b) provides some preliminary evidence on the finite sample performance of the likelihood ratio test for models with Gaussian innovations. He finds that the test has good size in samples as small as 100 observations, and that the rejection frequencies under the alternative are close to the asymptotic local power implied by the asymptotic distribution theory. Since the innovations for the model in Table 2 display only mild departures from normality, it seems likely that the standard errors reported in Table 2 are reliable. This may not be true, however, for the standard errors in Table 3. Studies that fit short-memory stochastic volatility models to log squared returns via Gaussian QML report that this produces highly inefficient inferences (see, e.g., Alizadeh et al., 2002). This probably holds for long-memory models as well.

As an additional robustness test, we investigate whether fitting the model to squared returns rather than log squared returns alters the basic message of the analysis. This test is motivated by a well-known drawback of the logarithmic transformation: it translates any squared return sufficiently close to zero into a negative outlier. In general, the parameter estimates (not reported) are similar to those obtained using log squared returns, and again the estimates of δ_1 are substantially smaller than those in Table 2. As an aside, we should point out that the assumption of i.i.d. innovations is almost certain to be violated for the

squared-return specification.[6] Although this would be problematic from the standpoint of conducting formal hypothesis tests, it is not a significant concern for the limited purposes of the application considered here.

4 Model forecasting performance

We now investigate the volatility forecasting performance of our long memory models. We are specifically interested in two issues. First, we want to measure the incremental forecasting performance associated with using realized variances instead of daily squared returns to fit the model. Second, we want to measure the incremental performance of volatility estimates obtained by conditioning on contemporaneous trading volume.

4.1 Methodology

Our methodology is based on the Andersen et al. (2003) regression approach for comparing the performance of volatility forecasts produced by different models. Let $h_{t|t-1}$ denote the forecast of the log variance for day t given the (model specific) information observed through day $t-1$. We use an OLS regression of the form

$$\log RV_t = a + bh_{t|t-1} + u_t, \tag{18}$$

to assess how much of the variation in the log volatility is captured by a given model. Although the regression R-squared is biased towards zero because the variance of $\log RV_t$ is greater than the variance of the true log volatility (see Andersen et al. (2005) for details), this is irrelevant for evaluating relative forecasting performance because the ratio of the R-squared values produced by different models is bias free.

We use this approach to compare the volatility forecasts from our long memory model fit to the HFBC estimates of realized volatility to those from the model fit to daily squared returns. For the realized volatility specification, the value of $h_{t|t-1}$ is simply the fitted conditional mean of $\log RV_t$ produced by the model. For the squared return specification, we adjust the fitted values from the model to account for the difference between the unconditional mean of $\log r_t^2$ and the unconditional mean of the log variance.

---

[6] To see this, write $r_t^2 = h_t \nu_t$, where $h_t \equiv \mathrm{E}[r_t^2 | \mathcal{I}_{t-1}]$, $\nu_t \equiv r_t^2/h_t$, and $\mathcal{I}_{t-1}$ denotes the date $t-1$ information set. Since $\log r_t^2 = \log h_t + \log \nu_t$, it is reasonable to assume that a log linear model has i.i.d. innovations. Note, however, that $r_t^2 = h_t + z_t$, where $z_t = h_t(\nu_t - 1)$. Even if ν_t is i.i.d., z_t is not unless h_t is constant.

The standard way to do this is to assume that we can express the squared demeaned return as $r_t^2 = h_t z_t^2$, where h_t is the unobserved variance and z_t is Gaussian white noise. Under this approach, it follows that $\text{E}[\log h_t] = \text{E}[\log r_t^2] + 1.27$. Thus, we add 1.27 to each of the fitted log variances to obtain the log variance forecasts.

We use a similar strategy to investigate the relation between the unexpected components of trading volume and volatility. Suppose that

$$\log RV_t = h_{t|t-1} + \eta_{1t} \tag{19}$$
$$\log V_t = m_{t|t-1} + \eta_{2t} \tag{20}$$

where $h_{t|t-1}$ and $m_{t|t-1}$ are the conditional means implied by our long memory model and $\eta_t = (\eta_{1t}, \eta_{2t})'$ is distributed i.i.d. $N(0, \Omega)$. Now let $\mathcal{I}_{t-1} = \{V_1, RV_1, \ldots, V_{t-1}, RV_{t-1}\}$ denote the day $t-1$ information set. With Gaussian innovations, we have

$$\log RV_t | \mathcal{I}_{t-1}, V_t \sim N(h_{t|t}, \Omega_{11|2}) \tag{21}$$

where $h_{t|t} = h_{t|t-1} + \Omega_{12}\Omega_{22}^{-1}(\log V_t - m_{t|t-1})$ and $\Omega_{11|2} = \Omega_{11} - \Omega_{12}^2\Omega_{22}^{-1}$.[7] Thus, we can use a regression of the form

$$\log RV_t = a + bh_{t|t} + u_t, \tag{22}$$

to investigate how conditioning on contemporaneous trading volume improves our volatility estimates. In particular, we can use the ratio of the R-squared values for the regressions in equations (18) and (22) as a measure of the information content of unexpected volume.

4.2 Empirical results

Table 4 summarizes the results of the volatility regressions. Panel A focuses on the realized volatility specification of the long memory model. The first three columns report the intercept, slope, and R-squared estimates for the regression in equation (18), the next three columns report the same quantities for the regression in equation (22), and the last column reports the ratio of the R-squared values for the two regressions. Panel B reports the same set of results for the squared returns specification of the long memory model.

The results in Panel A show that the one-step-ahead log variance forecasts produced by the realized volatility specification exhibit little evidence of bias. The intercept estimates are close to zero and the slope estimates are close to one. This is expected given

---

[7] More generally, for models with non-Gaussian innovations, $h_{t|t}$ represents the minimum-mean-square linear estimate of $\log RV_t$ given V_t after we first condition on $\mathcal{I}_{t-1}$.

the nature of the first-order conditions imposed during the QML estimation procedure. Moreover, the R-squared values for the forecasting regression range from 31 percent for General Motors to 58 percent for AT&T, which indicates that the long memory model produces forecasts that capture a sizeable fraction of the time-series variation in the log realized variances. This is consistent with the evidence presented in Andersen et al. (2003).

The results in Panel B indicate that the one-step-ahead log variance forecasts produced by the squared returns specification perform more poorly. The intercept and slope estimates suggest that the forecasts are biased, and the R-squared values now range from 19 percent for General Motors and Johnson & Johnson to 48 percent for AT&T. Comparing the R-squared values in Panels A and B, there is a reduction in explanatory power for every firm, and the reduction in R-squared is typically on the order of 0.1 or 0.2. These comparisons provide an indication of the value of using high-frequency return data to construct volatility forecasts.

Now consider the impact of updating the volatility forecasts from the realized volatility specification (Panel A) using the contemporaneous realization of trading volume. The forecasts still show little evidence of bias, but there is a marked increase in the explanatory power for most firms. The regression R-squared values range from 50 percent for Johnson & Johnson to 71 percent for General Electric and AT&T. The ratio of the R-squared for the regression in equation (18) to the R-squared for the regression in equation (22) is, on average, 0.73. Since this ratio is unaffected by measurement error, it should provide an accurate reflection of the extent to which observing contemporaneous volume resolves the uncertainty about volatility. Another way to look at this is to compute the reduction in the MSE obtained by updating the log variance forecasts to account for the observed trading volume. On average, the reduction is 30 percent.

We observe a similar increase in explanatory power by updating the forecasts from the squared returns specification (Panel B) with contemporaneous volume. In this case, the average increase in the R-squared value is 0.12, and the ratio of R-squared values for the regressions in equations (18) and (22) is 0.72. Interestingly, for most firms, the R-squared for the updated forecasts is comparable to the R-squared for the one-step-ahead forecasts from the realized volatility specification.

Thus, the realization of contemporaneous trading volume is quite informative about unpredictable shocks to return volatility, even at the daily sampling frequency. This sug-

gests that daily trading volume can be used in conjuction with daily returns to obtain more precise estimates of daily return volatility for situations in which the intradaily returns data needed to construct realized volatilities are unavailable. Of course, the updated forecasts are no longer forecasts in the usual sense. Like the log realized variances, they are estimates of the unobserved log variance based on both past and contemporaneous information. However, the motivation is the same as using the realized volatility approach to obtain a more precise proxy for daily volatility on a given day.

Volume data might also prove useful in the econometric analysis of realized volatility. For example, realized volatility is often modeled as the true volatility plus measurement error. In this case, we can obtain more precise volatility estimates by employing a filtering technique. Barndorff-Nielsen and Shephard (2002) propose a simple approach based on the Kalman filter. A similar strategy could be employed in a long-memory setting. For example, we might model the log realized variances as a fractionally-integrated process subject to additive measurement error, and then estimate a joint system that includes log trading volume via linear state-space methods.[8] This would allows us to exploit the contemporaneous correlation between volatility and volume shocks to extract more efficient "smoothed" estimates of the latent volatility process.

5   Concluding Remarks

Recent empirical work suggests that return volatility and trading volume exhibit similar long-run dynamics. Building on this research, we examine the performance of a long-memory model for return volatility and trading volume that permits both common long-run dependencies and flexible short-run interactions. In particular, we assume that the logarithms of volatility and volume are described by a trend-stationary fractionally-integrated process that displays VAR(p) dynamics after fractional differencing. Since our approach allows the trend, short-memory, and long-memory components to compete on an equal footing, it should be well suited to analyzing the relative importance of each component in explaining the observed characteristics of the data.

When we fit the model to daily realized volatilities and daily trading volumes for the 20 MMI firms, the results paint an interesting picture of the role of long memory in

---

[8]   Chan and Palma (1998) show that these methods perform well for univariate long-memory specifications.

volatility and volume dynamics. Both the log realized volatilities and log trading volumes appear to be fractionally integrated with an order of integration of about 0.35. However, the evidence on whether the two series share a common order of fractional integration is mixed, with one fifth of the firms producing rejections of this hypothesis at the 5 percent significance level. This is a considerably higher rejection rate than that reported previously for the S&P 100 firms. Although drawing definitive conclusions about this issue would require a better understanding of the finite sample performance of the tests used in this paper and in previous work, our findings suggest that it is premature to treat common long-run dependencies as a stylized fact of the volume-volatility literature.

With respect to short-run dynamics, we find that they play only a minor role in explaining the serial dependence in the data. The main short-run phenomenon is a strong contemporaneous correlation between the log volatility and log volume innovations. In light of this correlation, it may be possible to use trading volume data to improve estimates of daily return volatility for cases in which the high-frequency return data needed to construct realized volatilities are unavailable. On the other hand, it seems that there is little to be gained from modeling short-run dynamics if the objective is to improve short-range volatility forecasts. Fractional differencing by itself is sufficient to remove most of the serial dependence in the realized volatilities.

There are a couple of issues that hold potential for future research in this area. First, it would be useful to document the finite-sample size and power of asymptotic tests in the long-memory setting. This would facilitate comparison between studies that use different estimation techniques, and would help shed light on why the QML estimates of the memory parameter for volatility are sensitive to whether we use realized volatilities or squared returns to fit the models. Second, it would be interesting to study the implications of long memory for long-range volatility forecasts. If these forecasts can be improved by adopting long-memory specifications, then we can potentially improve the pricing and hedging of long-term financial contracts.

References

Alizadeh, S., M. Brandt, and F.X. Diebold, 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 57, 1041–1558.

Andersen, T.G., 1996, Return volatility and trading volume: An information flow interpretation of stochastic volatility, *Journal of Finance* 51, 169–204.

Andersen, T.G., and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance* 4, 115–158.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001, The distribution of stock return volatility, *Journal of Financial Economics* 61, 43–76.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42–55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579–625.

Andersen, T., T. Bollerslev, and N. Meddahi, 2005, Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities, *Econometrica* 73, 279–296.

Baillie, R., T. Bollerslev, and H. Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 74, 3–30.

Bandi, F.M., and J.R. Russell, 2005, Microstructure noise, realized volatility, and optimal sampling, Working paper, University of Chicago.

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

Bollerslev, T., and D. Jubinski, 1999, Equity trading volume and volatility: Latent information arrivals and common long-run dependencies, *Journal of Business and Economic Statistics* 17, 9–21.

Bollerslev, T., and H. Mikkelsen, 1996, Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73, 151-184.

Campbell, J., M. Lettau, B. Malkiel, and Y. Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of Finance* 56, 1–43.

Chan, N., and W. Palma, 1998, State space modeling of long-memory processes, *Annals of Statistics* 26, 719–740.

Deo, R., and C. Hurvich, 2001, On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models, *Econometric Theory* 17, 686–710.

Ding, Z., C. Granger, and R. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.

Dacorogna, M., U. Muller, R. Nagler, R. Olsen, and O. Pictet, 1993, A geographical model for the daily and weekly seasonal volatility in the FX market, *Journal of International Money and Finance* 12, 413–438.

Diebold, F.X., and A. Inoue, 2001, Long memory and regime switching, *Journal of Econometrics* 105, 131–159.

Fleming, J., C. Kirby, and B. Ostdiek, 2006, Stochastic volatility, trading volume, and the daily flow of information, *Journal of Business* 79, 1551–1590.

Gallant, R., P. Rossi, and G. Tauchen, 1992, Stock prices and volume, *Review of Financial Studies* 5, 199–242.

Geweke, J., and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221–238.

Giraitis, L., and H. Koul, 1997, Estimation of the dependence parameter in linear regression with long-range dependent errors, *Stochastic Processes and their Applications* 71, 207–224.

Granger, C.W.J., and N. Hyung, 2004, Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance* 11, 399–421.

Hansen, P.R., and A. Lunde, 2004, An unbiased measure of realized variance, Working paper, Stanford University.

Hansen, P.R., and A. Lunde, 2005, A realized variance for the whole day based on intermittent high-frequency data, *Journal of Financial Econometrics* 3, 525–554.

Hyung, N., S. Poon, and C. Granger, 2006, A source of long memory in volatility, Working paper.

Luu, J.C., and M. Martens, 2003, Testing the mixture-of-distributions hypothesis using "realized" volatility, *Journal of Futures Markets* 23, 661–679.

Merton, R.C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Newey, W.K., and K.D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Nielsen, M., 2004a, Efficient likelihood inference in nonstationary univariate models, *Econometric Theory* 20, 116–146.

Nielsen, M., 2004b, Efficient inference in multivariate fractionally integrated time series models, *Econometrics Journal* 7, 63–97.

Tauchen, G.E., and M. Pitts, 1983, The price variability-volume relationship on speculative markets, *Econometrica* 51, 485–505.

Taylor, S., 1986, Modeling Financial Time Series (Wiley, Chichester).

Table 1
Model fitting results using the standard estimator of realized volatility

The table reports the results of fitting a bivariate linear regression model with deterministic regressors and long-memory errors to the daily realized variance and daily trading volume for the MMI firms. The model is

$$y_t = \beta x_t + \varepsilon_t,$$
$$\Phi(L)\Delta(L)\varepsilon_t = \eta_t,$$

where $y_t = (\log RV_t, \log V_t)'$ with RV_t based on a five-minute sampling frequency and a window length of zero, $x_t = (1, t/T, t^2/T^2)'$, $\Phi(L) = I - \Phi L$, $\Delta(L) = \mathrm{diag}((1-L)^{\delta_1}, (1-L)^{\delta_2})$, and $\eta_t \sim \mathrm{NID}(0, \Omega)$ with $\eta_t = 0\ \forall t \leq 0$. We estimate δ and $\phi = \mathrm{vec}(\Phi)$ by Gaussian quasi-maximum likelihood, which is equivalent to minimizing $|\hat{\Omega}(\delta, \phi)|$ where

$$\hat{\Omega}(\delta, \phi) = \sum_{t=1}^{T}[\Phi(L)\Delta(L)(y_t - \hat{\beta}x_t)][\Phi(L)\Delta(L)(y_t - \hat{\beta}x_t)]' \quad \text{with} \quad \hat{\beta} = \left(\sum_{t=1}^{T} y_t x_t'\right)\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}.$$

Panel A reports the elements of $\hat{\delta}$ and their asymptotic standard errors, the elements of $\hat{\phi}$ and their asymptotic t-ratios, the quasi-log likelihood value ($\mathcal{L}_T$), and the quasi-likelihood-ratio statistic (LR) for H_0: $\delta_1 = \delta_2$. Panel B reports the elements of $\hat{\beta}$ and summary statistics for the fitted innovation sequence $\{\eta_t\}_{t=1}^{T}$ implied by the parameter estimates. The summary statistics are the unique elements of $\hat{\Omega}$, the coefficients of skewness (CS) and excess kurtosis (CK) for the fitted innovations, and the Leung-Box statistics computed from the first five and ten sample autocorrelations of the fitted innovations (Q_5 and Q_{10}). An asterisk indicates a likelihood ratio or Leung-Box statistic that is statistically significant at the 5 percent level. The realized variances are expressed as squared annualized percentage rates (assuming 252 trading days per year). Trading volume is measured in millions of shares. The sample period is January 1, 1993 to December 31, 2003.

Panel A: Long-memory and VAR parameters

Firm	$\hat{\delta}_1$	$\hat{\delta}_2$	$se(\hat{\delta}_1)$	$se(\hat{\delta}_2)$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$t(\hat{\phi}_1)$	$t(\hat{\phi}_2)$	$t(\hat{\phi}_3)$	$t(\hat{\phi}_4)$	$\mathcal{L}_T$	LR
AXP	0.45	0.39	0.02	0.02	-0.06	0.10	0.03	0.01	-1.78	2.48	1.64	0.27	-587.5	5.19*
CVX	0.42	0.29	0.02	0.02	-0.10	0.05	0.00	0.07	-3.33	1.46	0.02	1.94	-182.0	23.15*
DD	0.43	0.33	0.02	0.02	-0.07	0.05	0.00	0.12	-2.15	1.19	-0.20	3.19	-289.7	11.05*
DIS	0.45	0.36	0.02	0.02	-0.15	-0.01	0.05	0.11	-5.04	-0.29	3.40	3.06	-369.2	11.44*
DOW	0.45	0.32	0.02	0.02	-0.13	0.10	0.02	0.02	-4.61	2.75	1.00	0.62	-1059.4	24.81*
EK	0.39	0.34	0.02	0.02	-0.11	0.01	0.08	0.18	-3.44	0.22	4.57	4.54	-1070.0	6.19*
GE	0.43	0.37	0.02	0.02	-0.08	0.06	0.06	0.03	-2.43	1.79	2.60	0.78	173.6	5.89*
GM	0.41	0.36	0.02	0.02	-0.10	-0.02	0.03	0.11	-3.16	-0.50	2.02	2.94	-631.2	4.95*
IBM	0.46	0.39	0.02	0.02	-0.10	0.04	0.07	0.11	-2.76	0.88	3.04	2.77	-249.1	11.61*
IP	0.42	0.32	0.02	0.02	-0.12	-0.02	0.02	0.14	-4.19	-0.41	1.69	3.99	-619.1	11.51*
JNJ	0.43	0.33	0.02	0.02	-0.12	0.01	0.05	0.11	-3.74	0.19	2.79	3.12	-218.9	15.18*
KO	0.43	0.25	0.02	0.03	-0.09	0.06	0.04	0.22	-2.75	1.45	2.71	5.73	74.7	39.39*
MCD	0.42	0.34	0.02	0.02	-0.10	0.04	0.01	0.09	-3.10	1.07	0.55	2.52	-479.4	22.62*
MMM	0.41	0.34	0.02	0.02	-0.08	0.03	0.04	0.10	-2.61	0.65	2.07	2.68	-560.8	24.47*
MO	0.41	0.33	0.02	0.02	-0.13	0.04	0.10	0.19	-3.92	1.08	4.95	5.01	-744.1	11.67*
MRK	0.42	0.30	0.02	0.02	-0.07	-0.01	0.03	0.19	-2.23	-0.24	1.68	4.96	-168.0	18.73*
PG	0.43	0.33	0.02	0.02	-0.05	0.05	0.03	0.09	-1.45	1.20	1.56	2.57	-282.3	12.89*
S	0.41	0.38	0.02	0.02	-0.08	0.04	0.03	0.08	-2.46	1.04	2.19	2.23	-927.7	1.84
T	0.42	0.39	0.02	0.02	-0.07	0.03	0.03	0.11	-2.17	0.64	1.90	3.16	-531.3	0.91
XOM	0.46	0.27	0.02	0.02	-0.11	0.07	0.02	0.10	-3.75	2.13	0.95	2.94	452.5	49.71*

Panel B: Trend parameters and innovation diagnostics

Firm	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{23}$	$\hat{\Omega}_{11}$	$\hat{\Omega}_{12}$	$\hat{\Omega}_{22}$	Innovations to $\log RV_t$				Innovations to $\log V_t$			
										CS	CK	Q_5	Q_{10}	CS	CK	Q_5	Q_{10}
AXP	3.50	-0.16	0.00	1.27	0.41	-0.15	0.05	0.04	0.13	0.34	1.72	5.44	12.24	0.13	1.65	3.65	10.31
CVX	2.77	1.77	-1.75	-0.09	-0.19	1.48	0.04	0.03	0.10	0.37	1.87	6.91	21.94*	0.40	3.14	4.03	11.14
DD	2.95	1.43	-1.20	0.40	1.38	-0.73	0.05	0.03	0.11	0.28	1.18	9.10	16.68	0.42	2.37	1.64	20.05*
DIS	3.32	-0.14	0.46	1.15	0.71	0.20	0.05	0.03	0.12	0.62	2.29	7.75	13.38	0.79	2.98	7.83	13.96
DOW	2.96	0.25	0.31	0.45	0.76	-0.13	0.06	0.04	0.14	0.38	2.17	13.68*	21.43*	0.47	3.11	4.14	6.64
EK	3.26	-0.02	0.33	0.07	-1.16	2.26	0.06	0.06	0.18	0.58	2.29	20.15**	29.22**	1.10	4.51	11.59*	15.52
GE	2.76	1.70	-1.21	2.41	0.14	0.57	0.05	0.03	0.08	0.35	1.48	3.80	19.40*	0.34	1.84	0.67	14.56
GM	3.43	-0.27	0.23	0.83	-0.96	1.79	0.05	0.04	0.13	0.37	1.40	9.30	29.78**	0.30	1.58	5.21	13.16
IBM	3.29	0.79	-0.78	2.18	0.36	-0.74	0.06	0.05	0.11	0.33	1.05	7.90	15.07	0.57	2.27	7.60	22.93*
IP	2.70	2.91	-2.52	-0.41	1.80	-0.31	0.05	0.03	0.13	0.20	1.07	3.50	17.65	0.23	0.64	0.46	12.64
JNJ	3.34	-0.60	0.37	1.53	-0.76	1.44	0.05	0.03	0.10	0.23	1.31	10.61	18.84*	0.40	1.90	6.59	19.64*
KO	3.16	0.39	-0.51	0.89	0.59	0.11	0.04	0.03	0.10	0.30	1.53	6.39	11.15	0.38	1.68	3.57	7.58
MCD	3.12	0.50	-0.38	0.90	0.72	0.11	0.05	0.03	0.13	0.51	2.50	5.04	9.11	0.55	1.78	4.59	12.69
MMM	2.55	2.58	-2.29	-0.08	1.25	0.37	0.05	0.04	0.13	0.21	1.07	7.46	15.18	0.71	3.22	1.53	6.74
MO	3.13	1.20	-1.14	1.84	-0.13	0.26	0.07	0.05	0.13	1.06	4.44	3.21	4.62	0.95	3.66	2.41	6.64
MRK	3.47	-0.81	0.60	1.69	-0.64	0.82	0.05	0.04	0.10	0.34	2.85	11.89*	15.21	0.57	2.55	12.94*	20.83*
PG	3.04	1.48	-1.70	0.26	1.30	-0.31	0.05	0.03	0.11	0.27	1.23	4.87	16.06	0.76	6.44	3.94	21.57*
S	3.23	1.03	-0.67	-0.14	0.25	1.13	0.05	0.04	0.15	0.25	1.41	16.31**	27.32**	0.62	2.54	5.00	12.50
T	3.01	1.36	-0.74	-0.94	3.07	-0.64	0.05	0.03	0.12	0.49	2.24	3.20	9.98	0.85	3.83	2.51	16.94
XOM	2.68	1.64	-1.39	1.22	1.63	-0.40	0.04	0.02	0.08	0.32	1.48	4.66	16.09	0.15	2.06	2.00	10.67

Table 2

Model fitting results using the higher-frequency, bias-corrected (HFBC) estimator of realized volatility

The table reports the results of fitting a bivariate linear regression model with deterministic regressors and long-memory errors to the daily realized variance and daily trading volume for the MMI firms. The model is

$$y_t = \beta x_t + \varepsilon_t,$$
$$\Phi(L)\Delta(L)\varepsilon_t = \eta_t,$$

where $y_t = (\log RV_t, \log V_t)'$ with RV_t based on a 30-second sampling frequency and 30-minute window length, $x_t = (1, t/T, t^2/T^2)'$, $\Phi(L) = I - \Phi L$, $\Delta(L) = \text{diag}((1-L)^{\delta_1}, (1-L)^{\delta_2})$, and $\eta_t \sim \text{NID}(0, \Omega)$ with $\eta_t = 0\ \forall t \leq 0$. We estimate δ and $\phi = \text{vec}(\Phi)$ by Gaussian quasi-maximum likelihood, which is equivalent to minimizing $|\hat{\Omega}(\delta, \phi)|$ where

$$\hat{\Omega}(\delta, \phi) = \sum_{t=1}^{T}[\Phi(L)\Delta(L)(y_t - \hat{\beta}x_t)][\Phi(L)\Delta(L)(y_t - \hat{\beta}x_t)]' \quad \text{with} \quad \hat{\beta} = \left(\sum_{t=1}^{T} y_t x_t'\right)\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}.$$

Panel A reports the elements of $\hat{\delta}$ and their asymptotic standard errors, the elements of $\hat{\phi}$ and their asymptotic t-ratios, the quasi-log likelihood value ($\mathcal{L}_T$), and the quasi-likelihood-ratio statistic (LR) for H_0: $\delta_1 = \delta_2$. Panel B reports the elements of $\hat{\beta}$ and summary statistics for the fitted innovation sequence $\{\eta_t\}_{t=1}^{T}$ implied by the parameter estimates. The summary statistics are the unique elements of $\hat{\Omega}$, the coefficients of skewness (CS) and excess kurtosis (CK) for the fitted innovations, and the Leung-Box statistics computed from the first five and ten sample autocorrelations of the fitted innovations (Q_5 and Q_{10}). An asterisk indicates a likelihood ratio or Leung-Box statistic that is statistically significant at the 5 percent level. The realized variances are expressed as squared annualized percentage rates (assuming 252 trading days per year). Trading volume is measured in millions of shares. The sample period is January 1, 1993 to December 31, 2003.

Panel A: Long-memory and VAR parameters

Firm	$\hat{\delta}_1$	$\hat{\delta}_2$	$se(\hat{\delta}_1)$	$se(\hat{\delta}_2)$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$t(\hat{\phi}_1)$	$t(\hat{\phi}_2)$	$t(\hat{\phi}_3)$	$t(\hat{\phi}_4)$	$\mathcal{L}_T$	LR
AXP	0.39	0.39	0.02	0.02	-0.07	0.10	0.03	-0.01	-2.01	3.09	1.36	-0.40	-1206.0	0.01
CVX	0.33	0.29	0.02	0.02	-0.05	0.02	0.00	0.08	-1.55	0.75	0.10	2.14	-963.4	2.86
DD	0.37	0.34	0.02	0.02	-0.08	0.06	-0.02	0.09	-2.37	2.01	-0.91	2.51	-1099.6	1.18
DIS	0.38	0.37	0.02	0.02	-0.09	0.00	0.03	0.10	-2.65	-0.03	1.23	2.66	-1172.2	0.38
DOW	0.38	0.32	0.02	0.02	-0.09	0.05	0.00	0.02	-2.94	1.87	0.01	0.59	-1917.9	5.10*
EK	0.33	0.36	0.02	0.02	-0.09	0.02	0.09	0.15	-2.46	0.58	3.35	3.55	-1733.6	1.19
GE	0.38	0.36	0.02	0.02	-0.11	0.05	0.06	0.04	-3.36	1.75	2.08	0.98	-371.7	1.01
GM	0.33	0.35	0.02	0.02	-0.07	-0.02	0.03	0.12	-2.14	-0.66	1.11	3.01	-1364.7	0.49
IBM	0.40	0.38	0.02	0.02	-0.09	0.03	0.07	0.12	-2.50	0.85	2.13	2.86	-876.0	1.53
IP	0.37	0.33	0.02	0.02	-0.09	0.01	0.01	0.13	-2.88	0.31	0.45	3.52	-1291.7	2.03
JNJ	0.36	0.32	0.02	0.02	-0.10	0.01	0.05	0.12	-2.93	0.23	1.87	3.16	-997.4	2.56
KO	0.37	0.25	0.02	0.02	-0.09	0.05	0.05	0.21	-2.77	1.62	1.88	5.37	-628.1	22.94*
MCD	0.33	0.32	0.02	0.02	-0.07	0.01	0.00	0.11	-2.00	0.19	0.10	2.90	-1254.6	0.09
MMM	0.33	0.33	0.02	0.02	-0.08	0.01	0.02	0.11	-2.24	0.32	0.92	2.90	-1385.2	0.00
MO	0.35	0.31	0.02	0.02	-0.08	0.01	0.11	0.22	-2.15	0.41	3.24	5.23	-1422.9	2.93
MRK	0.38	0.30	0.02	0.02	-0.08	0.03	0.01	0.18	-2.43	0.82	0.26	4.41	-768.6	10.39*
PG	0.36	0.33	0.02	0.02	-0.09	0.02	0.02	0.10	-2.75	0.59	0.92	2.60	-1014.2	1.85
S	0.36	0.39	0.02	0.02	-0.07	0.05	0.01	0.06	-2.08	1.65	0.63	1.55	-1814.3	1.67
T	0.37	0.40	0.02	0.02	-0.05	0.05	0.01	0.09	-1.53	1.46	0.52	2.34	-1175.7	1.61
XOM	0.40	0.27	0.02	0.02	-0.12	0.07	0.01	0.10	-3.84	2.82	0.38	2.71	-309.3	28.82*

Panel B: Trend parameters and innovation diagnostics

Firm	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{23}$	$\hat{\Omega}_{11}$	$\hat{\Omega}_{12}$	$\hat{\Omega}_{22}$	Innovations to log RV_t				Innovations to log V_t			
										CS	CK	Q$_5$	Q$_{10}$	CS	CK	Q$_5$	Q$_{10}$
AXP	3.05	1.11	-0.76	1.27	0.41	-0.15	0.09	0.06	0.13	0.29	1.22	4.55	7.81	0.13	1.64	3.59	10.15
CVX	2.55	2.32	-2.08	-0.09	-0.19	1.48	0.09	0.04	0.10	0.17	0.80	3.46	9.41	0.40	3.18	3.45	10.68
DD	2.63	2.33	-1.84	0.40	1.38	-0.73	0.09	0.05	0.11	0.33	1.11	8.20	15.53	0.41	2.35	1.84	19.61*
DIS	3.00	0.50	0.10	1.15	0.71	0.20	0.09	0.06	0.12	0.46	1.15	9.27	16.21	0.79	2.98	7.91	14.16
DOW	2.71	0.91	-0.15	0.45	0.76	-0.13	0.12	0.06	0.14	0.28	0.98	9.13	13.99	0.47	3.08	4.09	6.87
EK	2.77	1.10	-0.45	0.07	-1.16	2.26	0.11	0.09	0.18	0.61	1.77	19.16**	33.08**	1.10	4.50	13.05*	17.10
GE	2.51	2.17	-1.40	2.41	0.14	0.57	0.08	0.05	0.08	0.23	0.55	0.77	12.12	0.34	1.83	1.53	15.27
GM	3.13	0.46	-0.15	0.83	-0.96	1.79	0.10	0.06	0.13	0.20	0.45	10.14	20.18*	0.30	1.57	4.99	12.84
IBM	3.01	1.43	-1.23	2.18	0.36	-0.74	0.10	0.07	0.11	0.30	0.50	8.43	10.47	0.56	2.27	6.31	21.14*
IP	2.61	2.82	-2.30	-0.41	1.80	-0.31	0.08	0.05	0.13	0.18	0.72	1.92	14.79	0.23	0.64	0.33	12.84
JNJ	3.05	0.22	-0.20	1.53	-0.76	1.44	0.10	0.05	0.10	0.28	0.81	8.32	22.23*	0.41	1.91	5.98	19.60*
KO	2.73	1.60	-1.29	0.89	0.59	0.11	0.08	0.05	0.10	0.26	0.76	2.91	7.36	0.38	1.68	4.24	8.18
MCD	2.80	1.07	-0.54	0.90	0.72	0.11	0.10	0.06	0.13	0.40	1.04	3.20	6.31	0.55	1.77	4.12	11.81
MMM	2.35	2.99	-2.55	-0.08	1.25	0.37	0.10	0.06	0.13	0.24	1.02	4.46	9.88	0.70	3.21	1.01	6.56
MO	2.81	1.79	-1.49	1.84	-0.13	0.26	0.12	0.08	0.13	0.79	2.58	2.86	7.23	0.96	3.69	2.63	6.94
MRK	3.05	0.46	-0.25	1.69	-0.64	0.82	0.09	0.06	0.10	0.42	1.51	14.47*	16.03	0.58	2.53	12.09*	20.21*
PG	2.70	2.18	-2.19	0.26	1.30	-0.31	0.09	0.05	0.11	0.64	3.62	7.59	15.38	0.76	6.42	3.96	21.20*
S	2.94	1.40	-0.81	-0.14	0.25	1.13	0.11	0.07	0.15	0.27	0.84	17.65**	22.76*	0.62	2.51	5.38	12.43
T	2.63	1.97	-0.96	-0.94	3.07	-0.64	0.09	0.06	0.12	0.60	1.98	4.69	13.70	0.86	3.86	2.34	17.04
XOM	2.41	2.30	-1.79	1.22	1.63	-0.40	0.07	0.04	0.08	0.27	0.55	4.90	13.13	0.15	2.08	1.95	10.34

Table 3
Model fitting results using daily squared returns

The table reports the results of fitting a bivariate linear regression model with deterministic regressors and long-memory errors to the daily squared demeaned returns and daily trading volume for the MMI firms. The model is

$$y_t = \beta x_t + \varepsilon_t,$$

$$\Phi(L)\Delta(L)\varepsilon_t = \eta_t,$$

where $y_t = (\log r_t^2, \log V_t)'$, $x_t = (1, t/T, t^2/T^2)'$, $\Phi(L) = I - \Phi L$, $\Delta(L) = \text{diag}((1-L)^{\delta_1}, (1-L)^{\delta_2})$, and $\eta_t \sim \text{NID}(0, \Omega)$ with $\eta_t = 0 \ \forall t \leq 0$. We estimate δ and $\phi = \text{vec}(\Phi)$ by Gaussian quasi-maximum likelihood, which is equivalent to minimizing $|\hat{\Omega}(\delta, \phi)|$ where

$$\hat{\Omega}(\delta, \phi) = \sum_{t=1}^{T} [\Phi(L)\Delta(L)(y_t - \hat{\beta}x_t)][\Phi(L)\Delta(L)(y_t - \hat{\beta}x_t)]' \quad \text{with} \quad \hat{\beta} = \left(\sum_{t=1}^{T} y_t x_t'\right) \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}.$$

Panel A reports the elements of $\hat{\delta}$ and their asymptotic standard errors, the elements of $\hat{\phi}$ and their asymptotic t-ratios, the quasi-log likelihood value ($\mathcal{L}_T$), and the quasi-likelihood-ratio statistic (LR) for H_0: $\delta_1 = \delta_2$. Panel B reports the elements of $\hat{\beta}$ and summary statistics for the fitted innovation sequence $\{\eta_t\}_{t=1}^{T}$ implied by the parameter estimates. The summary statistics are the unique elements of $\hat{\Omega}$, the coefficients of skewness (CS) and excess kurtosis (CK) for the fitted innovations, and the Leung-Box statistics computed from the first five and ten sample autocorrelations of the fitted innovations (Q_5 and Q_{10}). An asterisk indicates a likelihood ratio or Leung-Box statistic that is statistically significant at the 5 percent level. The returns are expressed as annualized percentage rates and multiplied by $\sqrt{252}$. Trading volume is measured in millions of shares. The sample period is January 1, 1993 to December 31, 2003.

Panel A: Long-memory and VAR parameters

Firm	$\hat{\delta}_1$	$\hat{\delta}_2$	$se(\hat{\delta}_1)$	$se(\hat{\delta}_2)$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$t(\hat{\phi}_1)$	$t(\hat{\phi}_2)$	$t(\hat{\phi}_3)$	$t(\hat{\phi}_4)$	$\mathcal{L}_T$	LR
AXP	0.17	0.37	0.02	0.02	-0.13	0.03	0.11	0.04	-4.51	4.27	1.71	1.11	-5083.4	50.16*
CVX	0.10	0.28	0.02	0.02	-0.08	0.02	0.10	0.07	-2.84	3.81	1.41	2.29	-4864.9	38.94*
DD	0.12	0.29	0.02	0.03	-0.04	0.03	0.10	0.15	-1.32	4.51	1.42	4.41	-4946.4	25.98*
DIS	0.11	0.31	0.02	0.02	-0.07	0.02	0.22	0.14	-2.31	2.43	2.92	4.00	-5120.5	50.36*
DOW	0.15	0.31	0.02	0.02	-0.11	0.04	0.11	0.03	-3.72	5.24	1.81	0.93	-5339.5	35.50*
EK	0.10	0.30	0.02	0.03	-0.08	0.02	0.20	0.20	-2.47	1.93	3.02	5.44	-5640.8	47.74*
GE	0.12	0.32	0.02	0.02	-0.10	0.02	0.14	0.08	-3.33	4.21	1.75	2.48	-4399.3	52.63*
GM	0.14	0.31	0.02	0.02	-0.07	0.03	0.03	0.12	-2.46	4.03	0.42	3.65	-5117.7	32.50*
IBM	0.15	0.34	0.02	0.02	-0.15	0.03	0.25	0.14	-5.22	4.74	3.35	4.16	-4951.5	49.15*
IP	0.12	0.31	0.02	0.03	-0.06	0.01	0.13	0.14	-2.16	2.00	2.01	4.12	-5104.2	42.75*
JNJ	0.13	0.31	0.02	0.02	-0.07	0.02	0.01	0.11	-2.31	3.95	0.13	3.34	-4807.2	39.08*
KO	0.12	0.25	0.02	0.03	-0.10	0.03	0.27	0.21	-3.46	3.97	3.56	5.79	-4684.1	16.98*
MCD	0.11	0.30	0.02	0.02	-0.07	0.02	0.07	0.13	-2.27	2.37	0.96	3.67	-5162.0	33.52*
MMM	0.07	0.31	0.02	0.02	-0.05	0.02	0.13	0.12	-1.62	3.17	1.81	3.55	-5302.1	66.02*
MO	0.15	0.27	0.02	0.03	-0.11	0.01	0.36	0.26	-3.83	1.69	4.96	6.88	-4944.4	22.41*
MRK	0.15	0.27	0.02	0.03	-0.09	0.02	0.16	0.20	-2.92	3.62	2.12	5.56	-4790.8	17.05*
PG	0.11	0.32	0.02	0.02	-0.10	0.02	0.19	0.10	-3.59	3.03	2.53	2.99	-5001.3	52.67*
S	0.12	0.36	0.02	0.02	-0.07	0.01	0.04	0.10	-2.26	1.09	0.53	2.91	-5456.8	68.34*
T	0.16	0.35	0.02	0.03	-0.07	0.01	0.07	0.15	-2.44	2.37	0.80	4.40	-5517.9	33.18*
XOM	0.14	0.24	0.02	0.02	-0.11	0.02	0.12	0.13	-3.95	4.37	1.44	4.01	-4331.2	13.53*

Panel B: Trend parameters and innovation diagnostics

Firm	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{23}$	$\hat{\Omega}_{11}$	$\hat{\Omega}_{12}$	$\hat{\Omega}_{22}$	Innovations to $\log r_t^2$				Innovations to $\log V_t$			
										CS	CK	Q_5	Q_{10}	CS	CK	Q_5	Q_{10}
AXP	2.35	1.82	-1.34	1.27	0.41	-0.15	1.14	0.11	0.13	-1.15	2.24	5.33	6.38	0.13	1.69	3.51	10.45
CVX	2.04	2.05	-1.85	-0.09	-0.19	1.48	1.23	0.09	0.10	-1.33	3.88	3.71	10.56	0.39	3.13	2.88	10.40
DD	2.04	2.65	-2.19	0.40	1.38	-0.73	1.16	0.10	0.11	-1.00	1.13	5.75	8.00	0.40	2.34	1.71	20.72*
DIS	2.25	1.33	-0.58	1.15	0.71	0.20	1.29	0.13	0.12	-1.42	4.71	15.21**	20.24*	0.80	3.02	8.64	14.83
DOW	2.13	0.97	-0.15	0.45	0.76	-0.13	1.23	0.12	0.14	-1.22	2.81	8.48	16.21	0.47	3.16	4.12	6.83
EK	2.29	0.91	-0.51	0.07	-1.16	2.26	1.35	0.19	0.18	-1.10	1.83	0.64	6.41	1.08	4.43	9.91	14.33
GE	1.88	2.66	-1.85	2.41	0.14	0.57	1.12	0.09	0.08	-1.12	1.80	15.44**	26.14**	0.34	1.91	5.75	19.68*
GM	2.54	0.56	-0.22	0.83	-0.96	1.79	1.21	0.13	0.13	-1.07	1.44	7.24	14.49	0.29	1.66	3.39	10.32
IBM	2.40	1.50	-1.22	2.18	0.36	-0.74	1.24	0.13	0.11	-1.19	3.10	13.42*	14.94	0.59	2.35	3.71	17.98
IP	2.06	2.83	-2.30	-0.41	1.80	-0.31	1.15	0.11	0.13	-1.14	1.77	3.28	5.72	0.23	0.63	0.87	12.03
JNJ	2.42	0.64	-0.68	1.53	-0.76	1.44	1.25	0.11	0.10	-1.28	2.94	10.26	25.37**	0.39	1.89	4.61	20.40*
KO	2.06	2.51	-2.33	0.89	0.59	0.11	1.14	0.11	0.10	-1.04	1.22	12.90*	22.52*	0.37	1.76	3.85	7.52
MCD	2.17	1.40	-0.92	0.90	0.72	0.11	1.25	0.12	0.13	-1.09	1.41	5.83	10.37	0.54	1.75	3.86	11.64
MMM	1.81	2.64	-2.18	-0.08	1.25	0.37	1.37	0.12	0.12	-1.27	2.98	4.19	9.64	0.69	3.16	0.50	5.92
MO	2.17	2.38	-2.13	1.84	-0.13	0.26	1.13	0.16	0.13	-0.84	1.35	9.98	19.09*	0.95	3.69	4.97	10.23
MRK	2.42	1.02	-0.86	1.69	-0.64	0.82	1.22	0.11	0.10	-1.18	2.17	4.33	11.39	0.58	2.57	12.58*	21.36*
PG	2.07	2.61	-2.68	0.26	1.30	-0.31	1.27	0.12	0.11	-1.33	3.43	2.67	11.28	0.76	6.43	4.44	21.26*
S	2.52	0.99	-0.65	-0.14	0.25	1.13	1.30	0.15	0.15	-1.12	2.15	7.24	12.13	0.62	2.56	4.15	12.67
T	1.75	2.73	-1.46	-0.94	3.07	-0.64	1.66	0.14	0.12	-1.44	2.39	2.71	4.52	0.85	3.85	4.05	16.67
XOM	1.83	2.40	-1.84	1.22	1.63	-0.40	1.12	0.08	0.08	-1.08	1.61	6.17	12.22	0.15	2.04	1.07	10.74

Table 4
Realized volatility regressions

The table summarizes the results of fitting a linear regression of $\log RV_t$, where RV_t is constructed using our HFBC estimator of realized volatility, on the log variance estimates obtained from our fitted long memory models. Panel A reports the results for the long memory model fit based on realized volatilities and Panel B reports the results for the long memory model fit based on daily squared returns. For each model, we consider two realized volatility regressions. The explanatory variable in the first regression is a one-step-ahead forecast of the log variance. The explanatory variable in the second regression is an estimate of the log variance obtained by updating the one-step-ahead forecast to reflect the information revealed by the realization of contemporaneous trading volume. For each regression we report the intercept estimate ($\hat{\alpha}$), slope estimate ($\hat{\beta}$), and sample R-squared (R^2). The final column reports the ratio of the two values of the sample R-squared.

Panel A: Log variance estimates based on realized volatilities

| Firm | One step ahead | | | After observing V_t | | | |
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2	Ratio
AXP	-0.20	1.03	0.48	-0.11	1.02	0.64	0.75
CVX	-0.03	1.00	0.39	-0.02	1.00	0.52	0.75
DD	-0.10	1.02	0.49	-0.11	1.02	0.63	0.78
DIS	0.00	1.00	0.49	-0.05	1.01	0.64	0.76
DOW	-0.07	1.01	0.53	-0.05	1.01	0.63	0.84
EK	0.16	0.97	0.39	-0.02	1.00	0.64	0.61
GE	-0.07	1.01	0.56	-0.04	1.01	0.71	0.79
GM	-0.06	1.01	0.31	-0.07	1.01	0.52	0.59
IBM	-0.01	1.00	0.43	-0.09	1.01	0.68	0.64
IP	-0.10	1.02	0.53	-0.06	1.01	0.63	0.85
JNJ	-0.14	1.02	0.32	-0.06	1.01	0.50	0.64
KO	-0.19	1.03	0.48	-0.08	1.01	0.63	0.76
MCD	0.05	0.99	0.36	-0.03	1.00	0.55	0.66
MMM	-0.10	1.02	0.46	-0.03	1.01	0.62	0.75
MO	-0.13	1.02	0.42	-0.07	1.01	0.65	0.65
MRK	0.03	0.99	0.34	-0.09	1.01	0.58	0.59
PG	-0.24	1.04	0.46	-0.10	1.02	0.61	0.75
S	0.02	1.00	0.41	-0.04	1.01	0.57	0.72
T	0.04	0.99	0.58	0.00	1.00	0.71	0.82
XOM	-0.10	1.02	0.54	-0.06	1.01	0.65	0.83

Panel B: Log variance estimates based on squared demeaned daily returns

| Firm | One step ahead | | | After observing V_t | | | |
	$\hat{\alpha}$	$\hat{\beta}$	R^2	$\hat{\alpha}$	$\hat{\beta}$	R^2	Ratio
AXP	0.71	0.87	0.30	1.00	0.70	0.46	0.65
CVX	-0.58	1.08	0.27	1.29	0.64	0.33	0.81
DD	0.06	0.96	0.35	0.57	0.74	0.46	0.78
DIS	0.70	0.87	0.32	1.83	0.59	0.41	0.76
DOW	-0.23	1.01	0.40	0.23	0.78	0.46	0.87
EK	-1.45	1.21	0.28	1.97	0.57	0.42	0.65
GE	0.00	0.97	0.45	0.44	0.75	0.57	0.79
GM	1.07	0.82	0.19	1.95	0.58	0.37	0.50
IBM	0.11	0.96	0.29	1.59	0.63	0.47	0.60
IP	-0.28	1.02	0.43	1.09	0.68	0.45	0.95
JNJ	0.37	0.92	0.19	2.13	0.54	0.32	0.57
KO	0.88	0.84	0.29	1.89	0.57	0.41	0.71
MCD	0.27	0.94	0.25	1.54	0.62	0.39	0.64
MMM	-0.93	1.14	0.35	0.58	0.74	0.44	0.78
MO	0.61	0.88	0.31	1.40	0.64	0.52	0.60
MRK	0.74	0.86	0.22	1.68	0.60	0.42	0.53
PG	0.45	0.91	0.29	1.54	0.61	0.41	0.71
S	-2.51	1.35	0.31	1.64	0.63	0.36	0.87
T	1.82	0.73	0.48	1.83	0.61	0.58	0.83
XOM	0.08	0.96	0.38	0.77	0.70	0.45	0.85