The Stochastic Component of Realized Volatility

Wai Mun Fong [a*] and Wing-Keung Wong [b]

[a] Department of Finance and Accounting, National University of Singapore

[b] Department of Economics, National University of Singapore

Abstract

Volatility-volume regressions provide a convenient framework to study sources of volatility predictability. We apply this approach to the daily realized volatility of common stocks. We find that unexpected volume plays a more significant role in explaining realized volatility than expected volume, and accounts for about one-third of the nonpersistent component in the volatility process. Contrary to the findings of Lamoureux and Lastrapes (1990), the ARCH effect is robust even in the presence of volume. However, this component explains only about half of the variations in realized volatility. Thus, large portion of realized volatility is clearly stochastic. This presents a significant challenge to the goal of achieving precise realized volatility forecasts.

Statement of Relationship to Practice

Practitioners often rely heavily on autoregressive volatility models to estimate and predict the true volatility of financial securities. This paper shows that this reliance is misplaced. True volatility is much more random than is suggested by these popular models. Research should continue to ask what key structural factors drive volatility.

JEL classification: C22, G10

Keywords: Realized volatility, trading volume, autoregressive models, ARCH.

* Corresponding author: Tel: ++ 65 6874 6693. Fax: +65-6779-2083.
Email: bizfwm@nus.edu.sg

1. Introduction

Volatility-volume regressions provide a convenient framework to test microstructure theories and forecast volatility. The literature has generally focused on the first aspect, as exemplified by the works of Bessembinder and Seguin (1993) and Daigler and Wiley (1999). Bessembinder and Seguin (1993) examine the impact of trading volume and market depth on daily returns volatility of eight futures markets. They use two-stage regression to obtain absolute residuals of the daily returns, which they then regress against lagged absolute residuals, volume and open interest. Unexpected volume was found to have a significant effect on volatility but deeper markets (larger open interest) mitigate volatility.

Daigler and Wiley (1999) use volatility-volume regressions to test the reaction of futures volatility to trades by informed and uninformed investors. Consistent with the predictions of dispersion models (Foster and Viswanathan 1990, and Shalen 1993), trades by the general public (uninformed traders) increases volatility, but trades by informed traders mitigate volatility.

In this paper, we use volatility-volume regressions as a tool for understanding volatility predictability. Recent studies have emphasized the use of precise volatility proxies such as realized volatility for evaluating predictive accuracy of volatility models (Andersen and Bollerslev 1998, Andersen, Bollerslev, Diebold and Labys 2003). We focus on daily realized volatility computed by summing many intra-day squared returns. This estimator is much more efficient compared to standard proxies such as absolute returns, squared returns, or price range. Of course, realized volatility estimates contain measurement errors due to market microstructure frictions such as bid-ask bounce and price discreteness. Techniques to purge realized volatility from microstructure noise have been an active area of research in recent years (e.g., Ait-Sahalia and Mykland and Zhang 2005, Bandi and Russell 2005).

We construct realized volatility corrected for microstructure biases to proxy for the latent volatility of common stocks. Consistent with previous studies (Andersen et al. 2001), the log of daily realized volatility is nearly Gaussian, in contrast to the highly non-Gaussian distribution of squared and absolute daily returns. We exploit the Gaussian property of log realized volatility and model it as an autoregression with expected and unexpected trading volume as regressors. This approach produces simple and efficient estimates of volatility persistence (the ARCH effect) and allows us to readily quantify the impact of stochastic components (unexpected volume and volatility innovations) to overall volatility predictability.

Our approach to modeling volatility may be contrasted with the standard framework based on parametric models such as GARCH and stochastic volatility models. Volatility estimates produced by these models are strictly valid only if the assumptions of the model are correct and in general, there is no consensus as to which is the best model to proxy for latent volatility. Although some empirical studies e.g., Andersen and Bollerslev (1998) show that simple models like the GARCH(1,1) perform reasonably well in out-of-sample tests when evaluated against realized volatility, GARCH models appear to explain no more than half of the daily fluctuations in realized volatility. This indicates that true volatility is highly stochastic, and contains a large unpredictable component that may be missed by deterministic GARCH models. Meanwhile, stochastic volatility models are in general cumbersome to estimate because volatility is treated as latent in these models (see Ghysels, Harvey and Renault 1996). While there are feasible techniques such as Gaussian maximum likelihood methods (Harvey, Ruiz and Shepard 1994), these may yield biased and inefficient inferences about the latent volatility because the conditional distribution of log absolute or squared returns, upon which QML estimation is based, is highly non-Gaussian (Andersen and Sorensen 1997).

Direct volatility measures like absolute and squared returns are easy to compute, but are highly imprecise because they ignore information contained in intraday prices. This makes them unsuitable for forecasting true volatility or as benchmarks for evaluating volatility models. Against this backdrop, the appeal of realized volatility becomes clear. In particular, since the latent volatility is observable via realized volatility, we can model it directly as a stochastic process without the complications of latent factor stochastic volatility models. The fact that realized volatility is a precise volatility estimator mitigates severe errors-in-variables problem associated with more noisy estimators such as absolute or squared returns.

As mentioned, a key objective of this paper is to link realized volatility with information from trading volume. There is of course a large literature on the relationship between volatility and trading volume, but most of this literature is concerned with the empirical validity of microstructure theories that attempt to explain the volume-volatility relation. An early study which examines the information content of volume for volatility prediction is Lamoureux and Lastrapes (1990). Using a simple GARCH(1,1) model, they find that volatility persistence virtually disappears when volume is entered into the conditional volatility equation. Using a regression approach, Bessembinder and Seguin (1993) also report relatively low volatility persistence when absolute returns are regressed on lagged absolute returns and contemporaneous volume and open interest. Fleming et al. (2001) criticize these studies, and show that entering volume directly into a volatility equation is tantamount to assuming that past volume and volatility shocks generate the same degree of volatility persistence. This assumption is not supported by empirical studies that take into account of joint dependence between volume and volatility. In particular, using a bivariate model of volatility and volume with AR(1) dynamics, Fleming et al. (2006) find that the persistent (ARCH) component of volatility is only weakly related to volume. At the same time, a large portion of volatility is stochastic, and is closely tied to unexpected volume. Their results are consistent with a version of the Mixture of Distributions Hypothesis (MDH) in which asset volatility is driven by random information arrivals or news and unexpected volume is a good proxy for news. This contrasts with the standard approach (e.g., Andersen 1996) where information flow is assumed to be highly serially correlated to rationalize the ARCH effect. However, the assumption that information flow is as serially correlated as volatility is theoretically hard to justify. Attempts to add other factors to account for ARCH such as traders' responsiveness to news have also not been very successful empirically (Liesenfeld 2001). In contrast to the standard approach, our regressions assume that realized volatility has an ARCH component which is orthogonal to expected and unexpected volume. This setup is more in line with the findings of Fleming et al. (2006).

As a preview of the results, we find that realized volatility dynamics can be captured very well by an autoregression with fifteen lags. This finding is consistent with Bollerslev and Wright (2001) who show that simple autoregressions can produce realized volatility forecasts that are almost as accurate as those generated using more complicated frequency domain regressions.

Secondly, adding volume in the volatility regression does not lead to a disappearance of the ARCH effect, contrary to the findings of Lamoureux and Lastrapes (1990). This implies that past realized volatility is useful for forecasting future realized volatility even after accounting for volume dynamics.

Thirdly, the ARCH effect explains only about half of the variability of realized volatility. Thus, a large portion of the fluctuations in realized volatility appears to be unpredictable. Volatility innovations (regression residuals) contribute two-thirds to this unpredictable component, with unexpected volume accounting for the rest. These results suggest that even if we could reduce the influence of volatility innovations by half, about one third of daily variations in realized volatility will still be unforecastable. Consistent with theory, expected volume plays a very minor role in explaining realized volatility, and is of no help in achieving greater forecast accuracy.

With the increasing availability of high frequency price data, realized volatility may become a choice volatility benchmark in a variety of settings. This study should provide some perspectives on the extent to which this volatility estimator can be predicted.

The rest of the paper is organized as follows. Section 2 reviews the theory and measurement of realized volatility, including our approach in dealing with microstructure noise. Section 3 introduces our sample of firms, variables of interest in this study, and the data sources. This is followed by a preliminary analysis of the data in Section 4. The results of volatility-volume regressions are discussed in Section 5 and 6. Section 7 concludes the paper with a summary of the key findings.

2. Realized Volatility: Theory and Measurement

2.1 Asymptotic Theory

Very high frequency data can be used to construct ex-post volatility estimates of continuously compounded returns of assets obeying a standard diffusion. Specifically, assuming the return process satisfy no-arbitrage and has finite instantaneous mean, the log price follows a Brownian semi-martingale diffusion process as proved in Back (1991).

Let σ_t^2 denote the instantaneous volatility of returns. We are primarily interested in the variance of returns over discrete horizons e.g., a day. The true daily variance, conditional on the sample path $\{\sigma_{t-\tau}^2\}_0^1$ is

$$\overline{\sigma}_t^2 = \int_0^1 \sigma_{t-\tau}^2 d\tau \,. \tag{1}$$

In the literature, $\overline{\sigma}_t^2$ is known as the integrated or quadratic variance[1]. The integrated variance measures the realized sample path variation of the squared return process. A natural estimator of the integrated variance is obtained by summing intraday

[1] The integrated variance is widely used in the options pricing literature where the option price depends on the distribution of the integrated variance of the underlying asset of the life of the option (see e.g., Hull and White 1987).

squared returns over many small intervals within the day. This estimator is known as realized variance, denoted as[2]:

$$RV(m) = \sum_{k=1,2,\ldots m} r_{t+k/m}^2 \quad (2)$$

where m is the number of equally spaced discrete intervals in a day.

Assume for the moment that there are no microstructure frictions, this implies that the observed intraday log price is also the true or informationally efficient price. Under general regularity conditions, Andersen et al. (2001, 2003) and Barndorff-Nielsen and Shephard (2001, 2002), among others, show that $RV(m)$ converges in probability to the integrated variance as $m \to \infty$. Thus, by summing up high frequency squared returns, it is possible to construct realized volatility measures that are asymptotically free of measurement errors. Moreover, this result holds even if the underlying price process contains jumps as long as the price process is a Brownian semi-martingale.

Daily squared returns ($m = 1$) is clearly a special case of daily realized volatility. But, it is a very noisy or inefficient estimator because it ignores the information contained in intraday stock prices. The same is true for daily absolute returns. This explains the typically low R^2s obtained when daily squared returns are regressed against GARCH-generated forecasts even though GARCH models perform reasonably well when evaluated against realized volatility (see e.g., Andersen and Bollerslev 1998).

Noises in squared or absolute returns also affect inference for stochastic volatility models because these models primarily use squared or absolute returns to proxy for the latent volatility (Ghysels et al. 1996, Fleming et al. 2006). Moreover, since the conditional distribution of squared or absolute returns is known to be highly non-Gaussian (Andersen and Sorensen 1997), estimates of stochastic volatility parameters via

[2] Realized volatility is not a new concept as shown by the works of French et. al. (1987) and Schwert (1989) who compute monthly realized volatilities using daily returns. Theoretical justifications for the realized volatility approach first appeared in Merton (1980) in the context of estimating the diffusion coefficient for continuous time diffusions, and later extended to the theory of quadratic variance by Karatzas and Shreve (1991), Barndorff-Nielsen and Shepard (2001, 2002) and Andersen et al. (2001).

Gaussian quasi-maximum likelihood will be highly inefficient and biased in finite samples. By contrast, the distribution of log realized volatility is much closer to a Gaussian distribution (Andersen et al. 2001). In section 3, we show that this lognormal property holds for our sample.

2.2 Measurement

In practice, realized volatility contains measurement error due to microstructure noise in the recorded price. The main sources of microstructure noise are: nonsynchronous trading, bid-ask bounce, price discreteness and order clustering. Nonsynchronous trading leads to spurious autocorrelations in returns which may induce biases in estimates of returns moments and co-moments (Lo and MacKinlay 1990). This suggests that realized volatility should only be computed for actively traded stocks.

Bid-ask bounce occurs when the transacted price bounces back and forth between bid and ask prices, producing spurious volatility and autocorrelation in returns even if the true price is efficient. This suggests the use of mid-quotes rather than transaction prices for computing realized volatility.

Another source of microstructure noise is price discreteness. Although the asymptotic theory for realized volatility assumes that prices are observed continuously, actual prices are recorded at discrete intervals, as are changes in mid-quotes. The posting of quotes by market makers is also discrete in nature because market makers respond to a variety of factors such as information arrivals, changes in liquidity and order clustering, all of which occur in discrete time.

Microstructure noise can cause realized volatility to diverge to infinity as $m \to \infty$, causing it to be inconsistent. Since increasing the sampling interval will reduce the impact of microstructure noise, this might seem to be an obvious solution to the problem. However, increasing the sampling interval also increases the standard error of realized volatility. Thus, the best practical solution is to choose sampling intervals that optimally balance these two conflicting objectives. Several methods of choosing optimal sampling intervals have been proposed in the literature. The method used in this paper is based on the approach in Bandi and Russell (2005) who develop a general but easy to implement method for computing sampling intervals that are optimal in the mean square error (MSE) sense[3].

The basic idea is as follows. Suppose p_t is the observed the stock price on day *t*. and let $p_t = p_t^e \eta_t$ where p_t^e is the equilibrium price and η_t denotes microstructure noise in the log price. Thus, returns are contaminated by noise as follows:

$$r_t = r_t^e + \varepsilon_t, \quad \varepsilon_t = \ln(\eta_t) - \ln(\eta_{t-1}) \ . \tag{3}$$

Assuming that ε_t is i.i.d mean zero with bounded eighth moment, Bandi and Russell (2005) show that, for actively traded stocks, the following rule of thumb is valid for choosing optimal sampling intervals:

$$m^* \approx \left(\frac{Q_t}{(E(\varepsilon_t^{\,2}))^2} \right)^{1/3} \tag{4}$$

where m^* is the optimal number of intraday sub-periods and $Q_t = \int_0^t \tilde{\sigma}_s^4 ds$ is the so-called quarticity of the log price process.

Intuitively, equation (4) is a signal-to-noise ratio, where the signal and noise are the quarticity and squared second moments of intraday return innovations respectively. Equation (4) implies that the higher is the signal-to-noise ratio, the smaller is the optimal sampling interval (or the larger is *m**).

The sample counterpart of equation (4) is:

[3] Several other papers have also analyzed the problem of choosing optimal sampling intervals. However, most of papers impose restrictions on the stochastic process for microstructure noise such as an MA(1) Gaussian structure e.g., Ait-Sahalia and Mykland and Zhang (2005). Oomen (2003) proposes a simulation-based method to generate mean square error (MSE) plots of realized volatility for different sampling intervals. Bandi and Russell (2005) show that the Gaussian assumption is not necessary and that optimal sampling intervals can be estimated by simply using sample moments of the (contaminated) high frequency data to identify moments of the underlying noise process.

$$\widehat{m}^* = \left(\frac{\widehat{Q}_t}{\widehat{\lambda}}\right)^{1/3} \tag{5}$$

where

$$\widehat{Q}_t = \frac{m}{3}\sum_{j=1}^{m}\widehat{r}_{j,t}^4 \ , \tag{6}$$

$$\widehat{\lambda} = \left(\frac{\sum_{t=1}^{T}\sum_{j=1}^{m}\widehat{r}_{j,t}^2}{Tm}\right)^2 \tag{7}$$

and *T* is the number of daily observations.

We use the above rule of thumb to estimate optimal sampling intervals for each stock. Bandi and Russell (2005) show that using relatively low intra-period frequencies such as 15 minutes to compute $\hat{Q}_t$ lead to very reliable estimates of the mean square error. We adopt this approach.

3. Data

Following Andersen et al. (2001), we focus on the 30 stocks that comprise the Dow Jones Industrial Average stock index (hereafter, the Dow Jones 30). As these stocks are actively traded, nonsynchronous trading is not an issue. To mitigate the bid-ask bounce effect, we compute intra-day returns using mid quotes.

Intra-day price data are obtained from the Trade-and-Quotations (TAQ) database. The TAQ contains continuously recorded data of all trades and quotes for stocks listed on the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) and the National Association of Security Dealers Automated Quotation system (NASDAQ). TAQ data is available from January 1993. Our sample period starts from January 1993 and ends in June 2004.

Below, we detail how returns, trading volume and realized volatility are computed in this study.

Returns

For each stock, we obtain its daily continuously compounded returns from the CRSP database. All returns are adjusted for stock splits and dividends. Daily unexpected returns are computed by subtracting daily expected returns from the actual daily returns. To model expected returns, we use a simple MA(1) model.

Trading Activity Variables

We measure trading volume in two ways: daily number of shares traded and daily number of trades. While the first definition is more commonly used in the literature, we also use number of trades since Jones et al. (1994) and Chan and Fong (2006) find that number of trades explains more of the variations in stock returns volatility than total volume. This result is consistent with the "stealth trading" hypothesis that informed investors tend to break up large trades into smaller ones to avoid revealing their private information (Barclay and Warner 1993).

Data on trading volume and number of trades are obtained from the TAQ database. We do not use CRSP for volume because the volume dataset in CRSP has numerous transcription errors. For instance, on certain days, volume is zero or omits the last two digits. TAQ appears to be more reliable. However, TAQ volume is not adjusted for stock splits. We adjust volume for stock splits using the adjustment factor provided in CRSP.

Realized Volatility

For realized volatility, intraday returns are computed using quote-to-quote midpoint prices from TAQ. For stocks that have NYSE as the primary market, quotes from NYSE and the Midwest (Chicago) are used. Nasdaq codes are used for two stocks in our sample (Intel and Microsoft). Only quotes from 10:00 to 16:00 are retained to ensure that our sample contains no opening and closing quotes. A quote is deleted if the bid-ask spread and/or price change exceeds 10%. It is also discarded if the bid price is equal or higher than the offer price, and if the offer or bid price is zero.

As a preliminary check on our computer codes, we replicated the results of Bandi and Russell (2005) using the same sample as theirs. This sample comprises firms in the S&P100 index for the month of February 2002. We find that our estimates of the optimal sampling intervals are very close to theirs. We then proceed to apply the methodology to derive optimal sampling intervals for the Dow Jones 30.

4. Preliminary Data Analysis

Table 1 presents descriptive statistics of realized volatility for the Dow Jones 30. The second column shows the optimal sampling frequency (in minutes and seconds). The mean and median sampling interval is about 7 minutes. This is higher than the mean of 4 minutes reported by Bandi and Russell for S&P 100 stocks. However, since their study covers only February 2002 whereas ours spans over ten years, the difference suggests that microstructure noise may have been more serious in earlier years. Most (twenty three) stocks have sampling intervals above 5 minutes, the interval that is used in many previous studies of realized volatility.

Table 1 shows that the distribution of realized volatility is positively skewed and highly leptokurtic. However, log realized volatility is approximately Gaussian. The mean kurtosis for log realized volatility is 3.47 compared to 7.46 for raw realized volatility. The distribution of log realized volatility is also significantly less skewed compared to the original series. These results are similar to Andersen et al. (2001) for the Dow Jones 30, albeit over a different sample period.

Table 2 presents summary statistics for daily returns and contain no surprises. The return distribution contains many outliers, as evidenced by high excess kurtosis for all stocks, but is approximately Gaussian when returns are standardized by log realized volatility. This result, together with the lognormality of realized volatility, suggests that daily returns are well described by a lognormal-normal mixture, consistent with the MDH introduced by Clark (1973).

5. Volatility-Volume Regressions

5.1 Basic Model

Bessembinder and Seguin (1993) use regressions to examine the impact of volume and open interest on daily absolute returns of eight futures markets. Daigler and Wiley (1999) perform volatility-volume regressions with daily absolute returns and log price range as volatility proxies. These studies are motivated by information theories that attribute the volatility-volume relation to information arrivals (Clark 1973, Andersen 1996), informed trading (Kyle 1985) and uninformed trading (Foster and Viswanathan 1990, Shalen 1993)[4].

In contrast to these papers, we are primarily interested in volatility-volume regressions as a tool to understand volatility predictability. To obtain more reliable inferences on volatility dynamics, we focus on realized volatility instead of absolute or squared returns which are highly inefficient. Realized volatility is also more efficient than the range if log prices follow Brownian diffusions and if realized volatility is obtained by sampling returns at sufficiently high frequencies (Andersen and Bollerslev 1998).

The lognormality of realized volatility suggests the use of standard linear Gaussian techniques for estimation and inference about the latent volatility process. We fit simple autoregressions to log realized volatility to capture the ARCH effect, and

[4] In the MDH, a mixing variable, typically the number of information arrivals causes volatility-volume correlation. Microstructure models such as Kyle (1985) argue that more informed trading takes place when the market is “thick”, and that price volatility increases when there is more informed trades. On the other hand, the dispersion of beliefs hypothesis (Shalen 1993) attributes the volatility-volume correlation mainly to uninformed trades because uninformed traders cannot distinguish between volume induced by changes in fundamentals from volume induced by non-informational causes. Informed traders, who generally have more homogeneous beliefs, mitigate volatility.

document for the first time, the influence of expected and unexpected volume on realized volatility predictability.

Our regressions takes the form $Log\ RV_t = E_{t-1}[\log RV_t \,|\, \boldsymbol{\Omega}_{t-1}] + \varepsilon_t$, where the information set Ω is a vector of regressors which include lagged realized volatilities and volumes, and ε_t is mean zero, serially uncorrelated and has unspecified distribution. It is well known that least square estimators are consistent even if ε_t is non-Gaussian (Greene 2000).

We begin with a simple AR(1) specification for log realized volatility in the spirit of the EGARCH(1,1) model from Nelson (1991):

$$\log RV_t = \omega_t + \alpha \,|\, z_{t-1} \,|\, + \delta z_{t-1} + \beta \log RV_{t-1} + \varepsilon_t \qquad (8)$$

where z_t denotes unexpected return, ε_t represents volatility innovation, and δ is used to account for a potential "leverage effect", i.e., the tendency for bad news to increase volatility more than good news of similar magnitude (Black 1976, Nelson 1991). Without the z_{t-1} terms, equation (8) becomes

$$\log RV_t = \omega_t + \beta \log RV_{t-1} + \varepsilon_t \qquad (9)$$

which is a discrete version of the lognormal stochastic volatility model introduced by Taylor (1986) and applied to option pricing by Hull and White (1987), Melino and Turnbull (1987) and Wiggins (1987)[5].

Equation (8) treats realized volatility as stochastic. This contrasts with GARCH models where volatility is specified as a deterministic function of past return shocks. Theory strongly suggests that volatility is stochastic since private information flow is a primary source of volatility (Glosten and Milgrom 1985, Ross 1989).

[5] Nelson (1990) and Nelson and Foster (1994) show that both the EGARCH(1,1) and lognormal SV models converge in probability to their diffusion limits as we approach continuous time. This implies, for example, that if the leverage effect is significant in discrete time, it will also be intrinsic to the volatility process in continuous time.

We estimate equation (8) using OLS. We use an MA(1) model to account for a small amount of negative serial correlation in daily returns, and compute unexpected returns using this MA(1) filter. For volatility, we add a vector of calendar dummy variables to the volatility equation to account for lower volatility on Mondays compared to other days of the week (Foster and Viswanathan 1990) as well as lower volatility around holidays[6],

Table 3 reports the OLS results. For brevity, estimates for calendar dummies are not reported, but consistent with past studies, we find statistically significant Monday and holiday effects. The average R^2 across the 30 stocks is 0.42, which is a reasonably good fit considering that the model has only one lag. The fit is much better than the R^2 range of 0.07 to 0.12 obtained by Fleming et al. (2001) from regressing daily squared returns on GARCH-fitted volatility. This underscores the importance of working with a precise volatility measure.

The parameter δ is negative and significant for all stocks, confirming the leverage effect documented by Black (1976) and Nelson (1991). Volatility persistence (the ARCH effect) is indicated by β or alternatively, the first order serial correlation coefficient of the fitted volatility which takes into account of the presence of dummy variables in the model. Estimates for these two parameters are very close, with a mean value of about 0.60. Volatility persistence implied by these estimates is higher than those derived using noisy volatility proxies. For example, Bessembinder and Seguin (1993) find that the sum of lagged absolute return coefficients from their AR(10) model range from only 0.31 for Treasury bill futures to 0.66 for cotton futures, with a mean estimate of 0.5 for all eight futures contracts. Measurement errors in their volatility proxy explain the low degree of volatility persistence.

To assess the relative impact of expected and unexpected volume on realized volatility, we add volume to the previous regression. Unlike previous studies, our

---

[6] Holiday dummies are added for one day prior to and after four public holidays: New Year, Independence, Thanksgiving and Christmas Day.

motivation is not to test theories about the volatility-volume relation, but to quantify the influence of expected and unexpected volume on overall volatility predictability.

Studies of the volatility-volume relationship typically insert contemporaneous volume into a volatility equation. For example, Lamoureux and Lastrapes (1990) insert volume into a GARCH(1,1) equation and find that volatility persistence virtually disappear. As mentioned, Bessembinder and Seguin (1993) also report relatively low volatility persistence when volume and open interests are added to a ten-lag volatility autoregression. Fleming et al. (2001) point out that simply adding contemporaneous volume to the volatility regression is problematic since the volume-volatility relationship is endogenous and this endogeneity leads to bias in the volume coefficient. In addition, inserting volume into the volatility equation implicitly restricts past volume and past return shocks to generate the same degree of volatility persistence. In our context, adding volume (V_t) to equation (8) gives:

$$\log RV_t = \omega_0 + \alpha |z_{t-1}| + \delta z_{t-1} + \beta \log RV_{t-1} + \lambda V_t \ . \qquad (10)$$

By repeated substitution, equation (10) can be also written as:

$$\log RV_t = \omega_0 + \sum_{i=0}^{\infty} \beta^i (\alpha |z_{t-i-1}| + \delta z_{t-i-1}) + \lambda \sum_{i=0}^{\infty} \beta^i V_{t-i} \ . \qquad (11)$$

Thus, for the same number of lags, the effect of past return shocks on current volatility is proportional to that of past volume shocks due to the common decay parameter β. This setup is restrictive if realized volatility is more strongly correlated with unexpected volume than with expected volume. Since unexpected volume is serially uncorrelated, it will *not* generate the same degree of volatility persistence as past return shocks.

Fleming et al. (2006) model volume and volatility as a bivariate AR(1) stochastic process with persistent and nonpersistent components and find that there is a very large nonpersistent component in daily stock volatility which is closely tied to unexpected volume. This is consistent with the original version of the MDH (Clark 1973) in which information is assumed to arrive randomly. The persistent component in volatility is found to be only weakly correlated with volume (expected or unexpected), indicating that the ARCH effect is not primarily driven by news per se.[7] As a result, the ARCH effect does not disappear even after accounting for the dynamics of trading volume.

The findings of Fleming et al. (2006) motivate a more flexible regression model which allows for an ARCH effect that is orthogonal to volume. Specifically, let $h_t = E_{t-1}RV_t$ be the conditional expectation of realized volatility for day *t*. We assume that

$$\log h_t = \omega_0 + \alpha \,|\, z_{t-1} \,| + \delta z_{t-1} + \beta \log h_{t-1} + \gamma E_{t-1}V_t \,. \qquad (12)$$

This equation says that volatility forecasts are formed using past returns, past volatility forecasts and expected volume, $E_{t-1}V_t$. To capture the impact of signed unexpected volume (v_t) on volatility, we assume that:

$$RV_t = h_t e^{\lambda v_t} \qquad (13)$$

which constrains volatility to be non-negative, while allowing unexpected volume to take both positive and negative values. We expect λ to be positive, which implies that an increase (decrease) in unexpected volume will lead to an increase (decrease) in realized volatility. The larger is λ, the more sensitive is realized volatility to changes in unexpected volume. Taking log for (13) and combining it with (12) leads to the following model:

$$\log RV_t = \omega_0 + \alpha \,|\, z_{t-1} \,| + \delta z_{t-1} + \beta \log RV_{t-1} + \gamma (EV_t - \beta EV_{t-1}) + \lambda (v_t - \beta v_{t-1}) \,. \qquad (14)$$

Equation (14) implies that volatility persistence is orthogonal to volume, which is roughly consistent with the findings of Fleming et al. (2006). To the extent that there is collinearity between the conditional variance and volume, this will bias our results against finding a significant relationship between volume and volatility.

There is no theoretical consensus on what should be the sign for expected volume. On the one hand, an increase in expected volume may reduce volatility because of improved liquidity. However, there is evidence that uninformed investors trade more when recent volume has been high (Easley et al. 1997). Furthermore, Dailey and Wiley (1999) show that the volatility-volume relation is mainly driven by uninformed trades.

[7] For fifteen out of twenty stocks in their sample, the correlation between the conditional return variances obtained by estimating a GARCH(1,1) model and the persistent factor common to both squared returns and volume is less than 0.4.

Taken together, these studies suggest that higher expected volume may lead to increased volatility. Our results show that this is the case.

In contrast to expected volume, there are strong theoretical reasons why volatility should be positively correlated with unexpected volume even though their serial correlation properties are very different. Information theories such as the MDH argue that information arrivals stimulate trading and increase volatility, and unexpected volume is a plausible proxy for information flow. Dispersion of beliefs models such as Foster and Viswanathan (1990) and Shalen (1993) posit that the volatility-volume relation depends not only on information arrivals, but also who trades and why. Specifically, it is argued that since uninformed traders lack precise information signals, they may react to all unexpected volume increases as if these are information. Dispersion models therefore suggest that the volatility-volume relationship is primarily due to uninformed trades.

To estimate equation (14), we need proxies for expected and unexpected volumes. An important first step in modeling expected volume is to test whether volume has unit roots, controlling for time trends. Using the Phillips-Perron unit root test with time trend, we find no evidence of unit roots in raw volume for all stocks. We therefore proceed to model expected volume using a stationary autoregressive model, adjusting for prominent seasonal effects (specifically, Monday and holiday effects)[8]. Our approach is similar to Bessembinder and Seguin (1993) who use an AR(10) specification for futures volume. Using the Schwarz (1978) criterion, we choose a lag limit to 15 for log volume to incorporate the range of significant lags among the thirty firms. Unexpected volume is just the difference between the actual log volume and the expected log volume based on estimates of the AR(15) model.

Tables 4 and 5 present regression results with unexpected volume. Table 4 defines volume as daily number of shares traded, while Table 5 defines volume as daily number of trades. We highlight two key results of these regressions. First, adding

[8] We also explore a specification that includes linear and quadratic time trends, but find very little difference in the results. We therefore report results based on the model without any time trend.

unexpected volume increases the adjusted R^2 quite appreciably for all stocks. The mean R^2 is 0.51 in Table 4 compared to 0.42 in Table 3. The mean R^2 is even higher when volume is defined by number of trades (0.55). This is consistent with the results of Jones et al. (1993) that number of trades explains more of the changes in daily volatility than average trade size[9]. The higher explanatory power of number of trades may be due to "stealth trading" whereby informed traders break up large trades into smaller ones to avoid revealing their private information to the market.

Secondly, adding unexpected volume does not reduce volatility persistence. Looking at Table 5 where volume is measured using number of trades, we see that the average lag-one autocorrelation of the fitted log realized volatility is 0.59, which is only slightly lower than the earlier reported value of 0.63 in Table 3. Thus, past volatilities have significant information content for predicting future volatility, even after accounting for the effect of unexpected volume.

Table 6 shows results when both expected and unexpected volumes are included in the volatility equation. The coefficient for expected volume is statistically significant but the mean R^2 for this specification is 0.56, which is only marginally higher than the mean R^2 for the previous regression. Consistent with theory, expected volume plays a relatively minor role in explaining changes in realized volatility.

Our results are contrary to those of Lamoureux and Lastrapes (1990) who find that volume subsumes the ARCH effect. Which of these inferences are correct? To explain, recall that Lamourex and Lastrapes insert volume directly into the GARCH equation. As mentioned, this implicitly assumes that past return and volume shocks generate the same degree of volatility persistence. Thus, one interpretation of their result is that volatility contains a large persistent component that is closely tied to volume. Since volume is also serially correlated, it substitutes for the ARCH effect.

[9] Using a nonparametric method, Ane and Geman (2000) show that number of trades is the stochastic clock that recovers normality in daily stock returns. This result is consistent with the MDH with number of trades as the indicator of the intensity of information flow.

A problem with this interpretation is that one has to assume that *total* volume is a good proxy for information flow. This does not seem plausible since theory suggests that volatility should correlate more highly with news (unexpected volume) than with anticipated information (expected volume). Therefore, an alternative explanation of the Lamoureux-Lastrape effect is that volatility contains a large *random* component, which is closely related to unexpected volume. Inserting volume into the GARCH equation injects this nonpersistent component into the volatility process, causing the ARCH effect to diminish.

Empirical results support the second interpretation. If unexpected volume is the correct proxy for information flow, then it should have a greater role in explaining volatility than expected volume, which is indeed what we find. Furthermore, Fleming et al (2006) show that empirically, "a large nonpersistent component of return volatility is closely related to the contemporaneous nonpersistent component of the trading process". Our results are in line with theirs, and confirm that the ARCH effect is indeed a robust feature of realized volatility. Nonetheless, this does not imply that realized volatility is highly predictable based on past realized volatilities. The fact that unexpected volume is an important component in realized volatility limits overall volatility predictability. We address this issue more fully in the next section.

5.2 Higher Order Autoregressions

Higher order autoregressions can better able to capture the long memory dynamics of realized volatility than the AR(1) specification discussed in the previous section.[10] Previous studies have successfully applied such models to different asset classes. Bessembinder and Seguin (1993) find that an AR(10) model adequately captures

[10] The motivation for using long autoregressions to model realized volatility may be understood as follows. Suppose we model log realized volatility as $\phi(L)\log RV_t = \varepsilon_t$ where $\phi(L)$ is a possibly infinite-order lag polynomial satisfying $\sum_{j=0}^{\infty}|\phi_j|\, j^{0.5} < \infty$, L is the lag operator, and ε_t is a white noise process with zero mean, finite second and finite fourth-moment. With sufficient lags, this process approximates a long memory (fractionally integrated) process arbitrarily well.

the dynamics of absolute returns for eight futures markets. Bollerslev and Wright (2001) show that for the DM-dollar exchange rate, an AR(10) model works better in forecasting future realized volatility than standard GARCH models. For the Dow Jones 30, we find that as many as fifteen lags are needed to obtain white noise residuals. The results of fitting an AR(15) model for realized volatility are shown in Table 7.

On the whole, the AR(15) model fits the data quite well. First, the mean adjusted R^2 is 0.52, compared to 0.42 for the one-lag specification[11]. Second, the Durbin-Watson statistic is close to 2.0 for all stocks, indicating that there is no significant lag-one serial correlation in the residuals. Results of a Ljung-Box test (not shown) also show that there are no significant higher order serial correlations.

The sum of all fifteen lagged volatility coefficients is a measure of the ARCH effect for this model. We report this statistic along with t-statistics that are averaged over the fifteen lags. The ARCH effect averages 0.87, and is never less than 0.80 for any stock. It is interesting to note that the level of volatility persistence implied by these estimates is much higher than those reported by Bessembinder and Seguin (1993) using absolute returns.

When we add expected volume to the model, the adjusted R^2 increases very marginally. Adding unexpected volume, however leads to a significant improvement in adjusted R^2 for all firms. These results are qualitatively similar to those of the AR(1) model. For brevity, we only report final regression results for the AR(15) model incorporating both expected and unexpected volume. This is shown in Table 8.

The sum of lagged volatility coefficients has an average value of 0.90, confirming that the ARCH effect is robust to the inclusion of volume. The lag-one autocorrelation coefficient of the fitted volatility, shown in the last column, has a mean value of 0.83

---

[11] Andersen et al. (2001) report R^2s of between 0.35 and 0.50 for realized volatility of daily exchange rates based on a Gaussian long memory model. Our results are therefore roughly in line with theirs.

compared to 0.92 in Table 7. Therefore, the presence of unexpected volume reduces overall volatility predictability.

The influence of all random components on realized volatility can be seen in Figure 2 for IBM stock. The top panel plots the daily log realized volatility for IBM and the middle panel shows the fitted log realized volatility generated by regression estimates of the AR(15) model without volume. We see that the fitted volatility tracks the actual series quite well (the correlation is 0.71 between them). Despite this, it fails to capture large swings in actual realized volatility. These swings, shown in the bottom panel, reflect the combined effect of unexpected volume and volatility innovations (regression residuals). The relative importance of these predictable and random components will be discussed in the next section.

6. Random Components in Realized Volatility

To quantify the relative impact of predictable and random components, we decompose the variance of log realized volatility into (a) calendar effects, (b) volatility persistence related to lagged volatilities and lagged return shocks, and (c) a nonpersistent component comprising of unexpected volume and volatility innovations (regression residuals). The first two components are predictable ex-ante, while the last component is random. Table 9 shows the results of this decomposition based on estimates of the AR(15) model with expected and unexpected volumes. The percentage contribution of each component is plotted in Figure 3. Note that due to measurement errors, the percentage contributions of the predictable and random components do not add up to one exactly (the mean deviation is 9.3%).

The bottom row of Table 9 shows that the mean value of the predictable and random component is 0.406 and 0.326 respectively. In percentage terms, a little more than half of the variations in realized volatility are predictable. As expected, most of this predictability is due to the ARCH effect. About 15% and 35% of variations in realized volatility are due to fluctuations in unexpected volume and volatility innovations respectively. This implies that 30% of the stochastic component in realized volatility reflects information flow as captured by unexpected volume. Since unexpected volume is serially uncorrelated, this component sets an important limit on how far we can improve the accuracy of volatility predictions. To be more specific, suppose we can reduce the relative impact of regression residuals half e.g., by adding other exogenous predictors. This still leaves about one third of the variance of realized volatility unforecastable. The prospect of finding structural models which have high predictive content for volatility does not seem good as shown by Cutler, Poterba and Summers (1989), Pagan and Schwert (1990) and Haugen, Talmor and Torous (1991) among others. Summarizing this literature, Schwert (1989) observes that even with a century of data, "there is (only) weak evidence that macroeconomic volatility provides incremental information about future stock return volatility". Thus, despite having a large ARCH component, predicting realized volatility accurately remains to be a highly challenging task.

7. Conclusion

Availability of high-frequency data has made possible the direct and precise measurement of asset returns volatility. In particular, realized volatility, corrected for microstructure biases, provides an accurate proxy for latent volatility. This paves the way for realized volatility to be used in standard applications such as derivative pricing, portfolio management decisions, and value-at-risk calculations (Cotter 2004). These applications all rely on accurate volatility forecasts. In this paper, we provide some perspectives on the extent to which realized volatility is predictable. Two specific issues relating to predictability were examined. The first issue is whether there is a robust ARCH effect in realized volatility. Previous studies using GARCH models (Lamoureux and Lastrapes 1990) models dispute the robustness of the ARCH effect in the presence of trading volume. Other studies such as Bessembinder and Seguin (1993) report relatively low volatility persistence in volatility-volume regressions when volatility is measured using absolute returns. In contrast to these studies, we find that realized volatility does exhibit significant ARCH effects when volume is properly included in the volatility equation. In fact, ARCH effects account for slightly over 50% of daily movements in realized volatility. Hence, past realized volatilities do have predictive content. Moreover, this predictive content can be adequately captured by fitting simple autoregressions.

The second issue that we examined is the size of the stochastic component in realized volatility. If realized volatility contain a large stochastic component, then volatility predictability may be low even if ARCH effects are strong. Our results show unexpected volume explains much more of the temporal variations in realized volatility than expected volume, and accounts for about 30% of the total stochastic component. While using more complicated structural models may help to reduce the influence of volatility innovations, the prospect that any of these models can produce highly accurate volatility predictions seems bleak given that most of volatility changes cannot be associated with economic fundamentals. Achieving high predictive accuracy for dailly realized volatility remains to be a daunting task.

References

Ait-Sahalia, P. Mykland and Zhang, L. (2005), ‘How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise’, *Review of Financial Studies*, 18: 351-416.

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 Bollerslev, T. (1998), ‘Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts’, *International Economic Review*, 39: 885-905.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001), ‘The Distribution of Realized Stock Return Volatility’, *Journal of Financial Economics*, 61: 43-76.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), ‘Modeling and Forecasting Realized Volatility’, *Econometrica*, 71: 579-625.

Andersen, T.G. and Sorensen, B.E. (1997), ‘GMM and QML Asymptotic Standard Deviations in Stochastic Volatility Models’, *Journal of Econometrics*, 76: 397-403.

Ane, T. and Geman, H. (2000), ‘Order Flow, Transaction Clock and Normality of Asset Returns’, *Journal of Finance*, 55: 2259-2284.

Barclay, M.J., and Warner, J.B. (1993), ‘Stealth Trading and Volatility: Which Trades Move Prices?’, *Journal of Financial Economics*, 34: 281-305.

Back, K. (1991), ‘Asset Prices for General Processes’, *Journal of Mathematical Economics*, 20: 317-395.

Bandi, F., and Russell, J.R. (2005), ‘Separating Microstructure Noise from Volatility and Optimal Sampling’, *Journal of Financial Economics*, 79(3): 537-568.

Barndorff-Nieslen, O.E., and Shephard, N. (2001), 'Non-Gaussian Ornstein-Uhlenbeck-Based Models and Some of Their Uses in Financial Economics', (with discussion), *Journal of the Royal Statistical Society*, Series B, 63: 167-241

Barndorff-Nieslen, O.E., and Shephard, N. (2002), 'Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models', *Journal of the Royal Statistical Society*, Series B, 64, Part 2: 253-280.

Bessembinder, H. and Seguin, P. (1993), 'Price Volatility, Trading Volume, and Market Depth: Evidence from Futures Markets', *Journal of Financial and Quantitative Analysis*, 28: 21-29.

Black, F. (1976), 'Studies in Stock Price Volatility Changes, Proceedings of the 1976 Business Meeting of the Business and Statistics Section', *American Statistical Association*, 177-181.

Bollerslev, T. and Wright, J.H. (2001), 'High-Frequency Data, Frequency Domain Inference and Volatility Forecasting', *Review of Economics and Statistics*, 83: 596-602.

Chan, C.C., and Fong, W.M. (2006), 'Realized Volatility and Transactions', *Journal of Banking and Finance*, 30(7): 2063-2085.

Clark, P.K. (1973), 'A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices', *Econometrica*, 41: 135-155.

Cotter, J. (2004), 'Minimum Capital Requirement Calculations for U.K. Futures', *Journal of Banking and Finance*, 24: 193-220.

Cutler, D.M., Poterba, J.M. and Summers, L.H. (1989), 'What Moves Stock Prices?', *Journal of Portfolio Management*, 15(3): 4-12.

Daigler, R.F., and Wiley, M. (1999), 'The Impact of Trader Type on the Futures Volatility-Volume Relation', *Journal of Finance*, 54: 2297-2316.

Easley, D., Keifer, N.M. and O'Hara, M. (1997), 'The Information Content of the Trading Process', *Journal of Empirical Finance,* 4:159-186.

Fleming, J., Kirby, C. and Ostdiek, B. (2001), 'Stochastic Volatility, Trading Volume and the Daily Flow of Information', working paper, University of New South Wales.

Fleming, J., Kirby, C. and Ostdiek, B. (2006), 'Stochastic Volatility, Trading Volume and the Daily Flow of Information', *Journal of Business*, 79(3): 1551-1590.

Foster, F.D. and Viswanathan, S. (1990), 'A Theory of Intraday Variations in Volume, Variance and Trading Costs in Securities Markets', *Review of Financial Studies*, 3: 593-624.

French, K.R. and Roll, R. (1986), 'Stock Return Variances: the Arrival of Information and the Reaction of Traders', *Journal of Financial Economics*, 17: 5-26.

French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987), 'Expected Stock Returns and Volatility', *Journal of Financial Economics*, 19: 3-29.

Ghysels, E., Harvey, A. and Renault, E. (1996), 'Stochastic Volatility', in G.S. Maddala and C.R. Rao (eds.), *Statistical Methods in Finance, Handbook of Statistics*, Volume 14, Amsterdam: North-Holland.

Glosten, L.R., and Milgrom, P.R. (1985), 'Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders', *Journal of Financial Economics*, 14: 71-100.

Greene, W.H. (2000), 'Econometric Analysis', NY: Prentice Hall.

Harvey, A., Ruiz, E. and Shephard, N. (1994), 'Multivariate Stochastic Variance Models', *Review of Economic Studies*, 61: 247-264.

Haugen, R.A., Talmor, E. and Torous, W.N. (1991), 'The Effect of Volatility Changes on the Level of Stock Prices and Subsequent Expected Returns', *Journal of Finance*, 46: 985-1007.

Hull, J. and White, A. (1987), 'The Pricing of Options on Assets with Stochastic Volatilities', *Journal of Finance*, 42: 281-300.

Jones, C.M., Kaul, G. and Lipson, M.L. (1994), 'Transactions, Volume and Volatility', *Review of Financial Studies*, 7: 631-651.

Karatzas, L. and Shreve, S.E. (1991), 'Brownian Motion and Stochastic Calculus', 2nd edition, New York: Springer-Verlag.

Kyle, A. (1985), 'Continuous Auctions and Insider Trading', *Econometrica*, 53: 1315-1335.

Lamoureux, C.G. and Lastrapes, W.D. (1990), 'Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects', *Journal of Finance*, 45: 221-229.

Liesenfeld, R. (2001), 'A Generalized Bivariate Mixture Model for Stock Price Volatility and Trading Volume', *Journal of Econometrics*, 104: 141-178.

Lo, A., and C. MacKinlay, C. (1990), 'An Econometric Analysis of Nonsynchronous Trading', *Journal of Econometrics*, 45: 181-212.

Melino, A., and Turnbull, S.M. (1987), 'Pricing Foreign Currency Options with Stochastic Volatility', *Journal of Econometrics*, 45: 239-265.

Nelson, D.B. (1990), 'ARCH Models as Diffusion Approximations', *Journal of Econometrics*, 45: 7-39.

Nelson, D.B. (1991), 'Conditional Heteroskedasticity in Asset Returns: a new approach', *Econometrica*, 59: 347-370.

Nelson, D.B., and Foster, D.P. (1994), 'Asymptotic Filtering Theory for Univariate ARCH Models', *Econometrica*, 62: 1-41.

Newey, W.K., and West, K.D. (1987), 'A Simple Positive Definite, Heteroskasdicity and Autocorrelation Consistent Covariance Matrix', *Econometrica*, 55: 703-708.

Oomen, R. (2003), 'Properties of Realized Variance for Pure-Jump Processes: Calendar Time Sampling versus Business Time Sampling'. Mimeo.

Ross, S. (1989), 'Information and Volatility: The No-Arbitrage Martingale Approach to Timing and Resolution Irrelevancy', *Journal of Finance*, 44: 1-18.

Schwarz, G. (1978), 'Estimating the Dimensions of a Model', *Annals of Statistics*, 6: 461-464.

Schwert, G.W. (1989), 'Why Does Stock Market Volatility Change Over Time?', *Journal of Finance*, 44: 1115-1153.

Pagan, A., and Schwert, G.W. (1990), 'Alternative Models for Conditional Stock Volatility', *Journal of Econometrics*, 45: 267-290.

Shalen, C.T. (1993), 'Volume, Volatility and the Dispersion of Beliefs', *Review of Financial Studies*, 6: 405-434.

Taylor, S.J. (1986), 'Modeling Financial Time Series', Wiley, New York.

Wiggins, J.B. (1987), 'Option Values under Stochastic Volatility: Theory and Empirical Estimates', *Journal of Financial Economics*, 19: 351-372.

Table 1. Summary Statistics for Realized Volatility

Summary statistics are presented for daily realized variance (RV) and log of realized variance (Log RV) for the Dow Jones 30. The sample period is January 1, 1993 to June 30, 2004. Realized variance is computed using optimal intraday sampling intervals as described in the text. The second column reports the optimal sampling interval (in minutes and seconds).

Firm	Sampling	RV				Log RV			
	Interval	Mean	Stdev	Skew	Kurtosis	Mean	Stdev	Skew	Kurtosis
AA	08:52	2.73	2.88	4.31	36.07	0.66	0.81	0.10	3.42
AIG	06:16	2.08	2.26	6.87	6.25	0.41	0.78	0.16	3.25
AXP	09:11	3.08	3.76	5.28	6.71	0.73	0.87	-0.01	3.71
BA	08:30	2.48	3.38	11.87	6.54	0.53	0.82	0.29	3.54
C	06:45	2.95	3.99	9.65	6.68	0.71	0.82	0.29	3.68
CAT	06:17	2.47	2.48	3.62	6.18	0.57	0.80	0.09	3.18
DD	07:05	2.32	2.40	4.15	6.02	0.50	0.80	0.19	3.02
DIS	08:05	2.87	4.73	16.16	6.79	0.68	0.81	0.33	3.79
GE	06:00	2.22	2.74	6.79	6.48	0.44	0.81	0.32	3.48
GM	07:23	2.14	2.47	6.67	6.74	0.45	0.73	0.51	3.74
HD	07:21	2.86	3.52	8.04	6.63	0.70	0.79	0.40	3.63
HON	07:33	3.05	4.57	9.58	6.94	0.70	0.85	0.34	3.94
HPQ	06:01	4.63	5.58	3.69	6.00	1.10	0.90	0.33	3.00
IBM	05:44	2.33	2.66	7.88	6.37	0.51	0.79	0.20	3.37
INTC	03:57	3.88	4.87	4.25	5.74	0.85	0.99	0.17	2.74
JNJ	08:45	1.82	2.12	8.28	6.70	0.28	0.76	0.18	3.70
JPM	04:03	2.82	4.79	13.82	6.47	0.58	0.87	0.54	3.47
KO	07:33	1.82	1.90	5.11	6.58	0.30	0.74	0.31	3.58
MCD	10:24	2.14	2.50	5.98	6.80	0.44	0.76	0.30	3.80
MMM	08:25	1.82	2.03	4.53	6.39	0.23	0.84	0.12	3.39
MO	04:53	2.60	5.48	19.08	7.32	0.55	0.81	0.45	4.32
MRK	09:43	2.04	2.43	6.36	7.22	0.39	0.77	0.16	4.22
MSFT	03:38	2.53	3.34	9.53	5.93	0.47	0.94	0.06	2.93
PFE	05:46	2.31	2.61	8.33	6.78	0.54	0.73	0.28	3.78
PG	07:44	1.91	2.27	7.46	6.45	0.31	0.80	0.16	3.45
SBC	08:05	2.59	2.99	4.65	6.35	0.58	0.82	0.29	3.35
UTX	04:59	2.27	2.91	10.28	6.23	0.41	0.89	0.06	3.23
VZ	03:20	3.11	3.06	3.12	5.71	0.80	0.80	0.24	2.71
WMT	10:17	2.80	3.66	12.06	6.67	0.68	0.80	0.20	3.67
XOM	03:26	1.96	2.06	4.18	6.07	0.33	0.80	0.25	3.07
Mean	**06:52**	**2.55**	**3.22**	**7.72**	**7.46**	**0.55**	**0.82**	**0.24**	**3.47**

Table 2. Summary Statistics for Returns

Summary statistics are presented for daily continuously compounded open-to-close returns for the Dow Jones 30. The sample period is January 1, 1993 to June 30, 2004. Entries under "Standardized Returns" are statistics for daily returns divided by log of daily realized standard deviation.

Firm	Returns				Standardized Returns			
	Mean	Stdev	Skew	Kurtosis	Mean	Stdev	Skew	Kurtosis
AA	0.08	2.20	0.38	5.80	0.04	1.33	0.23	2.95
AIG	0.08	1.81	0.28	5.84	0.07	1.23	0.25	3.06
AXP	0.10	2.15	0.17	5.15	0.09	1.27	0.26	3.12
BA	0.06	2.06	-0.31	9.26	0.07	1.27	0.13	2.93
C	0.10	2.23	0.63	12.15	0.11	1.30	0.30	3.65
CAT	0.09	2.08	0.14	5.46	0.07	1.34	0.15	2.93
DD	0.05	1.86	0.16	5.65	0.05	1.23	0.20	3.14
DIS	0.05	2.16	0.12	9.40	0.06	1.29	0.27	4.20
GE	0.08	1.80	0.18	6.59	0.10	1.23	0.27	3.14
GM	0.05	2.03	0.08	5.16	0.06	1.42	0.26	3.00
HD	0.07	2.32	-0.58	13.64	0.08	1.34	0.13	3.88
HON	0.06	2.28	0.26	16.41	0.05	1.25	0.13	3.13
HPQ	0.08	2.80	0.03	6.60	0.08	1.40	-0.12	5.70
IBM	0.09	2.18	0.37	8.53	0.08	1.40	0.30	3.84
INTC	0.12	2.93	-0.10	7.20	0.15	1.77	-0.60	3.20
JNJ	0.07	1.64	-0.17	7.83	0.06	1.24	0.03	3.23
JPM	0.06	2.17	0.39	8.89	0.06	1.26	0.25	3.11
KO	0.05	1.66	0.05	6.36	0.06	1.23	0.22	3.17
MCD	0.05	1.79	0.08	7.17	0.04	1.21	0.20	3.20
MMM	0.07	1.59	0.25	6.44	0.05	1.20	0.22	3.65
MO	0.07	2.11	-0.46	13.81	0.10	1.26	0.12	4.56
MRK	0.05	1.81	0.06	5.30	0.06	1.31	0.15	3.21
MSFT	0.11	2.33	0.13	7.15	0.13	1.66	0.31	3.58
PFE	0.09	1.98	-0.02	4.68	0.09	1.31	0.21	3.02
PG	0.07	1.73	-2.14	43.18	0.09	1.18	0.13	3.39
SBC	0.04	1.97	0.08	5.56	0.06	1.23	0.22	3.09
UTX	0.10	1.89	-1.14	22.06	0.09	1.22	0.10	3.12
VZ	0.00	2.18	0.08	6.07	0.00	1.25	0.17	3.31
WMT	0.06	2.05	0.22	5.02	0.06	1.29	0.23	3.16
XOM	0.03	1.66	0.28	6.92	0.07	1.21	0.06	3.06
Mean	**0.07**	**2.05**	**-0.02**	**9.31**	**0.07**	**1.30**	**0.16**	**3.39**

Table 3. Realized Volatility Regression (AR1 Model)

Results of the following volatility regression for the Dow Jones 30 firms: $\log\sigma_t^2 = \boldsymbol{\omega}'\mathbf{d_t} + \alpha\,|\,z_{t-1}\,| + \delta z_{t-1} + \beta\log\sigma_{t-1}^2 + \varepsilon_t$. The sample period is from January 1, 1993 to June 30, 2004. σ_t^2 is the realized variance for day t , $\boldsymbol{\omega}'\mathbf{d_t}$ is the unconditional mean parameter, $\mathbf{d_t}$ is a vector of calendar (Monday and holiday) dummies, and z_t are unexpected returns computed using an MA(1) model for expected returns. The regression is estimated using OLS, with Newey-West t-statistics reported in parentheses. The column labeled ρ_{RV} shows the first-order autocorrelation of the fitted log realized variance.

	Parameter Estimates				**T-statistics**			**Adj. R^2**	ρ_{RV}
Firm	ω_t	α	δ	β	$t(\alpha)$	$t(\delta)$	$t(\beta)$		
AA	-0.06	0.04	-0.02	0.54	4.59	-3.78	22.83	0.34	0.57
AIG	-0.21	0.06	-0.03	0.55	5.85	-4.46	22.08	0.38	0.60
AXP	-0.18	0.08	-0.03	0.55	7.62	-5.39	20.90	0.41	0.63
BA	-0.21	0.05	-0.02	0.54	4.54	-2.73	23.49	0.36	0.59
C	-0.16	0.05	-0.01	0.60	4.55	-2.41	27.48	0.44	0.64
CAT	-0.16	0.04	-0.02	0.56	3.93	-3.29	23.91	0.37	0.59
DD	-0.20	0.04	-0.02	0.62	3.97	-3.07	32.79	0.44	0.66
DIS	-0.06	0.04	-0.01	0.57	4.86	-2.07	26.26	0.38	0.61
GE	-0.38	0.09	-0.04	0.59	7.56	-5.69	31.47	0.47	0.66
GM	-0.18	0.06	-0.02	0.46	5.90	-3.53	19.65	0.29	0.52
HD	-0.13	0.04	-0.02	0.57	3.55	-3.29	21.81	0.39	0.61
HON	-0.05	0.04	-0.01	0.51	3.87	-2.04	19.98	0.31	0.55
HPQ	-0.02	0.01	-0.02	0.73	0.64	-2.18	19.51	0.57	0.75
IBM	-0.25	0.05	-0.02	0.60	4.87	-3.73	26.28	0.43	0.64
INTC	-0.21	0.04	-0.01	0.76	5.94	-3.80	46.34	0.65	0.80
JNJ	-0.24	0.09	-0.04	0.49	7.24	-4.63	19.68	0.32	0.54
JPM	-0.28	0.06	-0.03	0.69	6.00	-4.95	39.27	0.57	0.74
KO	-0.32	0.07	-0.04	0.59	5.83	-5.37	27.82	0.44	0.65
MCD	-0.14	0.06	-0.02	0.46	4.65	-2.95	19.44	0.26	0.50
MMM	-0.23	0.07	-0.03	0.54	4.60	-3.73	23.12	0.35	0.58
MO	-0.11	0.05	-0.01	0.48	4.42	-0.88	21.35	0.28	0.53
MRK	-0.34	0.06	-0.02	0.46	5.22	-2.74	17.45	0.28	0.51
MSFT	-0.25	0.05	-0.02	0.74	6.95	-4.38	47.68	0.62	0.78
PFE	-0.14	0.04	-0.01	0.54	4.62	-1.84	23.01	0.35	0.58
PG	-0.31	0.03	-0.02	0.62	2.39	-2.68	25.70	0.43	0.65
SBC	-0.12	0.03	-0.01	0.60	2.65	-2.18	24.94	0.40	0.63
UTX	-0.21	0.04	-0.02	0.66	2.46	-2.24	29.71	0.48	0.68
VZ	-0.22	0.03	-0.02	0.69	2.40	-2.51	26.42	0.53	0.72
WMT	-0.10	0.07	-0.02	0.52	6.99	-4.19	21.78	0.35	0.57
XOM	-0.26	0.05	-0.02	0.74	3.59	-2.46	32.44	0.61	0.77
Mean	**-0.19**	**0.05**	**-0.02**	**0.59**	**4.74**	**-3.31**	**26.15**	**0.42**	**0.63**

Table 4. Realized Volatility Regression (AR1 Model) with Unexpected Number of Shares Traded

Results of the following regression for log of realized variance for the Dow Jones 30 firms: $\log\sigma_t^2 = \boldsymbol{\omega}'\mathbf{d_t} + \alpha|z_{t-1}| + \delta z_{t-1} + \beta\log\sigma_{t-1}^2 + \lambda(v_t - \beta v_{t-1}) + \varepsilon_t$. The sample period is from January 1, 1993 to June 30, 2004. σ_t^2 is realized variance for day t, $\boldsymbol{\omega}'\mathbf{d_t}$ is the unconditional mean parameter, $\mathbf{d_t}$ is a vector of calendar (Monday and holiday) dummies, z_t are unexpected returns computed using an MA(1) model for expected returns, and v_t is unexpected log volume. In this table, volume is defined as **number of shares traded**. Expected log volume is obtained by fitting an AR(15) model to log of number of shares traded. The regression is estimated using OLS, with Newey-West t-statistics reported in parentheses. The column labeled ρ_{RV} shows the first-order autocorrelation of the fitted log realized variance.

	Parameter Estimates					**T-statistics**				**Adj. R^2**	ρ_{RV}
Firm	ω_t	α	δ	β	λ	$t(\alpha)$	$t(\delta)$	$t(\beta)$	$t(\gamma)$		
AA	-0.10	0.05	-0.02	0.58	0.64	5.70	-3.56	24.58	17.94	0.47	0.56
AIG	-0.23	0.08	-0.02	0.58	0.64	8.57	-2.90	25.73	18.57	0.47	0.56
AXP	-0.21	0.08	-0.02	0.58	0.72	9.70	-3.62	23.46	20.62	0.52	0.58
BA	-0.22	0.07	-0.01	0.56	0.72	8.35	-1.97	25.91	20.67	0.48	0.54
C	-0.22	0.07	-0.01	0.63	0.63	7.61	-1.26	30.46	18.31	0.54	0.62
CAT	-0.22	0.06	-0.02	0.58	0.58	6.79	-3.71	28.03	20.37	0.47	0.55
DD	-0.20	0.06	-0.01	0.65	0.56	5.95	-2.26	35.59	16.76	0.52	0.63
DIS	-0.09	0.07	0.00	0.58	0.60	8.91	-0.93	30.79	17.05	0.47	0.58
GE	-0.38	0.10	-0.02	0.62	0.70	9.99	-4.05	36.18	17.64	0.55	0.64
GM	-0.21	0.08	-0.02	0.49	0.66	8.49	-2.98	22.68	20.94	0.41	0.45
HD	-0.21	0.06	-0.01	0.58	0.63	5.99	-2.21	24.63	19.13	0.50	0.58
HON	-0.07	0.05	-0.01	0.53	0.51	4.06	-1.15	21.89	17.10	0.40	0.50
HPQ	-0.07	0.04	-0.02	0.73	0.44	2.74	-2.05	20.29	6.00	0.61	0.75
IBM	-0.30	0.09	-0.02	0.64	0.83	10.78	-3.63	33.30	24.74	0.59	0.60
INTC	-0.28	0.07	-0.01	0.79	0.64	11.64	-3.18	58.34	19.30	0.72	0.81
JNJ	-0.24	0.10	-0.03	0.52	0.67	9.54	-4.04	22.96	16.79	0.42	0.51
JPM	-0.36	0.07	-0.02	0.71	0.61	9.74	-3.79	46.57	20.28	0.66	0.73
KO	-0.31	0.09	-0.03	0.61	0.56	8.69	-4.47	30.18	18.21	0.51	0.63
MCD	-0.15	0.07	-0.02	0.50	0.76	7.65	-3.17	23.10	22.95	0.41	0.43
MMM	-0.21	0.08	-0.02	0.58	0.69	6.78	-2.47	27.49	16.45	0.46	0.54
MO	-0.11	0.07	0.00	0.49	0.79	6.51	0.17	22.48	18.42	0.43	0.46
MRK	-0.35	0.07	-0.01	0.49	0.77	6.98	-1.87	20.33	18.92	0.39	0.45
MSFT	-0.32	0.09	-0.02	0.76	0.70	13.78	-3.67	58.45	21.50	0.71	0.78
PFE	-0.18	0.07	-0.01	0.57	0.65	7.90	-1.00	27.45	19.38	0.47	0.53
PG	-0.33	0.06	-0.01	0.64	0.54	4.32	-1.67	28.94	15.88	0.50	0.63
SBC	-0.12	0.04	-0.01	0.61	0.54	4.45	-1.42	26.53	14.17	0.46	0.59
UTX	-0.23	0.06	-0.01	0.67	0.50	4.04	-1.23	33.25	15.28	0.54	0.66
VZ	-0.23	0.06	-0.01	0.70	0.45	4.93	-2.09	27.40	7.19	0.58	0.72
WMT	-0.15	0.08	-0.02	0.56	0.69	9.50	-2.83	26.01	18.06	0.45	0.55
XOM	-0.28	0.08	-0.02	0.74	0.59	5.79	-2.08	36.03	9.86	0.65	0.77
Mean	**-0.22**	**0.07**	**-0.01**	**0.61**	**0.63**	**7.53**	**-2.50**	**29.97**	**17.62**	**0.51**	**0.60**

Table 5. Realized Volatility Regression (AR1 Model) with Unexpected Number of Trades

Results of the following regression for log of realized variance for the Dow Jones 30 firms: $\log\sigma_t^2 = \boldsymbol{\omega}'\mathbf{d_t} + \alpha\,|z_{t-1}| + \delta z_{t-1} + \beta\log\sigma_{t-1}^2 + \lambda(v_t - \beta v_{t-1}) + \varepsilon_t$ The sample period is from January 1, 1993 to June 30, 2004. σ_t^2 is realized variance for day t , $\boldsymbol{\omega}'\mathbf{d_t}$ is the unconditional mean parameter, $\mathbf{d_t}$ is a vector of calendar (Monday and holiday) dummies, z_t are unexpected returns computed using an MA(1) model for expected returns, and v_t is unexpected log volume. In this table, volume is defined as **number of trades**. Expected log volume is obtained by fitting an AR15 model to log number of trades. The regression is estimated using OLS, with Newey-West t-statistics reported in parentheses. The column labeled ρ_{RV} shows the first-order autocorrelation of the fitted log realized variance.

	Parameter Estimates					**T-statistics**				**Adj. R^2**	ρ_{RV}
Firm	ω_t	α	δ	β	λ	$t(\alpha)$	$t(\delta)$	$t(\beta)$	$t(\gamma)$		
AA	-0.17	0.06	-0.01	0.60	0.51	7.31	-2.38	29.08	29.27	0.51	0.50
AIG	-0.29	0.07	-0.01	0.62	0.55	8.96	-1.95	31.07	25.80	0.55	0.55
AXP	-0.28	0.09	-0.02	0.60	0.55	11.09	-3.66	25.25	23.90	0.55	0.58
BA	-0.24	0.09	-0.01	0.57	0.50	9.92	-1.13	28.61	23.65	0.50	0.55
C	-0.28	0.08	-0.01	0.63	0.56	9.36	-1.37	33.03	22.63	0.56	0.62
CAT	-0.28	0.08	-0.02	0.62	0.54	9.51	-2.87	31.28	27.45	0.54	0.54
DD	-0.27	0.07	-0.01	0.67	0.57	7.78	-1.50	40.58	22.56	0.57	0.62
DIS	-0.11	0.07	0.00	0.60	0.49	9.75	-1.01	31.87	16.98	0.49	0.59
GE	-0.46	0.12	-0.02	0.63	0.58	11.74	-3.60	39.57	23.55	0.58	0.65
GM	-0.26	0.08	-0.01	0.53	0.47	9.38	-2.62	25.51	24.19	0.47	0.44
HD	-0.29	0.07	-0.01	0.60	0.53	6.65	-2.50	26.99	21.47	0.53	0.59
HON	-0.16	0.05	-0.01	0.56	0.48	4.23	-1.04	24.10	24.72	0.48	0.47
HPQ	-0.11	0.05	-0.01	0.74	0.64	4.03	-1.64	21.69	8.22	0.64	0.75
IBM	-0.32	0.09	-0.02	0.65	0.60	11.80	-3.69	35.78	28.52	0.60	0.61
INTC	-0.32	0.07	-0.01	0.79	0.75	13.19	-3.04	59.33	23.00	0.75	0.81
JNJ	-0.32	0.11	-0.03	0.54	0.44	10.48	-3.65	24.26	19.62	0.44	0.50
JPM	-0.41	0.08	-0.01	0.74	0.72	10.45	-1.98	56.22	29.50	0.72	0.72
KO	-0.38	0.09	-0.02	0.63	0.54	9.30	-3.84	31.93	21.94	0.54	0.63
MCD	-0.20	0.08	-0.02	0.51	0.43	9.00	-3.16	24.39	22.86	0.43	0.43
MMM	-0.31	0.08	-0.01	0.61	0.51	7.01	-1.89	31.97	21.91	0.51	0.51
MO	-0.17	0.08	0.00	0.51	0.43	6.20	0.18	23.71	15.96	0.43	0.46
MRK	-0.38	0.08	-0.01	0.51	0.41	7.95	-1.51	21.82	19.66	0.41	0.46
MSFT	-0.37	0.10	-0.01	0.78	0.75	15.73	-3.49	58.93	25.20	0.75	0.78
PFE	-0.22	0.07	0.00	0.60	0.51	8.92	-0.88	29.12	21.11	0.51	0.53
PG	-0.41	0.06	-0.01	0.66	0.54	5.91	-1.20	32.65	21.26	0.54	0.62
SBC	-0.24	0.05	0.00	0.63	0.51	6.19	-0.41	28.54	18.97	0.51	0.59
UTX	-0.30	0.05	-0.01	0.70	0.61	4.16	-0.92	37.01	25.24	0.61	0.64
VZ	-0.31	0.06	-0.01	0.72	0.62	5.94	-2.04	28.55	10.78	0.62	0.73
WMT	-0.21	0.09	-0.02	0.56	0.46	10.82	-3.05	26.00	19.76	0.46	0.56
XOM	-0.40	0.09	-0.01	0.76	0.69	7.05	-0.98	40.38	13.80	0.69	0.77
Mean	**-0.28**	**0.08**	**-0.01**	**0.63**	**0.55**	**8.66**	**-2.09**	**32.64**	**21.78**	**0.55**	**0.59**

Table 6. Realized Volatility Regression (AR1 Model) with Expected and Unexpected Volume

Results of the following regression for log of realized variance for the Dow Jones 30 firms: $\log\sigma_t^2 = \boldsymbol{\omega}'\mathbf{d}_t + \alpha\,|z_{t-1}| + \delta z_{t-1} + \beta\log\sigma_{t-1}^2 + \gamma(EV_t - \beta EV_{t-1}) + \lambda(v_t - \beta v_{t-1}) + \varepsilon_t$. The sample period is from January 1, 1993 to June 30, 2004. σ_t^2 is realized variance for day t , $\boldsymbol{\omega}'\mathbf{d}_t$ is the unconditional mean parameter, $\mathbf{d}_t$ is a vector of calendar (Monday and holiday) dummies, z_t are unexpected returns computed using an MA(1) model for expected returns, EV_t and v_t denotes expected and unexpected log volume respectively, where volume is defined as daily number of trades. Expected log volume is obtained by fitting an AR15 model to log number of trades. The regression is estimated using OLS, with Newey-West t-statistics reported in parentheses. The column labeled ρ_{RV} shows the first-order autocorrelation of the fitted log realized variance.

	Parameter Estimates						T-statistics					Adj. R^2	ρ_{RV}
Firm	ω_t	α	δ	β	γ	λ	$t(\alpha)$	$t(\delta)$	$t(\beta)$	$t(\gamma)$	$t(\lambda)$		
AA	-0.96	0.04	-0.01	0.53	0.30	1.26	5.51	-2.55	23.67	10.44	30.68	0.54	0.57
AIG	-0.70	0.06	-0.01	0.60	0.17	1.47	7.45	-1.76	27.49	6.00	26.33	0.56	0.60
AXP	-0.51	0.09	-0.02	0.61	0.09	1.29	9.65	-3.66	24.12	2.10	24.41	0.55	0.63
BA	-1.57	0.06	-0.01	0.51	0.40	1.16	6.62	-1.47	21.53	12.42	24.99	0.53	0.59
C	-0.81	0.07	-0.01	0.63	0.20	1.03	8.51	-1.40	27.53	5.67	23.74	0.57	0.64
CAT	-0.99	0.07	-0.02	0.59	0.28	1.25	7.63	-3.00	26.02	7.15	28.64	0.55	0.59
DD	-0.99	0.06	-0.01	0.65	0.32	1.28	6.00	-1.37	31.46	7.52	23.69	0.59	0.66
DIS	-1.25	0.05	-0.01	0.56	0.37	1.05	5.76	-1.13	25.53	9.25	18.08	0.51	0.61
GE	-1.40	0.09	-0.02	0.58	0.30	1.19	8.66	-3.27	26.09	9.30	25.14	0.60	0.66
GM	-0.79	0.07	-0.01	0.53	0.16	1.27	8.01	-2.68	25.75	3.20	24.39	0.47	0.52
HD	-1.16	0.05	-0.01	0.56	0.29	1.06	4.79	-2.43	23.89	8.64	22.79	0.54	0.61
HON	-0.84	0.04	-0.01	0.50	0.25	1.30	2.95	-1.11	18.85	10.26	26.18	0.51	0.55
HPQ	-0.37	0.05	-0.01	0.75	0.13	1.08	3.71	-1.63	20.75	0.46	7.78	0.64	0.75
IBM	-0.78	0.08	-0.02	0.66	0.18	1.12	9.63	-3.52	34.58	4.06	29.34	0.61	0.64
INTC	-1.47	0.05	-0.01	0.70	0.47	0.95	10.21	-2.95	33.91	16.17	25.84	0.78	0.80
JNJ	-0.33	0.11	-0.03	0.54	0.00	1.16	10.20	-3.64	24.38	0.10	19.60	0.44	0.54
JPM	-0.92	0.06	-0.01	0.71	0.29	1.32	8.36	-2.09	36.49	8.43	31.08	0.73	0.74
KO	-1.24	0.08	-0.02	0.62	0.33	1.04	7.05	-3.59	29.19	6.77	23.78	0.55	0.65
MCD	-1.44	0.06	-0.02	0.46	0.34	1.36	6.59	-2.94	20.91	9.33	24.12	0.45	0.50
MMM	-0.62	0.08	-0.01	0.61	0.13	1.47	6.31	-1.97	31.02	3.01	21.93	0.52	0.58
MO	-0.76	0.06	0.00	0.52	0.17	1.09	4.57	0.32	23.40	3.34	17.07	0.44	0.53
MRK	-1.08	0.07	-0.01	0.52	0.20	1.18	6.39	-1.28	21.70	2.70	20.21	0.41	0.51
MSFT	-1.33	0.09	-0.01	0.69	0.38	1.03	13.75	-3.70	33.42	16.86	27.56	0.77	0.78
PFE	-0.66	0.06	0.00	0.58	0.14	1.16	7.69	-0.78	25.77	6.21	21.54	0.52	0.58
PG	-0.43	0.06	-0.01	0.66	0.01	1.08	5.68	-1.20	33.18	0.26	21.04	0.54	0.65
SBC	-1.14	0.04	0.00	0.58	0.33	1.38	4.31	-0.44	24.79	8.58	19.95	0.53	0.63
UTX	-0.85	0.04	-0.01	0.67	0.30	1.29	3.40	-1.00	29.61	8.69	26.64	0.63	0.68
VZ	-0.72	0.06	-0.01	0.73	0.20	1.17	5.17	-2.01	28.46	1.32	10.94	0.62	0.72
WMT	-0.24	0.09	-0.02	0.56	0.01	1.01	9.90	-3.04	26.29	0.23	19.40	0.46	0.57
XOM	-0.33	0.10	-0.01	0.75	-0.04	1.24	7.05	-1.00	38.95	-0.36	13.84	0.69	0.77
Mean	**-0.89**	**0.07**	**-0.01**	**0.60**	**0.22**	**1.19**	**7.05**	**-2.08**	**27.29**	**6.27**	**22.69**	**0.56**	**0.63**

Table 7. Realized Volatility Regression (AR15 Model)

Results of the following volatility regression for the Dow Jones 30 firms: $\log\sigma_t^2 = \boldsymbol{\omega}'\mathbf{d_t} + \alpha\,|z_{t-1}| + \delta z_{t-1} + \sum_{i=1}^{15}\beta_i \log\sigma_{t-i}^2 + \varepsilon_t$. The sample period is from January 1, 1993 to June 30, 2004. σ_t^2 is realized variance for day t, $\boldsymbol{\omega}'\mathbf{d_t}$ is the unconditional mean parameter, $\mathbf{d_t}$ is a vector of calendar (Monday and holiday) dummies, and z_t are unexpected returns computed using an MA(1) model for expected returns. The regression is estimated using OLS, with Newey-West t-statistics reported in parentheses. The column labeled ρ_{RV} shows the first-order autocorrelation of the fitted log realized variance.

	Parameter Estimates				**T-statistics**			**Adj. R^2**	**DW**	ρ_{RV}
Firm	ω_t	α	δ	$\sum\beta$	$t(\alpha)$	$t(\delta)$	$t(\beta)$			
AA	0.05	0.04	-0.02	0.86	4.35	-4.55	43.18	0.46	2.01	0.93
AIG	0.02	0.05	-0.04	0.86	5.26	-6.40	42.83	0.50	2.01	0.93
AXP	0.03	0.06	-0.04	0.86	6.79	-6.92	42.89	0.51	2.02	0.92
BA	0.03	0.06	-0.02	0.84	5.57	-3.54	41.52	0.45	2.02	0.90
C	0.05	0.05	-0.03	0.86	6.39	-6.09	42.75	0.54	2.02	0.91
CAT	0.02	0.05	-0.02	0.87	5.69	-3.40	42.70	0.49	2.00	0.92
DD	0.03	0.04	-0.03	0.90	4.09	-4.24	45.16	0.55	2.01	0.94
DIS	0.05	0.04	-0.01	0.87	4.85	-2.66	43.53	0.49	2.01	0.93
GE	0.02	0.06	-0.05	0.87	6.32	-9.00	44.56	0.57	2.04	0.91
GM	0.03	0.06	-0.02	0.80	6.32	-3.67	40.67	0.38	2.00	0.88
HD	0.09	0.03	-0.03	0.86	2.97	-5.56	41.86	0.49	2.01	0.91
HON	0.07	0.04	-0.02	0.85	4.19	-3.01	41.30	0.42	2.01	0.91
HPQ	0.07	0.00	-0.03	0.95	0.05	-3.01	20.64	0.67	1.99	0.96
IBM	0.02	0.05	-0.03	0.85	5.23	-4.92	41.62	0.52	2.01	0.90
INTC	0.01	0.03	-0.02	0.93	5.91	-6.21	44.67	0.73	2.00	0.96
JNJ	-0.03	0.09	-0.04	0.81	6.67	-5.47	41.42	0.42	2.01	0.88
JPM	0.02	0.06	-0.03	0.89	6.29	-7.06	45.64	0.66	2.03	0.94
KO	0.00	0.06	-0.04	0.86	5.79	-6.30	43.06	0.53	2.02	0.92
MCD	0.04	0.05	-0.02	0.82	5.03	-3.15	41.46	0.36	2.02	0.89
MMM	-0.03	0.07	-0.04	0.86	6.00	-4.85	42.15	0.47	2.01	0.93
MO	0.05	0.05	-0.02	0.81	5.91	-2.85	42.24	0.38	2.01	0.90
MRK	0.03	0.06	-0.03	0.82	5.39	-4.05	40.79	0.38	2.01	0.89
MSFT	-0.02	0.05	-0.03	0.91	6.55	-6.30	43.33	0.69	2.01	0.94
PFE	0.05	0.04	-0.02	0.84	4.97	-3.80	41.28	0.45	2.01	0.90
PG	0.03	0.03	-0.03	0.89	2.40	-4.55	44.35	0.54	2.00	0.93
SBC	0.05	0.03	-0.02	0.90	2.91	-3.90	46.29	0.53	2.00	0.95
UTX	0.01	0.04	-0.03	0.91	3.63	-4.94	43.31	0.60	2.01	0.95
VZ	0.05	0.03	-0.03	0.92	2.15	-3.83	29.48	0.65	2.03	0.94
WMT	0.04	0.06	-0.04	0.85	6.42	-6.40	41.77	0.47	2.01	0.92
XOM	0.01	0.04	-0.04	0.92	3.36	-4.69	29.91	0.70	2.04	0.95
Mean	**0.03**	**0.05**	**-0.03**	**0.87**	**4.92**	**-4.84**	**41.21**	**0.52**	**2.01**	**0.92**

Table 8. Realized Volatility Regression (AR15) with Expected and Unexpected Volume

Results of the following regression for log of realized variance for the Dow Jones 30 firms: $\log\sigma_t^2 = \boldsymbol{\omega}'\mathbf{d}_\mathbf{t} + \alpha|z_{t-1}| + \delta z_{t-1} + \sum_{i=1}^{15}\beta_i \log\sigma_{t-i}^2 + \gamma(EV_t - \sum_{i=1}^{15}\beta_i EV_{t-i}) + \lambda(v_t - \sum_{i=1}^{15}\beta_i v_{t-i}) + \varepsilon_t$. The sample period is from January 1, 1993 to June 30, 2004. σ_t^2 is realized variance for day t, $\boldsymbol{\omega}'\mathbf{d}_\mathbf{t}$ is the unconditional mean parameter, $\mathbf{d}_\mathbf{t}$ is a vector of calendar (Monday and holiday) dummies, z_t are unexpected returns computed using an MA(1) model for expected returns, EV_t and v_t denotes expected and unexpected log volume respectively, where volume is defined as daily number of trades. Expected log volume is obtained by fitting an AR15 model to log number of trades. The regression is estimated using OLS, with Newey-West t-statistics reported in parentheses. The column labeled ρ_{RV} shows the first-order autocorrelation of the fitted log realized variance.

	Parameter Estimates						**T-statistics**							
Firm	ω_t	α	δ	$\sum\beta$	γ	λ	$t(\alpha)$	$t(\delta)$	$t(\beta)$	$t(\gamma)$	$t(\lambda)$	**Adj. R^2**	**DW**	ρ_{RV}
AA	-0.17	-0.01	-0.02	0.879	0.97	1.36	-0.80	-4.84	3.31	10.99	31.48	0.62	2.015	0.80
AIG	-0.05	0.00	-0.02	0.895	1.43	1.63	-0.61	-3.35	3.29	13.91	28.91	0.66	2.008	0.81
AXP	-0.15	0.01	-0.03	0.888	1.04	1.48	1.67	-7.07	3.35	12.26	24.58	0.66	2.027	0.82
BA	-0.55	0.03	-0.02	0.841	0.66	1.21	3.70	-3.00	2.91	8.27	24.75	0.60	2.010	0.80
C	-0.25	0.02	-0.03	0.906	0.71	1.13	3.57	-6.12	3.23	11.09	24.29	0.66	2.035	0.84
CAT	-0.27	0.02	-0.02	0.899	0.79	1.33	3.08	-3.75	3.17	9.76	30.37	0.65	2.022	0.80
DD	-0.18	0.00	-0.01	0.918	1.02	1.37	-0.09	-2.56	3.28	11.82	24.41	0.67	2.012	0.85
DIS	-0.36	0.00	-0.01	0.890	0.92	1.14	-0.33	-2.22	3.18	11.04	19.00	0.60	2.014	0.86
GE	-0.16	0.01	-0.02	0.899	1.10	1.30	0.63	-4.59	3.35	12.02	24.25	0.68	2.026	0.86
GM	-0.42	0.02	-0.02	0.845	0.76	1.37	2.73	-3.63	3.15	9.10	26.13	0.55	2.018	0.73
HD	-0.29	0.00	-0.02	0.897	0.68	1.10	0.21	-3.91	3.10	10.33	23.50	0.62	2.020	0.82
HON	-0.09	-0.01	-0.01	0.823	1.00	1.47	-1.30	-2.05	3.31	10.98	28.73	0.60	2.014	0.77
HPQ	-0.16	0.00	-0.02	0.965	0.94	1.13	-0.18	-2.55	1.42	3.09	8.11	0.74	1.988	0.92
IBM	-0.25	0.02	-0.02	0.918	0.84	1.22	2.90	-4.52	3.19	11.53	29.92	0.69	2.019	0.81
INTC	-0.30	0.01	-0.01	0.958	0.79	0.98	2.79	-4.62	3.13	13.30	24.78	0.82	2.006	0.91
JNJ	-0.22	0.02	-0.03	0.881	1.19	1.34	1.96	-3.64	3.31	13.61	21.56	0.54	2.014	0.81
JPM	-0.04	0.00	-0.02	0.951	1.22	1.47	0.79	-5.05	3.39	17.48	37.84	0.80	2.034	0.87
KO	-0.40	0.02	-0.02	0.886	0.83	1.10	2.32	-4.01	3.12	11.37	22.49	0.63	2.017	0.86
MCD	-0.54	0.02	-0.02	0.838	0.70	1.41	2.75	-2.84	3.04	8.33	24.25	0.52	2.017	0.75
MMM	-0.10	0.00	-0.03	0.897	1.27	1.64	0.00	-4.20	3.28	11.67	25.13	0.62	2.013	0.80
MO	-0.30	0.00	-0.01	0.849	0.97	1.26	-0.39	-1.07	3.25	15.05	22.94	0.56	2.011	0.79
MRK	-0.48	0.01	-0.01	0.877	0.89	1.28	0.88	-1.82	3.08	8.45	21.75	0.51	2.007	0.78
MSFT	-0.20	0.02	-0.02	0.932	0.95	1.10	3.82	-6.42	3.18	17.43	28.50	0.81	2.021	0.89
PFE	-0.09	0.00	-0.01	0.900	0.96	1.30	0.44	-2.90	3.27	10.99	24.17	0.60	2.014	0.80
PG	-0.06	-0.02	-0.02	0.938	1.06	1.25	-1.46	-2.76	3.25	9.82	19.06	0.65	2.006	0.85
SBC	-0.25	0.01	-0.01	0.892	0.72	1.42	0.77	-2.49	3.20	6.73	19.67	0.62	2.013	0.86
UTX	-0.02	-0.01	-0.02	0.923	1.47	1.51	-0.92	-4.77	3.35	13.72	31.81	0.73	2.018	0.86
VZ	-0.25	0.01	-0.02	0.942	1.24	1.31	0.51	-3.90	1.98	6.79	12.69	0.72	2.032	0.91
WMT	-0.27	0.04	-0.03	0.879	0.39	1.02	4.81	-5.06	2.93	5.34	19.93	0.55	2.021	0.83
XOM	-0.27	0.03	-0.02	0.943	1.20	1.31	2.62	-3.00	2.25	6.87	13.98	0.77	2.042	0.92
Mean	**-0.24**	**0.01**	**-0.02**	**0.90**	**0.96**	**1.30**	**1.23**	**-3.76**	**3.08**	**10.77**	**23.97**	**0.65**	**2.02**	**0.83**

Table 9. Realized Variance Decomposition

Results of a variance decomposition of log realized volatility for the Dow Jones 30. The sample period is January 1, 1993 to June 30, 2004. The decomposition is based OLS estimates of the AR(15) model with expected and unexpected log volume (see Table 8). Predictable components include (a) calendar dummies (Monday and holidays), (b) expected log volume and (c) ARCH components comprising fifteen lags of log realized volatility and previous day return shocks. The random component comprises unexpected log volume and volatility innovations (regression residuals). Volume is defined as daily number of trades. Expected log volume is estimated using an AR15 model with Monday and holiday dummies.

	Predictable Components				Random Components		
Firm	**Calendar Dummies**	**Expected Volume**	**ARCH Components**	**Total Predictable**	**Unexpected Volume**	**Volatility Innovations**	**Total Random**
AA	0.001	0.019	0.350	0.370	0.123	0.243	0.366
AIG	0.000	0.025	0.347	0.372	0.111	0.208	0.319
AXP	0.001	0.024	0.437	0.461	0.126	0.257	0.383
BA	0.002	0.017	0.299	0.318	0.106	0.267	0.373
C	0.004	0.017	0.407	0.428	0.095	0.228	0.323
CAT	0.002	0.015	0.354	0.371	0.115	0.221	0.335
DD	0.002	0.016	0.387	0.406	0.087	0.209	0.297
DIS	0.002	0.024	0.341	0.366	0.082	0.258	0.340
GE	0.005	0.025	0.396	0.426	0.084	0.210	0.295
GM	0.003	0.013	0.235	0.252	0.108	0.239	0.347
HD	0.005	0.018	0.321	0.344	0.090	0.235	0.326
HON	0.001	0.028	0.347	0.376	0.152	0.290	0.442
HPQ	0.003	0.013	0.562	0.578	0.053	0.206	0.259
IBM	0.003	0.023	0.357	0.383	0.121	0.197	0.318
INTC	0.001	0.028	0.688	0.717	0.100	0.180	0.279
JNJ	0.001	0.024	0.289	0.314	0.087	0.267	0.354
JPM	0.002	0.027	0.560	0.589	0.121	0.154	0.274
KO	0.003	0.017	0.298	0.318	0.062	0.199	0.261
MCD	0.002	0.013	0.219	0.234	0.101	0.273	0.374
MMM	0.001	0.023	0.379	0.404	0.127	0.262	0.389
MO	0.001	0.042	0.277	0.321	0.127	0.290	0.418
MRK	0.005	0.020	0.240	0.264	0.092	0.289	0.381
MSFT	0.001	0.037	0.657	0.695	0.127	0.168	0.294
PFE	0.002	0.020	0.286	0.308	0.097	0.210	0.307
PG	0.002	0.022	0.384	0.409	0.079	0.224	0.303
SBC	0.002	0.009	0.370	0.381	0.070	0.252	0.323
UTX	0.001	0.032	0.510	0.543	0.117	0.209	0.326
VZ	0.006	0.017	0.413	0.436	0.051	0.174	0.225
WMT	0.004	0.005	0.318	0.327	0.064	0.281	0.345
XOM	0.004	0.011	0.469	0.484	0.045	0.145	0.190
Mean	**0.002**	**0.021**	**0.383**	**0.406**	**0.097**	**0.228**	**0.326**

Figure 1. Daily Unexpected Volume (Log of Number of Trades) for IBM Stock: 1993-2004

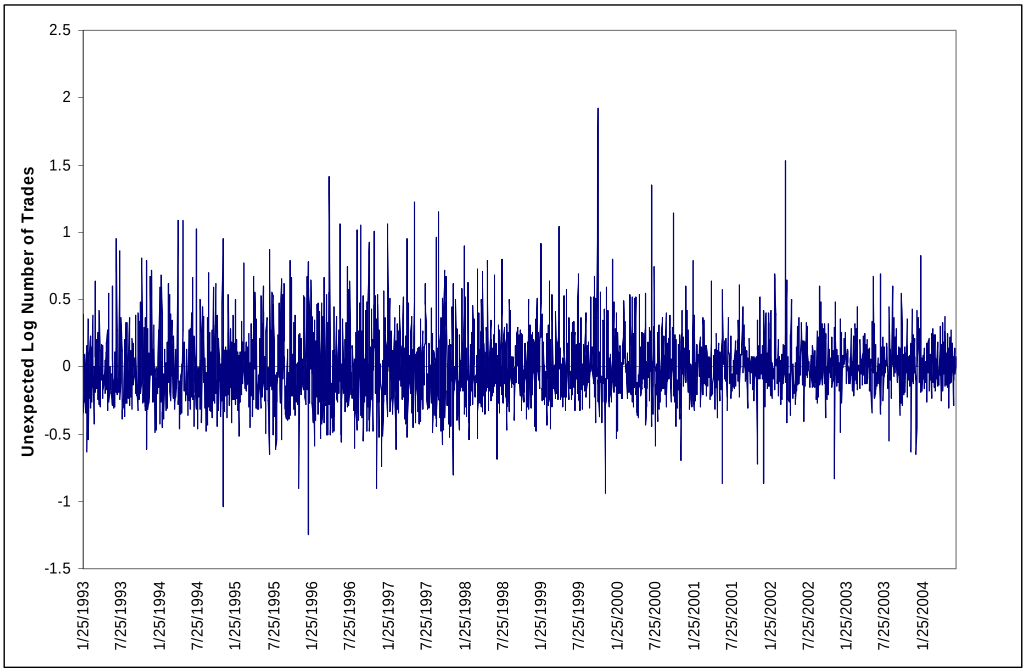


Figure 2. Percentage Contribution of Expected Volume, Unexpected volume and Volatility innovations to the Variance of Realized Volatility

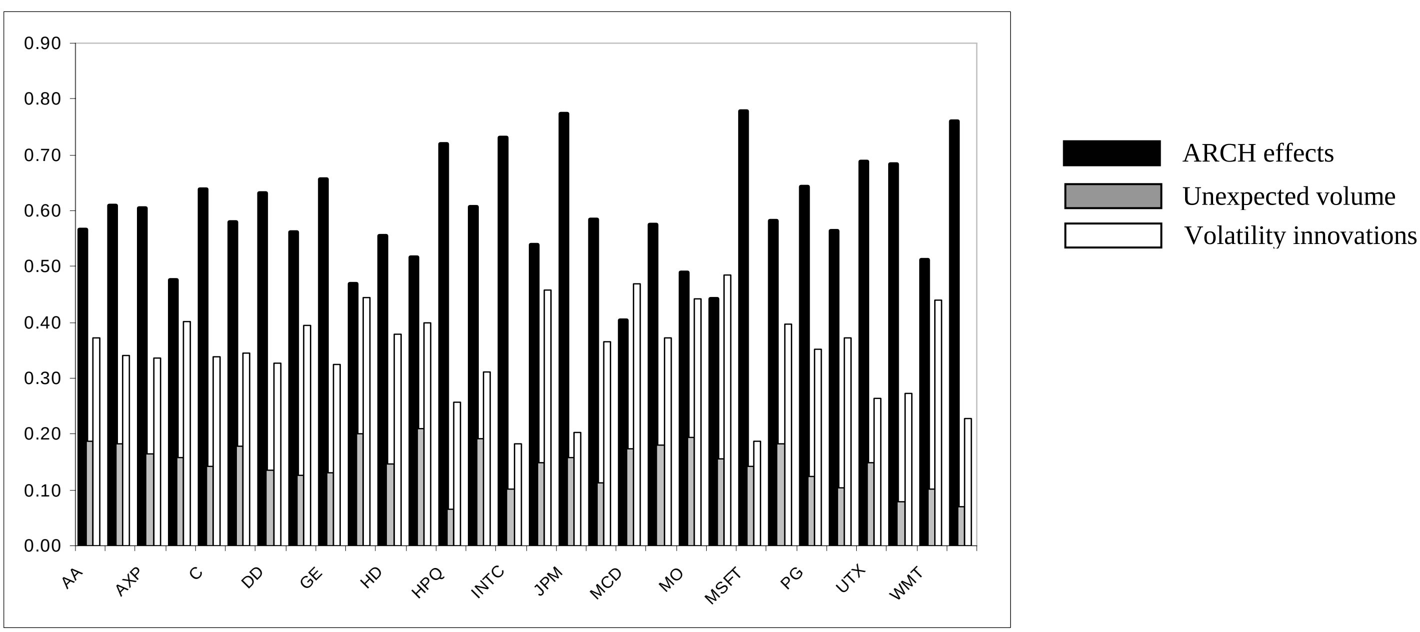
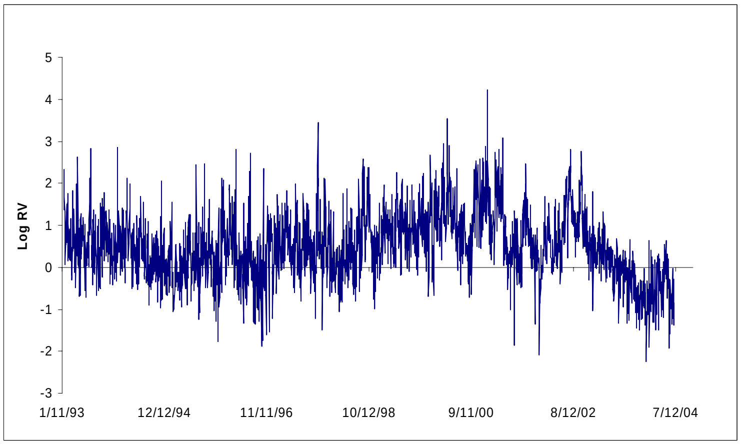


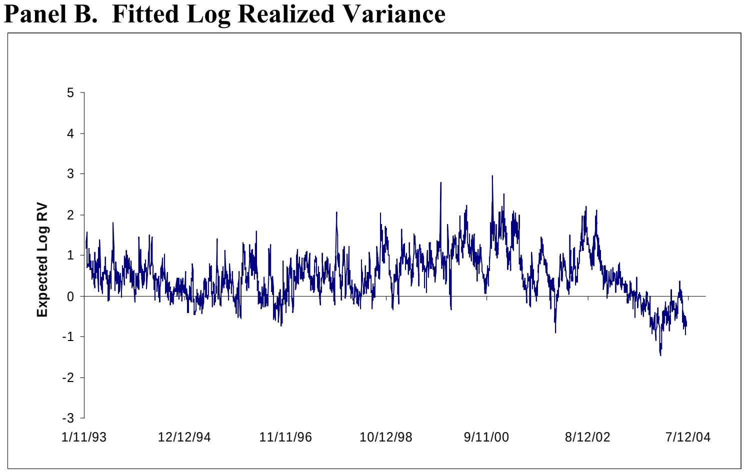
Figure 3. Realized Volatility Components: IBM Stock

Sample period is from January 1, 1993 to June 30, 2004. Panel A shows the daily log realized volatility of IBM stock. Panel B is the fitted log realized volatility based on regression estimates of an AR(15) model for log realized volatility. Panel C plots the time series of the regression residuals.

Panel A. Daily log Realized Variance



Panel B. Fitted Log Realized Variance



Panel C. Regression Residuals