ARCH Effects and Trading Volume ⋆

Jeff Fleming [a], Chris Kirby [b], Barbara Ostdiek [a]

[a] *Jesse H. Jones Graduate School of Management, Rice University, Houston, TX 77005*

[b] *John E. Walker Department of Economics, Clemson, SC 29631*

Abstract

Studies that fit volume-augmented GARCH models often find support for the hypothesis that trading volume explains ARCH effects in daily stock returns. We show that this finding is due to an unrecognized constraint imposed by the GARCH specification used for the analysis. Using a more flexible specification, we find no evidence that inserting volume into the conditional variance function of the model reduces the importance of lagged squared returns in capturing volatility dynamics. Volume is strongly correlated with contemporaneous return volatility, but the correlation is driven largely by transitory volatility shocks that have little to do with the highly persistent component of volatility captured by standard volatility models.

Key words: volume-volatility relation, information flow, two-component GARCH, bivariate mixture models, mixture of distributions hypothesis

⋆ Comments welcome. We thank Scott Baggett for helping us construct the price and volume datasets from TAQ data and Joel Hasbrouck and James Weston for providing useful insights and advice regarding TAQ data. Address correspondence to: Chris Kirby, John E. Walker Department of Economics, Clemson University, P.O. Box 341309, Clemson, SC 29634-1309.

Email addresses: `jfleming@rice.edu` (Jeff Fleming), `cmkirby@clemson.edu` (Chris Kirby), `ostdiek@rice.edu` (Barbara Ostdiek).

Rice University and Clemson University Working Paper *12 September 2005*

ARCH Effects and Trading Volume

ABSTRACT

Studies that fit volume-augmented GARCH models often find support for the hypothesis that trading volume explains ARCH effects in daily stock returns. We show that this finding is due to an unrecognized constraint imposed by the GARCH specification used for the analysis. Using a more flexible specification, we find no evidence that inserting volume into the conditional variance function of the model reduces the importance of lagged squared returns in capturing volatility dynamics. Volume is strongly correlated with contemporaneous return volatility, but the correlation is driven largely by transitory volatility shocks that have little to do with the highly persistent component of volatility captured by standard volatility models.

ARCH Effects and Trading Volume

This paper investigates the degree to which autoregressive conditional heteroscedasticity (ARCH) in stock returns is explained by the dynamics of trading volume. Our investigation is based on a new volume-augmented generalized ARCH (VA-GARCH) model in which return volatility is allowed to contain both short- and long-term components. In contrast to prior studies, we find no support for the hypothesis that inserting volume into the conditional variance function of the model reduces the importance of lagged squared returns in capturing volatility dynamics. Previous conclusions that trading volume largely explains ARCH effects appear to stem from an unrecognized constraint imposed by the econometric methodology.

Like most studies in the area, we use the mixture of distributions hypothesis (MDH) to guide our analysis. In one of the first studies to use VA-GARCH models in an MDH context, Lamoureux and Lastrapes (1990) report that ARCH effects tend to disappear when contemporaneous trading volume is added to the conditional variance function of a GARCH(1,1) specification. Although subsequent studies report a less dramatic attenuation of ARCH effects in such models, they generally find that incorporating trading volume produces a substantial drop in volatility persistence and at least some reduction in the significance of ARCH effects (see, e.g., Fujihara and Mougoue, 1997; Girma and Mougoue, 2002; Marsh and Wagner, 2003).

Fleming, Kirby, and Ostdiek (2006), on the other hand, use linear state-space methods to investigate the MDH. They fit a number of MDH-based specifications and find that accounting for the dynamics of trading volume has no impact on the significance of ARCH effects. Moreover, their analysis indicates that trading volume is primarily related to the nonpersistent component of return volatility. These findings, which are clearly at odds with the conclusions drawn by Lamoureux and Lastrapes (1990), call into question the robustness of the VA-GARCH methodology. Our paper attempts to identify the source of these conflicting results.

In general, there are two specification issues that may affect the results obtained using the VA-GARCH methodology. The first issue is the impact of simultaneity bias on the VA-GARCH coefficient estimates. Some researchers, such as Liesenfeld (1998), use this issue to help motivate econometric methods that are much more computationally intensive than fitting volume-augmented GARCH models. However, we show that the

MDH implies that the impact of simultaneity bias becomes negligible as the number of traders in the market and/or the number of daily information events becomes large. Hence, by focusing on widely held and frequently traded stocks that are often in the news, we can retain the tractability and intuitive appeal of the GARCH framework while minimizing concerns about the impact of simultaneity bias. Since Lamoureux and Lastrapes (1990) also examine actively traded stocks, it is unlikely that simultaneity bias plays a large role in their findings.

The GARCH(1,1) methodology used by Lamoureux and Lastrapes (1990), however, is vulnerable to a second specification problem that has escaped attention in the literature. Although it is natural to consider a GARCH(1,1) model given its success in other applications, a VA-GARCH(1,1) model imposes an important restriction that makes it difficult to interpret the model fitting results. Specifically, the coefficients on lagged volume and lagged squared returns are constrained to decline with the lag length at the same rate. Thus, if volume provides little information about future volatility, it might be necessary to downweight lagged volume (and hence lagged squared returns) to keep the fitted volatilities from becoming too noisy. This restriction may explain why ARCH effects tend to vanish when volume is added to the GARCH(1,1) model.

We overcome this problem by fitting a volume-augmented exponential GARCH (VA-EGARCH) model that allows for both short- and long-term volatility components. Our VA-EGARCH(2,2) model nests the VA-EGARCH(1,1) model as a special case, which allows us to directly assess the impact of relaxing the implicit restriction. We fit the model to daily returns on the 20 stocks in the major market index (MMI). We find no support for the hypothesis that volume explains ARCH effects. The results confirm that volume is strongly correlated with contemporaneous return volatility, but the correlation is driven by transitory shocks to the volatility process. Nothing in the model fitting results suggests that volume explains the highly persistent component of volatility that is captured by standard volatility models.

We provide further evidence on this issue by examining the relative performance of the EGARCH models in explaining realized volatility. Our approach, which follows Andersen and Bollerslev (1998), is to regress the realized variances on the fitted variances produced by each model. The regressions confirm that the VA-EGARCH(1,1) model provides misleading evidence regarding of the relation between volume dynamics and ARCH effects. The fitted variances from the VA-EGARCH(1,1) model produce a much lower

R-squared than the fitted variances from the basic EGARCH(1,1) model. In contrast, the fitted variances from the VA-EGARCH(2,2) model substantially outperform the fitted variances from both of these models as well as those from the basic EGARCH(2,2) model. Since the superior performance of the VA-EGARCH(2,2) model is primarily attributable to the undiminished role of ARCH effects, our results suggest that we need to look beyond volume in order to identify the features of the trading process that give rise to ARCH effects in daily stock returns.

The remainder of the paper is organized as follows. Section 1 covers general background issues, analyzes the large-market implications of the most widely studied bivariate mixture model, and introduces the two-component EGARCH model that we use for the empirical analysis. Section 2 describes the dataset and reports the model fitting results. Section 3 offers some concluding remarks.

1 Background and Methodology

Lamoureux and Lastrapes (1990) is arguably the most influential study of the relation between ARCH effects and trading volume. Using a sample of daily data for 20 U.S. firms, they fit a volume-augmented GARCH(1,1) model of the form

$$R_t = \mu + h_t^{1/2} z_t, \tag{1}$$

$$h_t = \omega + \beta h_{t-1} + \alpha r_{t-1}^2 + \gamma V_t, \tag{2}$$

where R_t is the daily stock return, V_t is the daily trading volume, z_t is an i.i.d. $\mathrm{N}(0,1)$ standardized innovation, and $r_t = R_t - \mu$ is the demeaned return. They find that they cannot reject $\alpha = \beta = 0$ for 16 of the firms and the parameter estimates for the remaining four firms suggest much lower levels of volatility persistence than those obtained under the constraint $\gamma = 0$. As a result, they conclude that "lagged squared residuals contribute little if any additional information about the variance of the stock return process after accounting for the rate of information flow, as measured by contemporaneous volume."

There are two aspects of the methodology that raise concerns about the robustness of the results. First, it treats volume as exogenous. This can give rise to an undetermined simultaneity bias if R_t and V_t are jointly determined. Second, it relies on a model that constrains volume effects to decay at the same rate as ARCH effects, i.e., the coefficients on V_{t-s} and r_{t-s-1}^2 are both proportional to β^s for all $s > 0$. This lack of flexibility can affect the coefficient estimates. Although simultaneity bias is generally viewed as the primary robustness issue, we argue that the lack of flexibility exhibited by the GARCH(1,1) model is likely to be a more serious concern in most applications.[1]

The criticism of simultaneity bias in the VA-GARCH methodology arises because most market microstructure models imply that trading volume is endogenous. It may be reasonable, however, to treat V_t as exogenous under conditions in which the magnitude of the resulting bias is likely to be small. The lack of flexibility in the VA-GARCH(1,1) model, on the other hand, is difficult to overcome. Suppose, for example, that V_t is strongly correlated with the volatility of R_t, but contains no information beyond that contained in R_t about the volatility of R_{t+s} for any $s > 0$. Obtaining an estimate of β close to zero might be evidence that volume largely subsumes ARCH effects, or it could simply indicate that putting large weights on lagged volume (and hence lagged squared returns) makes the fitted volatilities too noisy, thereby reducing the likelihood.

To avoid this kind of ambiguity, we propose a more flexible approach for investigating the extent to which volume explains ARCH effects. We begin by examining the issue of simultaneity bias in the context of bivariate mixture models. By analyzing the asymptotic properties of these models, we identify circumstances under which a strategy of fitting VA-GARCH models should pose minimal concerns about simultaneity bias. Once this is established, we show how to overcome the structural constraints of the VA-GARCH(1,1) model without sacrificing the tractability of the GARCH methodology.

1.1 Bivariate mixture models

Let v_t denote detrended trading volume. Much of the recent empirical work on the relation between volume and volatility, such as Liesenfeld (1998) and Watanabe (2000), is motivated in large part by Andersen's (1996) modified MDH. The modified MDH implies a bivariate mixture model of the form

$$r_t = \sigma(IJK_t)^{1/2} z_{rt}, \tag{3}$$

$$v_t = \tau(\nu + \lambda IJK_t) + \tau(\nu + \lambda IJK_t)^{1/2} z_{vt}, \tag{4}$$

[1] The results of Fleming, Kirby, and Ostdiek (2006), which are robust to simultaneity bias, indirectly support this argument.

where I is the number of informed (i.e., non-liquidity) traders in the market, J is the number of information arrivals on the day used as a benchmark, K_t is the intensity of information arrivals on day t relative to the benchmark day, z_{rt} is an i.i.d. $\mathrm{N}(0,1)$ standardized innovation, and z_{vt} is an i.i.d. standardized innovation — distributed independently of z_{rt+s} for all t and s — such that $v_t|K_t \sim \tau \cdot \mathrm{Po}(\nu + \lambda IJK_t)$.

If K_t is serially correlated, then the model in Equations (3) and (4) is capable of generating autocorrelation in return volatility that is explained to some extent by the dynamics of trading volume. Unfortunately, the fact that K_t is unobservable makes it difficult to fit the model using standard methods and tends to obscure the precise nature of the relation between volume dynamics and ARCH effects in returns. One way to bring the role of volume into sharper focus is to adopt an asymptotic perspective. In particular, Equation (4) implies that for large I and/or J,

$$\frac{v_t}{\tau\lambda IJ} \simeq K_t. \tag{5}$$

Hence, if a stock is heavily traded and/or frequently in the news, it might be reasonable to use volume as a proxy for K_t and approximate the return generating process as

$$r_t = (\gamma v_t)^{1/2} z_{rt}, \tag{6}$$

where $\gamma = \sigma^2/(\tau\lambda)$. This is basically a subordinated stochastic process model of stock returns in which trading volume functions as the directing variable (see Clark, 1973).

Although asymptotic analysis does not justify the use of VA-GARCH models in general, it does allow us to identify circumstances in which the impact of simultaneity bias is likely to be small. The model in Equation (6) is consistent with the model in Equations (1) and (2) provided that $\alpha = \beta = 0$. Thus, even if we believe that the data are generated by a bivariate mixture model, we can make a case for fitting a VA-GARCH(1,1) model as long as it is reasonable to assume that I and/or J are large. This is convenient because the computational demands of GARCH models are generally much lower than those of bivariate mixture models.[2] Of course, this argument applies only to the simultaneity bias

[2] The econometric analysis of bivariate mixture models is typically carried out using specialized and computationally-intensive Monte Carlo techniques (see, e.g., Liesenfeld, 1998; Watanabe, 2000). Fleming, Kirby, and Ostdiek (2006) show, however, that it is often possible to apply the more tractable framework of the Kalman filter. Their approach provides an alternative to the GARCH methodology developed here.

issue. It says nothing about whether a GARCH(1,1) model is well-suited to studying the role of volume in explaining ARCH effects. For robustness we need a model that allows volume effects to decay at a different rate than ARCH effects. This leads us to propose a two-component specification.

1.2 A two-component VA-EGARCH model

A natural way to generalize the Lamoureux and Lastrapes (1990) methodology is to specify a VA-GARCH model that allows for both short-term and long-term volatility components. We employ a model of the form

$$r_t = \sqrt{h_t} z_{rt} \tag{7}$$

$$\Delta \log h_t = \Delta m_t + \kappa_h (m_{t-1} - \log h_{t-1}) + \sigma_h u_{t-1} + \gamma_h w_t \tag{8}$$

$$\Delta m_t = \kappa_m (\varsigma - m_{t-1}) + \sigma_m u_{t-1} + \gamma_m w_t, \tag{9}$$

where $u_t = (|z_{rt}| - \mathrm{E}[|z_{rt}|])/\sqrt{\mathrm{var}(|z_{rt}|)}$ and $w_t = (\log v_t - \mathrm{E}[\log v_t])/\sqrt{\mathrm{var}(\log v_t)}$ denote the standardized values of $|z_{rt}|$ and $\log v_t$, respectively.[3]

To see the origins of the model, suppose $\kappa_m = \kappa_h$ and $\gamma_h = \sigma_m = \gamma_m = 0$. In this case, Equations (7) – (9) collapse to

$$r_t = \sqrt{h_t} z_{rt} \tag{10}$$

$$\Delta \log h_t = \kappa_h (\varsigma - \log h_{t-1}) + \sigma_h u_{t-1}, \tag{11}$$

which is simply an EGARCH(1,1) model expressed in a form that lends a convenient interpretation to each parameter.[4] Specifically, ς is the unconditional mean of $\log h_t$, κ_h determines the speed at which $\log h_t$ reverts towards ς, and σ_h is the volatility of the innovations to $\log h_t$. If we relax the constraint $\gamma_h = 0$, then we obtain a volume-augmented model similar to that of Lamoureux and Lastrapes (1990). The main differences are that we work in logarithms to enforce the nonnegativity of h_t and we standardize trading volume to facilitate discussion of the empirical results.

[3] We standardize $|z_{rt}|$ and $\log v_t$ to make it easier to interpret and compare the coefficient estimates. This has no effect on the dynamic implications of the model.
[4] Unlike the EGARCH specification of Nelson (1991), our model does not allow for leverage effects. This is simply for ease of exposition. Leverage effects tend to be small for individual stocks and allowing them does not have much impact on our findings.

Now consider the full model in Equations (7) – (9). Its underlying structure is still that of an EGARCH specification, but instead of reverting towards a fixed mean ς, the log variance is pulled towards a stochastic mean m_t whose dynamics are described by an autoregressive process. The idea behind this generalization, which follows Engle and Lee (1999), is that m_t captures low-frequency variations in volatility, while high-frequency variations are captured by $\log h_t - m_t$. This gives the model the flexibility to incorporate volume effects that decay at a different rate than ARCH effects. Suppose, for example, that volume has a very transitory impact on volatility, with most of the volatility persistence due to ARCH effects. We would expect to find that κ_h is large relative to κ_m, σ_h is small relative to γ_h, and σ_m is large relative to γ_m.

We gain additional insights into the dynamic properties of the model by expressing the conditional variance in a way that eliminates m_t from explicit consideration. To do this, we substitute Equation (9) into Equation (8), and then substitute for m_{t-1} in the resulting expression using the original Equation (8). After consolidating terms we obtain

$$\Delta \log h_t = \kappa_1(\varsigma - \log h_{t-1}) + \kappa_2(\varsigma - \log h_{t-2}) + \sigma_1 u_{t-1} + \sigma_2 u_{t-2} + \gamma_1 w_t + \gamma_2 w_{t-1}, \quad (12)$$

where $\kappa_1 = \kappa_h + \kappa_m - 1$, $\kappa_2 = (1 - \kappa_h)(1 - \kappa_m)$, $\sigma_1 = \sigma_h + \sigma_m$, $\sigma_2 = -(\kappa_h \sigma_m + \kappa_m \sigma_h)$, $\gamma_1 = \gamma_h + \gamma_m$, and $\gamma_2 = -(\kappa_h \gamma_m + \kappa_m \gamma_h)$. Hence, the model has a VA-EGARCH(2,2) representation that imposes a set of nonlinear coefficient restrictions.

The VA-EGARCH(2,2) representation highlights both the similarities and differences between our methodology and that of Lamoureux and Lastrapes (1990). Both approaches augment the conditional variance function of a standard GARCH process with contemporaneous trading volume. However, the model developed here is less restrictive than their VA-GARCH(1,1) model because it does not force volume effects to decay at the same rate as ARCH effects. This added flexibility could be important in other applications as well. For example, it is not uncommon for researchers to fit GARCH(1,1) models that are augmented with other explanatory variables such as bid-ask spreads, open interest, and implied volatilities (see, e.g., Day and Lewis (1992), Lamoureux and Lastrapes (1993), Fujihara and Mougoue (1997), Blair, Poon, and Taylor (2001), and Girma and Mougoue (2002)). Our analysis indicates that, to the extent that the impact of the explanatory variables decays at a different rate than ARCH effects, the model fitting results in such studies could be misleading.

1.3 Model comparisons

Ultimately we want to compare how well the different models capture the dynamics of volatility. We conduct these comparisons using realized variances. The concept of realized variance was introduced by Merton (1980). Let $R_{t_{i,m}}$, $i = 1, ..., m$ denote the intraday returns on day t over m equally-spaced intervals. The realized variance on day t is the sum of the squared returns,

$$RV_t = \sum_{i=1}^{m} R^2_{t_{i,m}}. \tag{13}$$

The realized variance should be close to the true variance provided that certain conditions are satisfied. 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 $IV_t = \int_0^1 \sigma^2_{t+\tau} d\tau$ as a measure of the daily variance. Under weak regularity conditions, $RV_t - IV_t \to 0$ almost surely as $m \to \infty$ (see ABDL (2001) and Barndorff-Nielsen and Shephard (2002) for details). This suggests that by increasing the frequency at which we sample the returns, we can construct consistent nonparametric estimates of the integrated variance that in principle are arbitrarily efficient. [5]

Andersen and Bollerslev (1998) use realized variances to assess whether standard volatility models generate accurate forecasts. Their approach consists of regressing the realized variances on the fitted variances produced by a volatility model estimated using daily returns. To apply their approach, we fit regressions of the form

$$\log RV_t = a + b \log \hat{h}_t + e_t, \tag{14}$$

where $\log \hat{h}_t$ denotes the fitted log variance for day t produced by one of our EGARCH models. Although the regression R-squared will be 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 does not affect our model comparisons because the ratio of the R-squared values produced by different models is bias free.

[5] Obviously the true price process is unobservable in practice and realized variances constructed according to Equation (13) can be biased by the influence of microstructure effects on observed prices and the absence of high-frequency returns during the nontrading periods overnight and on weekends. We discuss our approach for dealing with these issues in the next section.

2 Data and empirical findings

Our investigation of the properties of bivariate mixture models suggests that we should focus on widely held, frequently traded stocks that have high information flow. Rather than attempt to identify a suitable set of stocks through the use of some screening mechanism, we simply consider the 20 firms in the MMI.[6] This gives us a set of stocks that are likely to possess the desired characteristics without raising concerns about stock selection bias. The MMI stocks are widely held by both individual and institutional investors and generally exhibit a high level of trading activity.

The dataset consists of daily returns, trading volumes, and realized variances for the 20 MMI firms. We construct the dataset using data from two sources: intraday observations on transaction prices and trading volume from the Trade and Quote (TAQ) database of the New York Stock Exchange (NYSE), and information on daily returns, stock splits, stock dividends, and cash dividends from the Center for Research in Security Prices (CRSP) daily stock price file. Our sample period is January 4, 1993 to December 31, 2003 (2,770 observations).[7]

We delete 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). In addition, we apply two other screens designed to identify and eliminate price reporting errors. First, we exclude prices that are more than 20 percent higher or lower than the previous transaction price. Second, we flag prices that imply a price change greater than two percent in magnitude and are immediately followed by a price reversal greater than two percent in magnitude. We exclude the flagged price if the implied price change is more than two times the next largest price change for the day, or if the price falls outside the day's high-low range (ignoring the flagged price) by more than the next largest price change for the day.

We aggregate the transaction volume for all of the remaining TAQ records to construct the daily trading volume for each firm. We adjust this figure as necessary for

[6] 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).

[7] Philip Morris 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. We exclude this date from the sample.

stock splits and stock dividends using the information on stock distributions from the CRSP daily stock price file. We obtain the detrended volume series used to fit the various EGARCH models by using OLS to extract a quadratic time trend. Although other methods would provide more flexibility in fitting the trend, they would be more prone to overfitting as well, which could inadvertently remove components of volume that are important to the volume-volatility relation. In any case, we find that detrending the volume series has little impact on our results. We obtain similar results using the original series instead of the detrended series.

We construct the daily realized variances for each firm using the filtered transaction price records from the TAQ database. We use the Hansen and Lunde (2005) approach to construct the realized variance for the full day. First, we construct the trading-day realized variance using the Newey and West (1987) corrected realized variance estimator proposed by Hansen and Lunde (2004). The Newey-West approach provides an unbiased estimate of the integrated variance, even in the presence of microstructure effects for returns sampled at very high frequencies, and guarantees nonnegativity. We implement the estimator using a 30-second sampling frequency for returns and a 30-minute window length for the Newey-West correction. Next, we combine the trading-day realized variance with the squared return over the previous nontrading period. We estimate this return using the last transaction price on the previous day from the TAQ database and the first price on the current day, adjusted for cash dividends and stock distributions reported in the CRSP database. We follow Hansen and Lunde (2005) to determine the relative weights placed on the trading- and nontrading-period variances to obtain the full-day realized variance. More details regarding our construction of the realized variances are provided in the Appendix.

2.1 Estimation and inference for the EGARCH(1,1) model

We begin our empirical analysis by fitting the basic EGARCH(1,1) model in Equations (10) and (11). Specifically, we estimate the parameters via maximum likelihood and use the Bollerslev and Wooldridge (1992) approach to compute robust standard errors. Table 1 reports the parameter estimates and t-ratios along with several specification diagnostics. As expected, the model only partially accounts for the fat tails that characterize the distribution of daily returns. The excess kurtosis of the standardized returns is positive for all firms, with especially large values for Eastman Kodak, Phillip Morris, and Procter and Gamble. Finding a few large values is not unusual, however, given the extreme returns that occasionally occur for individual stocks.

The results clearly point to high levels of volatility persistence. Our estimate of κ_h is close to zero for all firms except Eastman Kodak, and only seven of the estimates have t-ratios of 2.0 or greater. In addition, the first-order sample autocorrelation of the fitted conditional volatilities is 0.96 or higher for all firms except Eastman Kodak. The diagnostics suggest that the low estimate of persistence for Eastman Kodak is probably due to a small number of influential observations. Eastman Kodak has the largest excess kurtosis of any firm, and an examination of the data reveals several instances of daily returns between 10 and 20 percent in magnitude. Overall, the model fitting results are consistent with those of previous studies in the volatility modeling literature (see, e.g., Kim and Kon (1994)).

The model itself does not appear to have much explanatory power. The R-squared value for a regression of the absolute demeaned returns on the fitted conditional volatilities ranges from 3 percent for Eastman Kodak to 15 percent for Dow Chemical.[8] Of course, as Andersen and Bollerslev (1998) point out, we expect such regressions to produce relatively low R-squared values because absolute returns are a noisy proxy for volatility. The realized variance regressions discussed later provide a better benchmark for assessing how well the EGARCH(1,1) specification captures volatility dynamics.

Next we consider a modified version of the EGARCH(1,1) specification in which contemporaneous volume is included as an explanatory variable. Table 2 reports the model fitting results. The most striking change from Table 1 is a sharp increase in the estimates of κ_h together with a sharp decline in the estimates of ρ, the first-order sample autocorrelation of the fitted conditional volatilities. The estimate of κ_h exceeds one for 19 of the 20 stocks and most of the estimates are highly statistically significant. The largest ρ estimate now is just 0.63 (Dow Chemical), while the smallest estimate is 0.25 (ChevronTexaco). Clearly, the addition of volume as an explanatory variable produces a marked drop in the degree of volatility persistence implied by the model.

More generally, our results confirm that volume is a significant factor in explaining contemporaneous volatility. In most cases, volume enters the model with a t-ratio greater

[8] We use absolute, rather than squared, demeaned returns in these regressions so the results are less sensitive to outliers. See Davidian and Carroll (1987).

than 15, the log likelihood values are substantially higher than those in Table 1, and the excess kurtosis is substantially lower as well. Incorporating volume also produces a jump in the R-squared values for a regression of the absolute demeaned returns on the fitted conditional volatilities. The majority of the R-squared values in Table 2 are greater than 20 percent (the largest is 39 percent for Procter and Gamble), while the majority of those in Table 1 are less than 10 percent. This increase in explanatory power suggests that the contemporaneous relation between volume and volatility is quite strong.

Despite these results, it would be premature to conclude that volume accounts for or subsumes ARCH effects in daily returns. The most obvious indication of this is that the absolute standardized return still enters the volume-augmented model with a positive and statistically significant coefficient for almost all of the stocks. In many cases, the t-ratio on the σ_h estimate is five or greater. The question is how to interpret this evidence given the structural constraints imposed by EGARCH(1,1) methodology. One possibility is that, even if the constraints were relaxed, ARCH effects would make only a small contribution to the explanatory power of the volume-augmented model. This would support the conclusions drawn by Lamoureux and Lastrapes (1990). Alternatively, the constraints could be masking the true contribution of ARCH effects to volatility dynamics. We turn now to a more detailed analysis of this issue.

2.2 Estimation and inference for the EGARCH(2,2) model

To investigate the effects of relaxing the constraints imposed by the EGARCH(1,1) methodology, we estimate a more flexible econometric specification that nests the EGARCH(1,1) model as a special case. In particular, we consider an EGARCH(2,2) model that allows for both short- and long-term volatility components. We initially estimate the model without incorporating volume to assess how its empirical implications differ from those of the basic EGARCH(1,1) model. Table 3 reports the results.

In general, the EGARCH(2,2) model fits better than the EGARCH(1,1) model reported on in Table 1. Most of the t-ratios for the κ_h and σ_h estimates are greater than two and the increase in the log likelihood is statistically significant at the 5 percent level for a majority of the stocks. Nonetheless, allowing for two volatility components does not have a major impact on the volatility dynamics implied by the model. Although we observe some decline in the first-order sample autocorrelation of the fitted conditional volatilities, the autocorrelation still exceeds 0.90 for 17 of the 20 stocks. Similarly, the R-squared for a regression of the absolute demeaned returns on the fitted conditional volatilities suggests little increase in the explanatory power of the model. Overall these results point to a relatively modest improvement in the goodness of fit.

Now we consider the main issue, which is how the EGARCH(2,2) model performs once we incorporate volume as an explanatory variable. Table 4 reports the results. Two aspects of the results stand out immediately. First, all of the log likelihood values are significantly higher than those for the VA-EGARCH(1,1) model reported in Table 2. The average increase in log likelihood across stocks is 103. Second, all of the R-squared values are substantially higher as well. Most of the increases are in the range of 6 to 10 percentage points, with increases of 13 percentage points for three of the firms. These findings point to a clear increase in explanatory power relative to the VA-EGARCH(1,1) model.

We also see an interesting pattern in the coefficient estimates. All of the κ_h and γ_h estimates are positive, highly statistically significant, and comparable in magnitude to the corresponding estimates in Table 2. More importantly, only four of the σ_h estimates are significantly different from zero at the 5 percent level. This indicates that the short-term dynamics of log volatility, which are captured by $\log h_t - m_t$, are explained almost exclusively by volume. In contrast, only one of the γ_m estimates is statistically significant at the 5 percent level, while the ς, κ_m and σ_m estimates look similar to the corresponding estimates in Table 1. This indicates that the long-term dynamics of log volatility, which are captured by m_t, are explained almost exclusively by the absolute standardized returns.

These findings suggest a much different role for volume than the results obtained using the VA-EGARCH(1,1) model. Specifically, nothing in the results obtained using the VA-GARCH(2,2) model indicates that volume accounts for or subsumes ARCH effects in daily returns. On the contrary, we find that ARCH effects are a key determinant of long-term volatility dynamics and that the long-term component of volatility displays the high level of persistence typically reported in the ARCH literature. Moreover, the evidence suggests that the long-term component of volatility in the VA-EGARCH(2,2) model behaves similar to the conditional volatility implied by the basic EGARCH(1,1) model.

Figure 1 illustrates this point more clearly. The figure compares the fitted values produced by the two models for American Express, the first stock alphabetically in the MMI. Panels A and B plot the fitted values of $\log h_t$ and m_t, respectively, for the VA-

EGARCH(2,2) model. Panel C plots the fitted values of $\log h_t$ for the basic EGARCH(1,1) model. The fitted log volatility in Panel A is highly variable, reflecting the strong short-term impact of trading volume. But there are also indications of an underlying autoregressive structure that seems characteristic of a slowly mean-reverting process. Once we isolate the associated component of log volatility (Panel B), we find that it tracks closely with the fitted values from the basic EGARCH(1,1) model (Panel C).

Although this evidence is clearly at odds with the main conclusion of Lamoureux and Lastrapes (1990), there is another aspect of the relation between volume and ARCH effects on which we agree. Regardless of the model, incorporating volume produces a large drop in the persistence of the fitted volatility series. Consider the ρ estimates in Table 4. Although they are higher than the corresponding estimates in Table 2, they are still well below the values reported in Tables 1 and 3. Since ARCH effects appear to be undiminished for the model analyzed in Table 4, this has two implications. First, the short-term component of volatility is much less persistent than is typical of the fitted volatility from ARCH models. Second, short-term dynamics account for a substantial fraction of the total variation in the volatility of daily returns.

Table 5 provides additional evidence on the short- and long-term volatility dynamics. The first four columns report the sample variance of the fitted $\log h_t$ for each of the models in Tables 1 – 4. Not surprisingly, both of the volume-augmented models imply substantially more variation in $\log h_t$. The more interesting comparison is between these two models. The VA-EGARCH(2,2) model yields the higher value for every stock, and the difference is often 20 percent or more. This is indicative of the impact of relaxing the structural constraints imposed by the VA-EGARCH(1,1) model. In the absence of these constraints, ARCH effects make an important contribution to the dynamics of the log variance series.

The three remaining columns of the table decompose the sample variance of the fitted $\log h_t$ for the VA-EGARCH(2,2) model into three components — short-term, long-term, and interaction — using the relation $\mathrm{var}(\log h_t) = \mathrm{var}(\log h_t - m_t) + \mathrm{var}(m_t) + 2\,\mathrm{cov}(\log h_t - m_t, m_t)$. The results show that most of the variation in $\log h_t$ is short-term in nature. But this is not due to the absence of strong ARCH effects. Compare the variance of m_t in column six to the variance of the fitted $\log h_t$ from the EGARCH(1,1) model in column one. The two sets of figures are relatively similar, which is consistent with the evidence from Figure 1. In general, the long-term component of volatility tends to closely mimic the conditional volatility implied by the basic EGARCH(1,1) model.

The relation between the short- and long-term volatility components is also of interest. The interaction term in the variance decomposition is negative for most of the firms. However, with the exception of AT&T, Eastman Kodak, and General Motors, the correlation between the components is such that a regression of one on the other would yield an R-squared of less than 5 percent. Therefore, it seems that the short- and long-term components of volatility are largely unrelated. Since the former is driven primarily by contemporaneous volume and the latter by lagged absolute returns, this lack of correlation is broadly consistent with volatility following a stochastic autoregressive process in which the unpredictable volatility shocks are strongly associated with the contemporaneous level of trading activity.

2.3 Regression-based model comparisons

Table 6 provides direct evidence on how well the various models capture the dynamics of volatility. The table reports the R-squared values for a regression of the log realized variances on the fitted log variances from each of the four EGARCH specifications. The R-squared values for the basic EGARCH(1,1) model range from 16.8 percent for Eastman Kodak to 50.1 percent for AT&T. This range is roughly consistent with the evidence reported by Andersen and Bollerslev (1998) for a GARCH(1,1) model. Since the basic model captures up to 50 percent of the variation in the log realized variances, it should represent a reasonable benchmark for assessing the performance of the other three models.

If contemporaneous volume largely subsumes ARCH effects, then we should find that the VA-EGARCH(1,1) model performs at least as well as the basic model. We find this not to be the case. The VA-GARCH(1,1) model produces a lower R-squared value for 15 of the 20 stocks and, in some cases, the reduction exceeds 20 percentage points. Apparently, the addition of trading volume forces the model to place too little weight on the lagged absolute returns, leading to variance estimates that have a lower correlation with the realized variances than the estimates from the basic model. Thus, the results support our earlier conclusions about the shortcomings of the VA-EGARCH(1,1) specification.

The R-squared values for the EGARCH(2,2) model are similar to those for the EGARCH(1,1) model. However, adding contemporaneous volume to the EGARCH(2,2) specification leads to a substantial increase in the R-squared value for most of the firms, typically on the order of 10 to 20 percentage points. In most cases, the R-squared for the VA-EGARCH(2,2) model exceeds 50 percent. This finding confirms the need to properly account for the short-term impact of the information contained in daily volume in order to uncover the true nature of the relation between ARCH effects and trading volume.

The implications of these findings go beyond the specific models considered here. There are many examples of studies in the literature that include explanatory variables other than volume in a GARCH(1,1) specification. Day and Lewis (1992) and Lamoureux and Lastrapes (1993), for example, add implied volatilities to the model. Fujihara and Mougoue (1997) and Girma and Mougoue (2002) add bid-ask spreads and open interest. Although a direct investigation of the empirical performance of these specifications is beyond the scope of the paper, our results suggest that they may deliver unreliable inferences if the variables added to the conditional variance function capture a component of volatility different from that captured by lagged squared returns.

3 Conclusions

The relation between volume and volatility has attracted a great deal of interest in the finance literature. We investigate a particular aspect of this relation: the ability of volume to explain ARCH effects. Studies such has Lamoureux and Lastrapes (1990) and Marsh and Wagner (2003) report that ARCH effects tend to vanish when contemporaneous volume is added to the conditional variance equation of a GARCH(1,1) model. We demonstrate that this is mainly due to structural constraints inherent in the GARCH(1,1) specification. When volume is added to the model, the associated coefficient is constrained to decline with the lag length at the same rate as the coefficient on the squared residual. Thus, if the impact of volume on volatility is strong but short lived, ARCH effects may appear to vanish because the squared residuals must be downweighted to adequately capture volume effects.

Using a more flexible EGARCH(2,2) model that allows for both short- and long-term volatility components, we find little support for the proposition that volume explains ARCH effects. The model does imply that volume is strongly associated with return volatility, but this is primarily a short-term phenomenon that has little to do with the highly persistent component of volatility that is characteristic of GARCH processes. We find that it is important to incorporate both contemporaneous volume and the persistent component of volatility in order to explain realized volatility. These results suggest that if we want to identify the features of the trading process that give rise to ARCH effects in daily stock returns, we need to look beyond trading volume.

More generally, our analysis has implications for any GARCH(1,1) model that includes additional explanatory variables. Since these models are subject to the same kind of specification issue identified for VA-GARCH models, it can be difficult to properly interpret the model fitting results. The methodology developed in this paper overcomes the shortcomings of the GARCH(1,1) methodology. Thus, it should prove more reliable in applications that use augmented GARCH models to examine the role of variables other than lagged squared returns in explaining volatility dynamics.

Appendix

This appendix describes our methodology for constructing realized variance, including our choice of sampling frequency and our method of dealing with trading- versus nontrading-period returns. In theory, we should construct realized variance by sampling returns continuously. However, as the sampling frequency increases, returns become more negatively serially correlated due to market microstructure effects, which leads to a biased variance estimate. Moreover, high-frequency returns are not available on weekends or overnight. Our general strategy is to deal with these issues separately. First we construct realized variance for the trading day using an estimator that is robust to serial correlation in returns, and then we construct the full-day realized variance by appropriately weighting the trading-day realized variance and the squared return during the nontrading period.

We construct the realized variance for the trading day using the Newey-West (1987) estimator proposed by Hansen and Lunde (2004),

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

where q denotes the window length for the autocovariance terms. Since this estimator is consistent in the presence of serial correlation, it allows us to sample returns at a higher frequency and thereby incorporate information that might otherwise be lost.

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

$$RV_t = \varphi \frac{\psi}{\psi_o} RV_{t[o]} + (1-\varphi)\frac{\psi}{\psi_c} R_{t[c]}^2, \tag{A.2}$$

where

$$\varphi = \frac{\psi_o^2 \eta_c^2 - \psi_o \psi_c \eta_{oc}}{\psi_c^2 \eta_o^2 + \psi_o^2 \eta_c^2 - 2\psi_o \psi_c \eta_{oc}}, \tag{A.3}$$

and $\psi = \mathrm{E}(R_t^2)$, $\psi_o = \mathrm{E}(RV_{t[o]})$, $\psi_c = \mathrm{E}(R_{t[c]}^2)$, $\eta_o^2 = \mathrm{var}(RV_{t[o]})$, $\eta_c^2 = \mathrm{var}(R_{t[c]}^2)$, and $\eta_{oc} = \mathrm{cov}(RV_{t[o]}, R_{t[c]}^2)$. The ratios ψ/ψ_o and ψ/ψ_c scale the variance estimates to have the same unconditional mean as 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 variance is typically lower during the nontrading period than the trading period and $R^2_{t[c]}$ is a relatively imprecise estimator of the nontrading-period variance. These effects can most easily be seen in Equation (A.3) by assuming $\eta_{oc} = 0$.

To implement Equations (A.1) and (A.2), we use intraday transaction prices from the TAQ database. We apply the price filters described in the paper to eliminate obvious reporting errors and then use the remaining prices to construct returns. The trading day for stocks is usually 390 minutes in length (9:30am to 4:00pm EST). We consider sampling frequencies as high as $m = 780$ (i.e., 30-second returns). For a given choice of m, we need to find the price at the beginning and end of each $m/390$-minute interval. We start with the first price in the TAQ database for that day and treat it as the beginning price for the interval in which it occurs. We then estimate the price at the end of this and each successive interval by linear interpolation of the prices nearest (on either side) to the end of the interval (see Andersen and Bollerslev (1997)). If one or more prices occurs exactly at the end of the interval, we use the average of these prices. We use the last transaction price of the day as the price at the end of the last interval. We construct the returns by differencing these log prices. As expected, the returns have a negative first-order serial correlation which increases with the sampling frequency. The average correlation coefficient across the 20 MMI stocks is -0.07 for five-minute returns and -0.15 for 30-second returns. [9]

We use the intraday returns to construct $RV_{t[o]}$, using values of q that correspond to four different window lengths: 0, 15, 30, and 60 minutes. Using a window length of 0, the bias caused by microstructure effects is readily apparent. Realized variances constructed using five-minute returns, which is common practice in the literature, are on average 13 percent greater than the average squared open-to-close return. The bias is much worse at higher sampling frequencies. However, increasing the window length counteracts the bias. Using a 15-minute window, the realized variances are still noticeably biased; but, using a 30-minute window, the average realized variances at every 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, as including unnecessary covariance terms in Equation (A.1) reduces efficiency.

[9] These serial correlation coefficients (based on interpolated prices) are substantially smaller than those obtained using the last transaction price in each intraday time interval. This is true even if we use an MA(1) model to filter returns as in ABDE (2001).

Based on these results, we use the realized variances constructed using 30-second returns and a 30-minute window length in our construction of the full-day realized variances.

We construct the full-day realized variances by substituting the sample analogs of ψ, ψ_o, ψ_c, η_o^2, η_c^2, and η_{oc} into Equations (A.2) and (A.3). Hansen and Lunde (2005) suggest removing outliers from the estimation to avoid obtaining a negative weight on $R^2_{t[c]}$. Accordingly, we exclude days in which either $RV_{t[o]}$ or $R^2_{t[c]}$ are among the largest 0.5 percent of the observations for each 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 approximately 20 percent of the daily variance occurs during the nontrading period. However, the φ estimate gives less weight than this to the nontrading-period variance estimate because the trading-period variance estimate is much more precise.

References

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., and T. Bollerslev, 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885–905.

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

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.

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

Bollerslev, T. and J.M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances, *Econometric Reviews* 11, 143–172.

Clark, P.K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–156.

Davidian, M. and Carroll, R.J., 1987, Variance function estimation, *Journal of the American Statistical Association* 82, 1079–1091.

Day, T.E. and C.M. Lewis, 1992, Stock market volatility and the information content of stock index options, *Journal of Econometrics* 52, 267-287.

Fujihara R., and M. Mougoue, 1997, Linear dependence, nonlinear dependence and petroleum futures market efficiency, *Journal of Futures Markets* 17, 75-99.

Engle, R.F., and G.J. Lee, 1999, A permanent and transitory component model of stock return volatility, in (eds) Engle, R.F. and H. White, Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger, Oxford University Press.

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

Girma, P., and M. Mougoue, 2002, An empirical examination of the relation between futures spreads volatility, volume, and open interest, *Journal of Futures Markets* 22, 1083–1102.

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*, forthcoming.

Kim, D., and S.J. Kon, 1994, Alternative models for the conditional heteroscedasticity of stock returns, *Journal of Business* 67, 563–598.

Lamoureux, C.G. and W.D. Lastrapes, 1990, Heteroskedasticity in stock return data: Volume versus GARCH effects, *Journal of Finance* 45, 221–229.

Lamoureux, C.G. and W.D. Lastrapes, 1993, Forecasting stock-return variance: Toward an understanding of stochastic implied volatility, *Review of Financial Studies* 6, 293-326.

Liesenfeld, R., 1998, Dynamic bivariate mixture models: Modeling the behavior of prices and trading volume, *Journal of Business and Economic Statistics* 16, 101–109.

Marsh, T.A., and N. Wagner, 2003, Return-volume dependence and extremes in international equity markets, Working paper, University of California at Berkeley.

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

Nelson, D, 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.

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

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

Watanabe, T., 2000, Bayesian analysis of dynamic bivariate mixture models: Can they explain the behavior of returns and trading volume? *Journal of Business and Economic Statistics* 18, 199–210.

Table 1. EGARCH(1,1) model

	Estimates			t-ratios			Diagnostics			
Firm	ς	κ_h	σ_h	ς	κ_h	σ_h	$\mathcal{L}$	R^2	ρ	C_K
AXP	1.56	0.03	0.11	12.2	2.7	5.5	− 5893.3	0.12	0.98	1.35
CVX	0.89	0.02	0.07	6.8	1.8	3.9	− 4950.8	0.06	0.98	0.87
DD	1.45	0.01	0.07	7.1	1.2	2.4	− 5512.3	0.09	0.99	2.01
DIS	1.87	0.01	0.07	7.6	1.3	2.1	− 5873.6	0.09	0.99	5.82
DOW	1.61	0.01	0.08	5.0	1.8	3.9	− 5352.6	0.15	0.99	1.78
EK	1.48	0.34	0.22	12.3	2.3	3.9	− 5820.4	0.03	0.51	15.69
GE	1.05	0.01	0.07	2.9	1.1	3.1	− 5280.5	0.13	0.99	1.26
GM	1.48	0.03	0.08	13.6	2.1	4.3	− 5798.8	0.07	0.98	1.44
IBM	1.92	0.02	0.08	8.7	2.0	3.8	− 5984.5	0.05	0.98	4.02
IP	1.47	0.01	0.05	6.2	1.0	2.1	− 5630.9	0.10	0.99	1.44
JNJ	1.08	0.02	0.09	7.2	2.3	4.3	− 5181.9	0.08	0.98	1.61
KO	1.21	0.01	0.06	4.9	1.6	3.7	− 5173.9	0.10	0.99	1.81
MCD	1.48	0.01	0.06	6.8	2.3	4.0	− 5417.6	0.06	0.99	2.25
MMM	1.20	0.01	0.04	7.4	1.7	2.3	− 5107.4	0.07	0.99	3.98
MO	2.95	0.00	0.04	1.5	0.7	4.0	− 5870.4	0.05	0.99	10.16
MRK	1.38	0.01	0.04	7.7	1.3	2.7	− 5531.5	0.05	0.99	2.18
PG	1.56	0.00	0.06	2.7	1.9	5.6	− 5175.5	0.09	0.99	10.72
S	1.90	0.03	0.11	11.2	1.9	3.4	− 6150.0	0.07	0.96	4.02
T	2.17	0.01	0.06	6.9	1.1	1.7	− 5909.6	0.13	0.99	7.59
XOM	0.87	0.01	0.07	3.8	2.0	4.4	− 4798.7	0.11	0.99	1.09

The table reports the results of fitting an EGARCH(1,1) model to daily percentage returns on the MMI stocks. The model is of the form

$$r_t = \sqrt{h_t} z_{rt},$$
$$\Delta \log h_t = \kappa_h(\varsigma - \log h_{t-1}) + \sigma_h u_{t-1},$$

where r_t is the demeaned return for day t, $u_t = (|z_{rt}| - \mathrm{E}[|z_{rt}|])/\sqrt{\mathrm{var}(|z_{rt}|)}$, and $z_{rt} \sim \mathrm{NID}(0,1)$. We fit the model via maximum likelihood and report the parameter estimates, the associated t-ratios, the maximized value of log-likelihood ($\mathcal{L}$), the sample R-squared for a regression of $\{|r_t|\}_{t=1}^{T}$ on the fitted conditional volatilities (R^2), the first-order sample autocorrelation of the fitted conditional volatilities (ρ), and the coefficient of excess kurtosis for the standardized returns (C_K). The t-ratios are based on robust standard errors. The sample period is January 5, 1993 to December 31, 2003.

Table 2. VA-EGARCH(1,1) model

Firm	Estimates				t-ratios				Diagnostics			
	ς	κ_h	σ_h	γ_h	ς	κ_h	σ_h	γ_h	$\mathcal{L}$	R^2	ρ	C_K
AXP	1.32	1.19	0.09	0.74	44.7	24.1	3.6	21.3	− 5757.2	0.19	0.41	0.21
CVX	0.66	1.16	0.04	0.58	21.9	20.0	1.6	15.7	− 4836.1	0.12	0.25	0.37
DD	1.07	1.17	0.13	0.63	32.5	14.3	4.6	17.3	− 5410.6	0.15	0.35	0.62
DIS	1.21	1.23	0.13	0.78	39.8	30.5	5.7	22.6	− 5600.4	0.25	0.31	0.22
DOW	1.07	0.66	0.26	0.45	23.3	2.4	5.1	3.0	− 5378.7	0.15	0.63	1.17
EK	0.90	1.22	0.09	0.89	27.9	31.0	3.3	27.8	− 5171.8	0.37	0.33	0.59
GE	0.91	1.20	0.07	0.72	29.9	19.1	2.8	21.2	− 5190.0	0.21	0.37	0.38
GM	1.21	1.31	0.07	0.72	39.8	30.3	2.9	19.0	− 5598.9	0.19	0.30	0.40
IBM	1.23	1.26	0.08	0.81	38.7	32.1	2.9	22.3	− 5624.0	0.28	0.35	0.57
IP	1.15	1.10	0.16	0.63	35.3	17.5	6.2	17.9	− 5522.5	0.18	0.45	0.46
JNJ	0.75	1.17	0.07	0.68	25.7	21.2	2.8	19.8	− 4965.6	0.24	0.39	0.17
KO	0.74	1.14	0.05	0.73	26.4	21.7	2.1	23.0	− 4950.3	0.25	0.41	0.04
MCD	0.90	1.24	0.14	0.69	28.4	26.2	5.7	20.7	− 5177.4	0.23	0.28	0.45
MMM	0.67	1.27	0.06	0.74	21.4	26.0	2.3	19.9	− 4855.5	0.21	0.28	0.56
MO	1.00	1.26	0.05	0.96	17.3	28.6	1.2	22.1	− 5311.6	0.37	0.34	6.96
MRK	0.95	1.29	0.05	0.73	34.8	29.5	2.2	25.0	− 5245.8	0.25	0.27	-0.04
PG	0.71	1.15	0.12	0.68	22.8	23.1	4.6	20.1	− 4910.2	0.39	0.39	0.35
S	1.42	1.34	0.05	0.77	44.7	26.4	2.2	22.5	− 5894.6	0.26	0.29	0.67
T	1.39	1.17	0.15	0.68	33.0	7.4	2.8	10.0	− 5831.4	0.24	0.41	2.45
XOM	0.56	1.02	0.15	0.62	18.1	15.0	5.5	17.1	− 4705.7	0.16	0.44	0.21

The table reports the results of fitting a volume-augmented EGARCH(1,1) model to daily percentage returns on the MMI stocks. The model is of the form

$$r_t = \sqrt{h_t} z_{rt},$$
$$\Delta \log h_t = \kappa_h(\varsigma - \log h_{t-1}) + \sigma_h u_{t-1} + \gamma_h w_t,$$

where r_t is the demeaned return for day t, $u_t = (|z_{rt}| - \mathrm{E}[|z_{rt}|])/\sqrt{\mathrm{var}(|z_{rt}|)}$, $w_t = (\log v_t - \mathrm{E}[\log v_t])/\sqrt{\mathrm{var}(\log v_t)}$, and $z_{rt} \sim \mathrm{NID}(0, 1)$. We fit the model via maximum likelihood and report the parameter estimates, the associated t-ratios, the maximized value of log-likelihood ($\mathcal{L}$), the sample R-squared for a regression of $\{|r_t|\}_{t=1}^{T}$ on the fitted conditional volatilities (R^2), the first-order sample autocorrelation of the fitted conditional volatilities (ρ), and the coefficient of excess kurtosis for the standardized returns (C_K). The t-ratios are based on robust standard errors. The sample period is January 5, 1993 to December 31, 2003.

Table 3. EGARCH(2,2) model

	Estimates					t-ratios					Diagnostics			
Firm	κ_h	σ_h	ς	κ_m	σ_m	κ_h	σ_h	ς	κ_m	σ_m	$\mathcal{L}$	R^2	ρ	C_K
AXP	0.10	0.10	1.59	0.00	0.03	3.2	4.7	8.0	1.4	2.4	− 5884.4	0.13	0.97	1.27
CVX	1.73	0.03	0.90	0.01	0.06	9.9	1.2	6.5	1.8	4.0	− 4949.8	0.06	0.96	0.87
DD	0.22	0.11	1.42	0.00	0.03	2.5	3.9	4.7	1.1	2.9	− 5497.1	0.09	0.95	1.63
DIS	1.37	0.08	1.88	0.01	0.05	7.1	2.3	7.3	1.8	2.4	− 5866.6	0.10	0.91	5.75
DOW	0.14	0.07	1.62	0.00	0.04	0.5	1.6	5.2	1.3	1.9	− 5345.3	0.15	0.98	1.63
EK	0.44	0.21	1.73	0.00	0.01	1.8	3.4	9.7	2.7	2.4	− 5810.9	0.03	0.48	16.32
GE	0.13	0.08	-0.13	0.00	0.03	1.4	3.3	-0.3	-0.8	2.9	− 5268.1	0.13	0.98	1.00
GM	0.06	0.07	1.54	0.00	0.02	1.9	3.5	10.0	0.9	1.1	− 5795.1	0.07	0.97	1.41
IBM	0.09	0.05	1.99	0.01	0.05	1.1	1.8	8.7	2.5	3.4	− 5982.4	0.06	0.97	4.08
IP	0.17	0.08	1.36	0.00	0.03	2.5	2.8	4.3	1.2	4.3	− 5618.0	0.11	0.97	1.20
JNJ	0.86	0.08	1.10	0.02	0.07	3.8	2.3	6.6	2.2	3.8	− 5178.0	0.08	0.92	1.60
KO	0.41	0.09	1.18	0.00	0.04	2.7	3.0	4.1	1.6	5.1	− 5164.3	0.10	0.96	1.70
MCD	1.07	0.11	1.46	0.01	0.05	3.4	3.1	6.7	2.2	3.5	− 5404.6	0.07	0.87	2.47
MMM	1.04	0.10	1.18	0.01	0.04	1.1	1.9	7.9	2.0	2.3	− 5098.5	0.07	0.89	3.62
MO	0.51	0.11	3.11	0.00	0.03	1.7	2.9	1.9	1.0	4.7	− 5848.7	0.06	0.91	10.68
MRK	0.63	0.06	1.41	0.01	0.03	0.2	1.5	5.3	0.6	1.1	− 5528.5	0.05	0.95	2.19
PG	1.46	0.04	1.54	0.00	0.06	3.1	1.0	2.6	1.9	5.0	− 5173.5	0.09	0.96	11.03
S	1.90	-0.02	1.89	0.04	0.11	28.8	-1.1	11.9	1.8	3.2	− 6148.3	0.07	0.95	3.82
T	0.21	0.11	2.03	0.00	0.03	3.8	4.7	7.2	1.9	3.4	− 5886.6	0.14	0.96	7.18
XOM	0.30	0.09	0.88	0.01	0.05	3.5	3.7	3.2	2.0	5.4	− 4788.4	0.11	0.96	1.01

The table reports the results of fitting an EGARCH(2,2) model to daily percentage returns on the MMI stocks. The model has a two-component representation of the form

$$r_t = \sqrt{h_t} z_{rt},$$

$$\Delta \log h_t = \Delta m_t + \kappa_h (m_{t-1} - \log h_{t-1}) + \sigma_h u_{t-1},$$

$$\Delta m_t = \kappa_m (\varsigma - m_{t-1}) + \sigma_m u_{t-1},$$

where r_t is the demeaned return for day t, $u_t = (|z_{rt}| - \mathrm{E}[|z_{rt}|])/\sqrt{\mathrm{var}(|z_{rt}|)}$, and $z_{rt} \sim \mathrm{NID}(0,1)$. We fit the model via maximum likelihood and report the parameter estimates, the associated t-ratios, the maximized value of log-likelihood ($\mathcal{L}$), the sample R-squared for a regression of $\{|r_t|\}_{t=1}^T$ on the fitted conditional volatilities (R^2), the first-order sample autocorrelation of the fitted conditional volatilities (ρ), and the coefficient of excess kurtosis for the standardized returns (C_K). The t-ratios are based on robust standard errors. The sample period is January 5, 1993 to December 31, 2003.

Table 4. Volume-augmented EGARCH(2,2) model

	Estimates							t-ratios							Diagnostics			
Firm	κ_h	σ_h	γ_h	ς	κ_m	σ_m	γ_m	κ_h	σ_h	γ_h	ς	κ_m	σ_m	γ_m	$\mathcal{L}$	R^2	ρ	C_K
AXP	1.21	-0.05	0.72	1.01	0.01	0.06	0.00	32.4	-1.8	20.7	6.0	1.9	4.1	0.7	− 5618.9	0.29	0.62	-0.03
CVX	1.15	-0.07	0.60	0.48	0.01	0.05	0.00	23.9	-2.4	18.2	3.2	1.5	3.5	0.1	− 4730.4	0.18	0.45	0.02
DD	1.25	0.01	0.62	0.78	0.01	0.06	0.00	24.8	0.3	20.3	3.6	1.3	3.0	-0.3	− 5246.3	0.25	0.57	0.06
DIS	1.24	0.01	0.79	0.78	0.00	0.05	0.00	36.2	0.2	23.5	4.0	2.4	6.6	-0.5	− 5437.6	0.35	0.49	-0.18
DOW	1.14	-0.01	0.61	0.92	0.00	0.06	0.00	21.5	-0.5	18.3	3.9	1.9	5.5	-1.4	− 5112.2	0.27	0.69	0.39
EK	1.18	-0.03	0.96	0.83	0.06	0.10	-0.02	28.5	-0.7	26.3	12.6	1.9	3.5	-1.6	− 5093.0	0.43	0.35	0.50
GE	1.24	-0.05	0.67	-0.32	0.00	0.03	0.00	28.4	-1.9	23.3	-1.0	1.8	5.2	1.6	− 4992.9	0.34	0.65	-0.29
GM	1.30	-0.03	0.76	1.05	0.02	0.06	-0.01	33.5	-1.0	23.9	11.2	2.9	5.3	-1.6	− 5492.0	0.26	0.41	-0.11
IBM	1.21	-0.09	0.89	0.83	0.01	0.06	-0.01	36.7	-3.2	24.6	5.2	2.3	5.0	-2.4	− 5429.8	0.41	0.44	-0.18
IP	1.20	0.02	0.64	0.56	0.00	0.04	0.00	26.8	0.5	19.4	0.9	0.6	1.7	-0.6	− 5377.7	0.27	0.65	0.05
JNJ	1.21	-0.03	0.70	0.61	0.02	0.06	0.00	29.3	-1.1	19.8	6.5	3.2	6.2	-1.4	− 4883.8	0.30	0.46	-0.10
KO	1.21	-0.04	0.69	0.37	0.01	0.04	0.00	30.7	-1.5	24.2	2.3	2.5	5.8	0.9	− 4835.8	0.34	0.56	-0.26
MCD	1.26	0.04	0.72	0.69	0.01	0.05	0.00	29.3	1.2	24.3	3.9	1.4	3.1	-0.6	− 5059.6	0.32	0.41	0.11
MMM	1.27	-0.02	0.73	0.49	0.01	0.04	0.00	27.4	-0.7	22.8	2.1	1.3	3.4	-0.9	− 4733.2	0.29	0.47	0.11
MO	1.26	-0.12	1.02	0.93	0.03	0.10	-0.01	43.6	-2.4	26.7	5.3	1.7	3.6	-0.9	− 5218.1	0.41	0.41	8.23
MRK	1.28	-0.03	0.73	0.69	0.01	0.06	0.00	32.2	-0.9	25.5	5.8	2.5	5.5	0.6	− 5154.3	0.30	0.41	-0.30
PG	1.18	0.00	0.72	0.17	0.00	0.05	0.00	28.8	-0.1	22.0	0.6	1.3	4.1	-0.5	− 4762.9	0.45	0.46	-0.07
S	1.26	-0.06	0.81	1.24	0.01	0.08	-0.01	29.5	-2.1	24.4	9.8	2.4	5.1	-1.6	− 5743.7	0.36	0.45	0.00
T	1.26	-0.02	0.83	0.82	0.01	0.07	-0.01	34.8	-0.8	24.4	2.7	1.7	5.3	-1.5	− 5420.7	0.37	0.60	0.18
XOM	1.13	-0.01	0.60	0.08	0.01	0.05	0.00	24.6	-0.4	21.9	0.3	1.5	5.5	1.0	− 4552.5	0.26	0.66	-0.25

The table reports the results of fitting a volume-augmented EGARCH(2,2) model to daily percentage returns on the MMI stocks. The model has a two-component representation of the form

$$r_t = \sqrt{h_t} z_{rt},$$
$$\Delta \log h_t = \Delta m_t + \kappa_h (m_{t-1} - \log h_{t-1}) + \sigma_h u_{t-1} + \gamma_h w_t,$$
$$\Delta m_t = \kappa_m (\varsigma - m_{t-1}) + \sigma_m u_{t-1} + \gamma_m w_t,$$

where r_t is the demeaned return for day t, $u_t = (|z_{rt}| - \mathrm{E}[|z_{rt}|])/\sqrt{\mathrm{var}(|z_{rt}|)}$, $w_t = (\log v_t - \mathrm{E}[\log v_t])/\sqrt{\mathrm{var}(\log v_t)}$, and $z_{rt} \sim \mathrm{NID}(0,1)$. We fit the model via maximum likelihood and report the parameter estimates, the associated t-ratios, the maximized value of log-likelihood ($\mathcal{L}$), the sample R-squared for a regression of $\{|r_t|\}_{t=1}^T$ on the fitted conditional volatilities (R^2), the first-order sample autocorrelation of the fitted conditional volatilities (ρ), and the coefficient of excess kurtosis for the standardized returns (C_K). The t-ratios are based on robust standard errors. The sample period is January 5, 1993 to December 31, 2003.

Table 5. Volume versus ARCH effects

	Estimated var($\log h_t$) for the different models				Components of var($\log h_t$) for the VA-EGARCH(2,2) model		
Firm	EGARCH(1,1)	VA-EGARCH(1,1)	EGARCH(2,2)	VA-EGARCH(2,2)	Short-term	Long-term	Interaction
AXP	0.301	0.484	0.315	0.663	0.425	0.214	0.024
CVX	0.164	0.304	0.168	0.444	0.320	0.153	-0.028
DD	0.271	0.371	0.289	0.581	0.318	0.268	-0.004
DIS	0.281	0.527	0.287	0.736	0.511	0.257	-0.031
DOW	0.516	0.445	0.509	0.790	0.327	0.478	-0.016
EK	0.129	0.654	0.133	0.724	0.777	0.141	-0.194
GE	0.439	0.448	0.509	0.741	0.371	0.358	0.012
GM	0.148	0.415	0.151	0.534	0.450	0.148	-0.063
IBM	0.217	0.526	0.217	0.729	0.634	0.299	-0.204
IP	0.293	0.392	0.304	0.598	0.344	0.260	-0.006
JNJ	0.190	0.411	0.195	0.504	0.406	0.124	-0.026
KO	0.280	0.475	0.290	0.606	0.393	0.173	0.040
MCD	0.214	0.416	0.218	0.543	0.425	0.190	-0.072
MMM	0.169	0.448	0.180	0.597	0.424	0.255	-0.018
MO	0.274	0.731	0.291	0.890	0.800	0.160	-0.070
MRK	0.115	0.432	0.119	0.537	0.421	0.137	-0.020
PG	0.369	0.431	0.367	0.691	0.435	0.251	0.005
S	0.230	0.453	0.228	0.613	0.509	0.227	-0.123
T	0.516	0.418	0.511	1.054	0.535	0.625	-0.105
XOM	0.328	0.415	0.343	0.627	0.329	0.268	0.030

The table examines the extent to which trading volume captures ARCH effects in daily returns on the MMI stocks. We report the sample variance of the fitted log volatility series from the models in Tables 1 – 4. In addition, we decompose the variance for the model in Table 4 into three components — short-term, long-term, and interaction — using the relation $\mathrm{var}(\log h_t) = \mathrm{var}(\log h_t - m_t) + \mathrm{var}(m_t) + 2\,\mathrm{cov}(\log h_t - m_t, m_t)$. The sample period is January 5, 1993 to December 31, 2003.

Table 6. Realized variance regressions

Firm	Regression R-squared for the different models EGARCH(1,1)	VA-EGARCH(1,1)	EGARCH(2,2)	VA-EGARCH(2,2)
AXP	0.376	0.348	0.384	0.555
CVX	0.283	0.170	0.284	0.377
DD	0.405	0.203	0.417	0.520
DIS	0.403	0.238	0.411	0.507
DOW	0.486	0.279	0.496	0.562
EK	0.168	0.318	0.223	0.439
GE	0.482	0.282	0.493	0.636
GM	0.240	0.261	0.246	0.424
IBM	0.309	0.283	0.322	0.578
IP	0.447	0.209	0.463	0.493
JNJ	0.232	0.293	0.243	0.404
KO	0.369	0.291	0.383	0.521
MCD	0.241	0.216	0.257	0.426
MMM	0.340	0.238	0.354	0.491
MO	0.267	0.416	0.303	0.529
MRK	0.246	0.289	0.259	0.488
PG	0.358	0.281	0.362	0.521
S	0.320	0.203	0.311	0.440
T	0.501	0.221	0.522	0.613
XOM	0.433	0.226	0.445	0.520

The table reports the R-squared for the regression

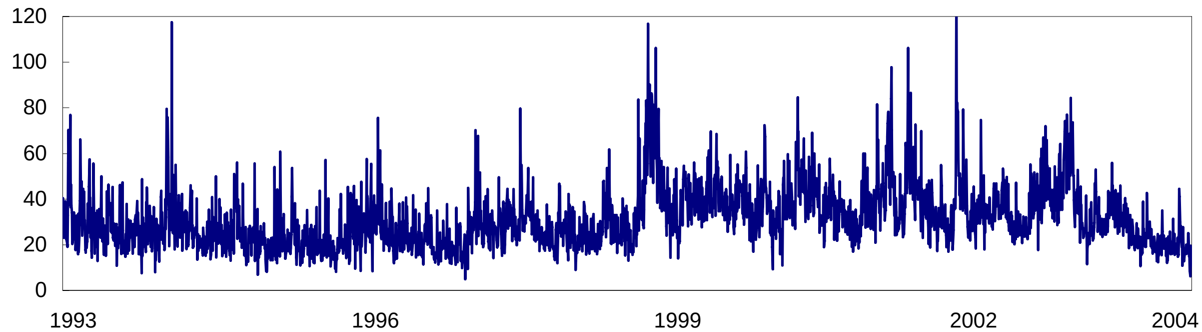
$$\log RV_t = a + b \log \hat{h}_t + e_t$$

where RV_t is the realized variance for day t and $\log \hat{h}_t$ is the fitted log variance for day t for each of the models in Tables 1 – 4. We construct the realized variance using the full-day Newey-West estimator described in the Appendix, with a 30-second sampling frequency and a window length of 30 minutes. The sample period is January 5, 1993 to December 31, 2003.

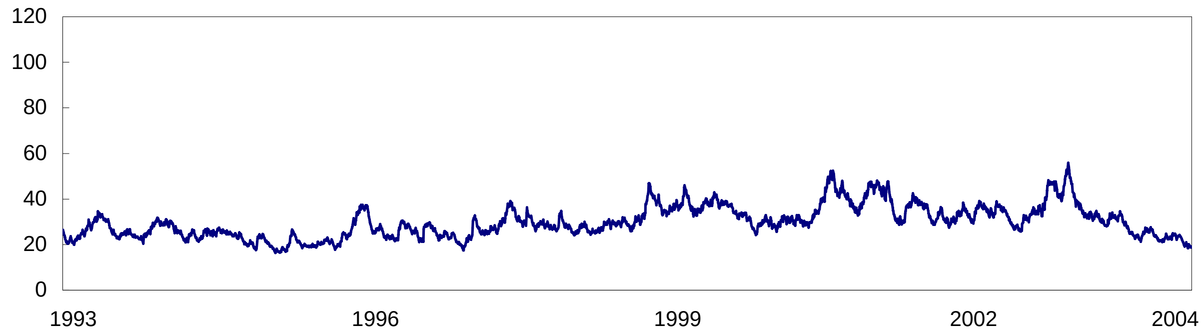
Figure 1
Comparison of volatility estimates for American Express

The figure plots the daily volatility estimates for American Express under the EGARCH(1,1) model and the volume-augmented (VA) EGARCH(2,2) model. Panel A shows the fitted volatility estimates for the VA-EGARCH(2,2) model, Panel B shows the fitted estimates of the long-term component of volatility under the VA-EGARCH(2,2) model, and Panel C shows the fitted volatility estimates for the EGARCH(1,1) model. Each series is expressed as an annualized percentage volatility. The sample period is January 5, 1993 to December 31, 2003.

Panel A: VA-EGARCH(2,2) volatility estimates



Panel B: VA-EGARCH(2,2) estimates of the long-term component of volatility



Panel C: EGARCH(1,1) volatility estimates