Journal of Financial Econometrics, 2012, Vol. 10, No. 1, 54–83

Forecasting intraday volatility in the US equity market. Multiplicative component GARCH

ROBERT F. ENGLE
New York University

MAGDALENA E. SOKALSKA
City University of New York

ABSTRACT

This paper proposes a new intraday volatility forecasting model, particularly suitable for modeling a large number of assets. We decompose volatility of high-frequency returns into components that may be easily interpreted and estimated. The conditional variance is a product of daily, diurnal, and stochastic intraday components. This model is applied to a comprehensive sample consisting of 10-minute returns on more than 2500 US equities. Apart from building a new model, we obtain several interesting forecasting results. We apply a number of different specifications. We estimate models for separate companies, pool data into industries, and consider other criteria for grouping returns. In general, forecasts from pooled cross-section of companies outperform the corresponding forecasts from company-by-company estimation. For less liquid stocks, however, we obtain better forecasts when we group less frequently traded companies together. (*JEL*: C22, C51, C53, G15)

KEYWORDS: ARCH, intraday returns, volatility

We would like to thank the editors: Eric Ghysels, René Garcia, Eric Renault, as well as anonymous referees for their insightful comments. We are grateful to Robert Ferstenberg and Ananda Chanda, who contributed to earlier versions of the paper. The authors gratefully acknowledge that financial support for the research was provided by Morgan Stanley to New York University's Stern School of Business. The views expressed herein are solely those of the authors and not those of any other person or entity. Address correspondence to Magdalena E. Sokalska, Department of Economics, Queens College, City University of New York, 65-30 Kissena Boulevard, Flushing NY 11367, or e-mail: msokalska@qc.cuny.edu.

doi: 10.1093/jjfinec/nbr005

Let us consider a situation when a trading desk of a large bank needs to forecast volatility for more than 2500 stocks each 10 minutes every day. These intraday volatility forecasts serve as input for algorithms that schedule trades and help place limit orders. Given an unprecedented surge in automated trading in major financial markets over the last years, the high-frequency volatility forecasting model is of substantial practical importance for the finance industry.

Conventional GARCH (Engle 1982, Bollerslev 1986) approaches were argued to be unsatisfactory for modeling intraday returns by authors at the Olsen conference on High Frequency Data Analysis in Zurich in March 1995. As shown in Andersen and Bollerslev (1997), estimation of intraday MA(1)-GARCH(1,1) model for different intraday frequencies gives parameters that are inconsistent between these frequencies and do not comply with theoretical results of Drost and Nijman (1993) on time aggregation of GARCH models. The pronounced diurnal patterns of volatility and trading activity are responsible for these difficulties.

A number of closely connected models were developed to take account of intraday volatility patterns, compare Ghose and Kroner (1996), Andersen and Bollerslev (1997, 1998), and Giot (2005). Andersen and Bollerslev (1997) build a multiplicative model of daily and diurnal volatility for five-minute returns on the Deutschemark-dollar exchange rate. The conditional variance is expressed as a product of daily and diurnal components. Andersen and Bollerslev (1998) add an additional component which takes account of the impact of macroeconomic announcements on the volatility. This specification proved popular in the literature on intraday foreign exchange volatility. Typically authors work with at most several time series and can easily monitor the relevant macroeconomic announcements.

Faced with a task of forecasting volatility of high-frequency returns for a huge number of equities, we felt that the specification of the intraday component as a function of dummy variables associated with particular announcements was not very practical. First, important macroeconomic announcements happen before the stock market opens. Second, idiosyncratic announcements can be expected to be particularly important for equities, and the timing of the majority of them can be difficult to predict. Third, the reaction of the market heavily depends on whether the news was genuinely unexpected. Finally, stock markets are generally considered more vulnerable to asymmetric information and we cannot rely on public announcements as the only information revelation channel. In particular, macroeconomic or public announcement dummies cannot account for information arrival through order flow.

Our model builds on the work of Andersen and Bollerslev (1997, 1998) and decomposes the volatility of high-frequency asset returns into multiplicative components, which may be easily interpreted and estimated. The conditional variance is expressed as a product of daily, diurnal, and stochastic intraday volatility components. This model is applied to a comprehensive sample consisting of 10-minute returns on more than 2500 US equities. We apply a number of different specifications. Namely, we construct models for separate companies, pool data into industries, and consider various criteria for grouping returns. It turns out that results for the pooled regressions seem to be more stable. The forecasts from the pooled specifications outperform the corresponding forecasts from company-by-company estimation, and we discuss several issues regarding the best way to pool. Additionally, in contrast to the aforementioned multiplicative component literature, we derive statistical properties of the multistep estimator of the model.

For most of their models, Andersen and Bollerslev (1998) specify the intradaily volatility component as deterministic. In contrast, we include two intradaily components: a deterministic diurnal pattern and stochastic intradaily ARCH.

Overall, in distinction to a huge volume of literature on daily volatility models, forecasting intradaily volatility is far less discussed. Apart from the papers quoted above, Giot (2005) estimates GARCH(1,1) (Bollerslev 1986) and EGARCH(1,1) (Nelson 1991) models for high-frequency equity returns, after accounting for the deterministic intraday patterns with cubic splines. He does not include a separate daily component, and the stochastic intraday component does not seem to have enough persistence to carry through across days. In another specification, Giot adds contemporaneous microstructure variables into the conditional variance equation. Taylor and Xu (1997) construct an hourly volatility model using an ARCH specification and supplementing the conditional variance equation by two additional elements: the implied volatility and the realized volatility computed from the high-frequency data. A long memory stochastic volatility approach was applied by Deo, Hurvich, and Lu (2005). Their paper diurnally adjusts in the frequency domain and then uses a local Whittle estimator on log of squared returns to estimate the parameters.

We expect our intraday model to be of particular interest for derivative traders or hedge funds who seek high-frequency measures of risk or time-varying hedge ratios. Volatility estimates on an intraday basis could be used to evaluate risk of slow trading (Engle and Ferstenberg 2007) or as input to measures of time-varying liquidity. Most importantly, however, intraday volatility estimates are useful for devising optimal strategies to place limit orders or schedule trades. The literature on order choice supplies sufficient evidence that volatility is an important factor in order submission strategies (cf. Ellul *et al.* 2007; Griffiths *et al.* 2000).

Our paper is organized as follows. Sections 1 and 2 present the model and discuss statistical properties of an estimator. Section 3 describes the data and gives results of estimation. This is followed by a forecasting section and conclusions.

1 THE MODEL

1.1 Notation

We use the following notation. Days in the sample are indexed by $t (t = 1, \ldots, T)$. Each day is divided into 10-minute intervals referred to as bins and indexed by

$i(i = 0, \ldots, N)$. The current period is (t, i). The price of an asset at the end of bin i of day t is denoted by $P_{t,i}$. The continuously compounded return $r_{t,i}$ is calculated as

$$\begin{aligned} r_{t,i} &= \ln\left(\frac{P_{t,i}}{P_{t,i-1}}\right) \qquad \text{for } i \geqslant 1 \\ &= \ln\left(\frac{P_{t,1}}{P_{t-1,N}}\right) \qquad \text{for } i = 0. \end{aligned} \tag{1}$$

The overnight return in bin zero is deleted leading to a total number of return observations, M=TN. The consequences of deleting overnight returns are discussed in Section 3.1.

1.2 Model

We propose a GARCH model for high-frequency intraday financial returns, which specifies the conditional variance to be a multiplicative product of daily, diurnal, and stochastic intraday volatility. Intraday equity returns are described by the following process:

$$r_{t,i} = \sqrt{h_t s_i q_{t,i}} \varepsilon_{t,i} \quad \text{and} \quad \varepsilon_{t,i} \sim N(0,1), \tag{2}$$

where

h_t is the daily variance component,
s_i is the diurnal (calendar) variance pattern,
$q_{t,i}$ is the intraday variance component, with $E(q_{t,i}) = 1$, and
$\varepsilon_{t,i}$ is an error term.

For the daily variance component, h_t, our paper uses commercially available volatility forecasts produced for each company in the sample. The volatility forecasts are made on the basis of a multifactor risk model—for a detailed analysis of one of the available models, see Fabozzi, Jones, and Vardharaj (2002). We chose such a solution because we see a great value in combining structural modeling with time series analysis. Multifactor models offer an interesting alternative to prediction based purely on past price, like realized volatility (RV) or daily GARCH. Risk factors are expected to capture the common variation among stock returns. For daily horizons, the common risk factors mainly include an industry factor, as well as common liquidity factors: capitalization, volume, and spread. A third group involves a momentum factor, a company's alpha and recent price performance (cf. Fabozzi, Jones, and Vardharaj 2002).

Andersen and Bollerslev (1997, 1998) took a different approach and forecasted the daily component from a daily GARCH. They used a longer daily sample, going back a number of years. Our implementation eliminates the need for longer series for the daily model. With the high turnover of corporate ownership, it is difficult to get consistent long series for a big universe of stocks.

Apart from daily GARCH or multifactor models, the daily variance component could be estimated based on daily realized variance as proposed by Engle and Gallo (2006).

Assuming that intradaily returns are serially uncorrelated, the daily conditional variance is the sum of the variances in each of the bins. Hence

$$E\left(\sum_{i=1}^{N} r_{t,i}^2/h_t\right) = \lambda. \tag{3}$$

If the daily variance estimator is unbiased and if overnight returns are included in the analysis, then $\lambda = 1$. If this is not the case and if the bias is constant over time, then it will take a value different from one. This will, however, not influence the subsequent analysis.

Likewise, there are alternative ways to capture s_i—the deterministic diurnal pattern. Andersen and Bollerslev (1997, 1998) use flexible Fourier transform, for an example of a functional data analysis of intraday volatility, see Mueller, Sen, and Stadtmuller (2007). We propose a simple specification with the diurnal component which allows complete freedom for the daily shape. We simply calculate s_i as the variance of returns in each bin after deflating by the daily variance

$$\frac{r_{t,i}^2}{h_t} = s_i q_{t,i} \varepsilon_{t,i}^2$$

and

$$E\left(\frac{r_{t,i}^2}{h_t}\right) = s_i E(q_{t,i}) = s_i. \tag{4}$$

Thus, the sum of the s_i, $\sum s_i = \lambda$ is given by equation (3). To facilitate notation in later sections, we will refer to the return adjusted by the daily standard deviation as $y_{t,i} = r_{t,i}/\sqrt{h_t}$.

We estimate the model in two steps. First, we estimate the diurnal component for each bin following equation (4):

$$\hat{s}_i = \frac{1}{T}\sum_{t=1}^{T} \frac{r_{t,i}^2}{h_t}. \tag{5}$$

After normalizing returns by daily and diurnal volatility components, we model the residual volatility as a GARCH(p,q) process

$$z_{t,i} = r_{t,i}/\sqrt{h_t s_i} = \sqrt{q_{t,i}}\varepsilon_{t,i}, \tag{6}$$

$$q_{t,i} = \omega + \alpha(r_{t,i-1}/\sqrt{h_t s_{i-1}})^2 + \beta\, q_{t,i-1}. \tag{7}$$

In summary, the GARCH specification can be rewritten as

$$\begin{aligned} z_{t,i}\,|F_{t,i-1} &\sim N\left(0, q_{t,i}\right), \\ q_{t,i} &= \omega + \alpha z_{t,i-1}^2 + \beta\, q_{t,i-1}, \\ z_{t,i} &= r_{t,i}/\sqrt{h_t s_i}. \end{aligned} \tag{8}$$

GARCH(1,1) proved to be an adequate specification in most of the cases. The unit GARCH might enforce the constraint $\omega = 1 - \alpha - \beta$ although in the empirical work this has not been done. In some of the empirical work, an asymmetric GARCH model is tested. Many other specifications could be considered.

After reviewing an extensive ARCH/GARCH literature, Bollerslev, Chou, and Kroner (1992) found that GARCH(1,1) is the most popular model. If any higher lags are included in the conditional variance specification, the models are not of a higher order than GARCH(1,2) or GARCH(2,1). Nelson (1990) offers a possible explanation of this empirical regularity. Whereas under mild regularity conditions, many ARCH/GARCH models could be consistent filters of a particular diffusion process, efficiency considerations favor models of a lower order.

2 ECONOMETRIC ISSUES

In this section, we will discuss statistical properties of a two-step estimator of the model outlined in the previous section. The estimation proceeds in two steps. First, we specify and estimate the diurnal component. The second step consists of standardizing $y_{t,i}$ by $\sqrt{\hat{s}_i}$ and estimating parameters of the GARCH(p,q) model, which describes the dynamics of the intraday stochastic component as in equation (7). Such a multistep estimation strategy is potentially misleading as errors in one stage can lead to errors in the next stage. Nevertheless, it will be shown below that the estimator is consistent but that the standard errors should be adjusted.

In deriving the asymptotic properties of the estimators in this sequential procedure, we will follow Newey and McFadden (1994) (later denoted as NM) and cast the above steps into the GMM framework. We will consider the GMM estimator of the moment conditions stacked one on the other. We will use the following notation. Vector $\psi = \begin{pmatrix} \phi \\ \theta \end{pmatrix}$ contains both the k_1 parameters ϕ, estimated in the first step, and the k_2 parameters θ, estimated in the second step. Let there be k_1 moment conditions $g_1(\phi)$ and k_2 moment conditions $g_2(\phi, \theta)$ comprising vector $g(\psi) = \begin{pmatrix} g_1 \ (\phi) \\ g_2 \ (\phi, \theta) \end{pmatrix}$. The corresponding sample sums are g_{1M} and g_{2M}, giving g_M=$(g'_{1M}, g'_{2M})'$. We will consider the GMM estimator of the parameter vector

$$\hat{\psi} = \begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} = \arg\min g_M W g_M = \arg\min g_M g_M. \tag{9}$$

Since it is a just-identified system, $W = I$. To solve this system, ϕ must solve the first set of equations and θ must solve the second set conditional on the estimated value of ϕ. Thus, it is a natural framework to analyze two-step estimators of this type. Newey and McFadden (1994) (cf. their Theorem 6.1, p. 2178), have shown that if $\hat{\phi}$ and $\hat{\theta}$ are consistent estimators of the true ϕ_0 and θ_0, respectively, and g_M satisfies a number of standard regularity conditions, the resulting GMM estimator is consistent and asymptotically normal:

$$\sqrt{M}\begin{pmatrix}\hat{\phi}\ \phi_0\\ \hat{\theta}\ \theta_0\end{pmatrix} \xrightarrow{d} N(0, G^{-1}\Omega G^{-1\prime}), \tag{10}$$

where $G = E\left(\frac{\partial g(\psi)}{\partial \psi'}\right)$ and $\Omega = E(g(\psi)g(\psi)')$.

As in Hansen (1982), the above matrices can be consistently estimated by replacing expectations by sample averages and parameters by their estimates.

The NM approach is very convenient and may be applied when parameters at some steps are estimated by ML. In this case, some of the GMM moment conditions are taken to be score functions. In the current two-step setting, the sample sums in the first and the second stages are

$$g_{1,M}(\phi) = g_{1,M} = \begin{pmatrix} 1/T\sum_{t,1}^{T}(y_{t,1}^2 - s_1) \\ \vdots \\ 1/T\sum_{t,N}^{T}(y_{t,N}^2 - s_N)\end{pmatrix}, \tag{11}$$

$$g_{2,M}(\hat{\phi},\theta) = g_{2,M} = 1/TN\sum_{t=1}^{T}\sum_{i=1}^{N}\nabla_\theta(\log(q_{t,i}) + (y_{t,i}^2/\hat{s}_i q_{t,i})), \tag{12}$$

$$G = \frac{1}{M}\sum\begin{pmatrix}\nabla_\phi g_1 & 0\\ \nabla_\phi g_2 & \nabla_\theta g_2\end{pmatrix} \quad \text{and} \quad \frac{1}{M}\sum\begin{pmatrix} g_{1,t}^2 & g_{1,t}g_{2,t}\\ g_{1,t}g_{2,t} & g_{2,t}^2\end{pmatrix} \xrightarrow{p} \Omega. \tag{13}$$

In order to apply NM's Theorem 6.1, we have to make sure that $\hat{\phi}$ and $\hat{\theta}$ are consistent estimators of the true parameter values at each stage. This is indeed the case for estimator (5). In random sampling from a stationary ergodic distribution, the sample mean is a consistent estimate of the expected value. Consistency of $\hat{\theta}$ follows from, for example, Lee and Hansen (1994) or Lumsdaine (1996). In sum, the consistency and asymptotic normality of the two-step estimator (12) is a corollary to Theorem 6.1 (p. 2178) in Newey and McFadden (1994). The above results could, in principle, be generalized to a multistep estimation.

3 EMPIRICAL RESULTS

3.1 Data

Our sample consists of price data on 2721 companies obtained from the TAQ database. Data span a three month period from April to June 2000. We analyze logarithmic returns standardized by a commercially available volatility forecast for each company and each day and the standard deviation of returns in each 10-minute bin. The returns were calculated using transaction prices. The overnight return in bin zero has been deleted. Although the inclusion of overnight returns could yield a more complete analysis, it would come at a price of a greater model complexity. Our paper was "practically" motivated and a preliminary estimation indicated a poor forecasting performance for the first trading period. Excluding overnight returns in empirical analysis is a fairly common practice. This is a modeling assumption we took, and we maintain that this step does not adversely affect the other parts of our analysis. The consequences of excluding overnight returns are algebraic. In particular, under the assumption of no autocorrelation in intraday returns and the overnight return included $\sum_{i=1}^{N} s_i = 1$ from equation (4). When we exclude the overnight return, we lose this elegant interpretation since the sum of the s_i components no longer equals 1. However, we never impose the condition in the estimation. As long as the overnight volatility is proportional to the daily volatility, this summation will be a constant smaller than one, and the average of the intradaily GARCH will be the reciprocal of this constant.

3.2 Results for a Single Stock

Some results will be presented using returns on a single randomly chosen NYSE-listed stock—Valero Energy Corporation (VLO). Valero Energy operates as an independent petroleum refining and marketing company. As outlined in Section 1.2, we divide 10-minute intraday returns by their respective daily commercially available volatility forecasts. Returns adjusted in this way exhibit very clear diurnal volatility pattern. Figure 1 plots the standard deviation of returns in each of 39 10-minute bins. There is a pronounced increased variation in the beginning of each day, a calmer period in the middle and somewhat increased variation towards the end. This diurnal pattern has been observed by many studies for all sorts of financial returns.

The sample variance for each bin will be our estimate of diurnal variance component s_i. In the second step, returns are normalized by their respective diurnal standard deviation. In order to take account of the remaining intraday dynamics, we fit a GARCH (1,1) model to returns standardized in that way. Table 1 presents GARCH results for Valero Energy. It reveals that the persistence measure $(\alpha + \beta)$ equals 0.814, which is lower than what is usual for daily GARCH models. This, however, does not contradict temporal aggregation results of Drost and Nijman (1993), since we have previously removed the daily volatility component, responsible for a longer persistence.

Volatility components estimated for VLO are displayed in five panels combined in Figure 2. Top panel shows logarithmic returns. This is followed by the square root of the daily volatility forecast, which is the same for all bins on a given day. Third panel represents the regular diurnal pattern, and the stochastic intraday component is drawn below. We may appreciate that this component is able to modify the regular deterministic diurnal pattern. The bottom of Figure 2 draws the square root of a composite variance component, being the product of the preceding three variance components.

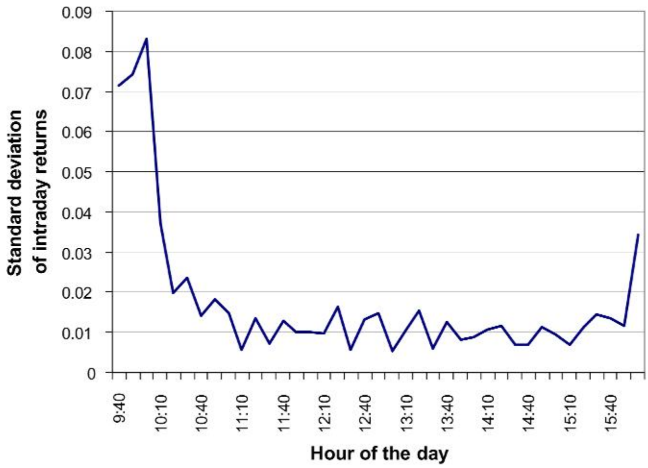

Figure 1 Standard deviation of returns across bins for the VLO stock. The horizontal axis labels denote hours during a trading day. Values depicted in this graph are calculated as a standard deviation of 10-minute returns in each bin. Returns have been previously divided by daily volatility component.

Table 1 Intraday GARCH results for the VLO stock

Parameter	Value	Standard error	*T statistic*
C	0.0065	0.0204	0.3161
ω	0.1865	0.0302	6.1835
β	0.7264	0.0387	18.7659
α	0.0876	0.0121	7.2185

This table presents estimation results for intraday GARCH(1,1) model for the VLO stock. Sample period April to May 2000. Symbols α, β, and ω denote GARCH parameters from the variance equation (7). C denotes a constant in the mean equation.

3.3 Results for a Sample of 2721 Stocks

3.3.1 Separate estimation results. Model (7) is estimated for 2721 US stock equity returns, which have been previously divided by a volatility forecast for a day and "diurnally adjusted" by the standard deviation for each bin. In order to eliminate autocorrelation present in 10-minute returns, we prewhiten (prefilter) them by fitting an ARMA(1,1) model and use the residuals in further analysis. Estimation is performed for the period April to May 2000, and the combined count of observations during this period exceeds 4.2 million data points. Since it is rather demanding to fit results of estimation for 2721 separate companies into a table

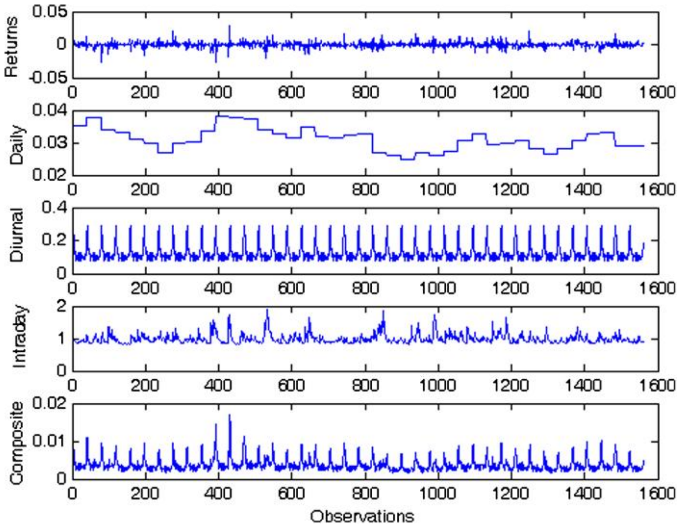

Figure 2 Volatility Components for the VLO stock. Estimation period April to May 2000. Top panel: logarithmic intraday returns on VLO stock. Second panel: the square root of the daily variance component. Third panel: the square root of the diurnal variance component. Fourth panel: the square root of the intraday variance component. Fifth panel: the square root of the composite variance component being the product of the proceeding three variance components.

of a manageable size, we report results of this procedure resorting to graphical methods. Figure 3 shows parameter values for companies sorted by their trading intensity. By a "GARCH parameter" and an "ARCH parameter", we refer to β and α coefficients from equation (7). The top and middle panels of Figure 3 depict β and α parameters, respectively. The bottom panel plots the sum of both parameters, thus informing us how persistent the volatility is. Figure 3 offers a histogram of this measure of persistence $(\beta + \alpha)$. In both figures, we may observe a fair amount of variation in the values of parameters and the measure of persistence. For the purpose of this graphical illustration, the companies are sorted according to their trading intensity. Here, we measure trading intensity by the average daily number of trades.

Companies at the right of Figure 3 are very actively traded and those at the very left seldom trade. It can be observed that the variability of the estimates decreases with the trading intensity. Further, there is an upward trend in the GARCH parameter and a downward tendency for the ARCH parameter. In fact, Figure 3 give us a rationale for grouping companies for the purpose of estimation. It turns out that for some companies, especially the least actively traded ones, separate GARCH estimation encounters difficulties, predominantly of a numerical nature. When we inspected the "troubled" companies more closely, estimation problems were usually resolved by removing one or two very influential observations (of a

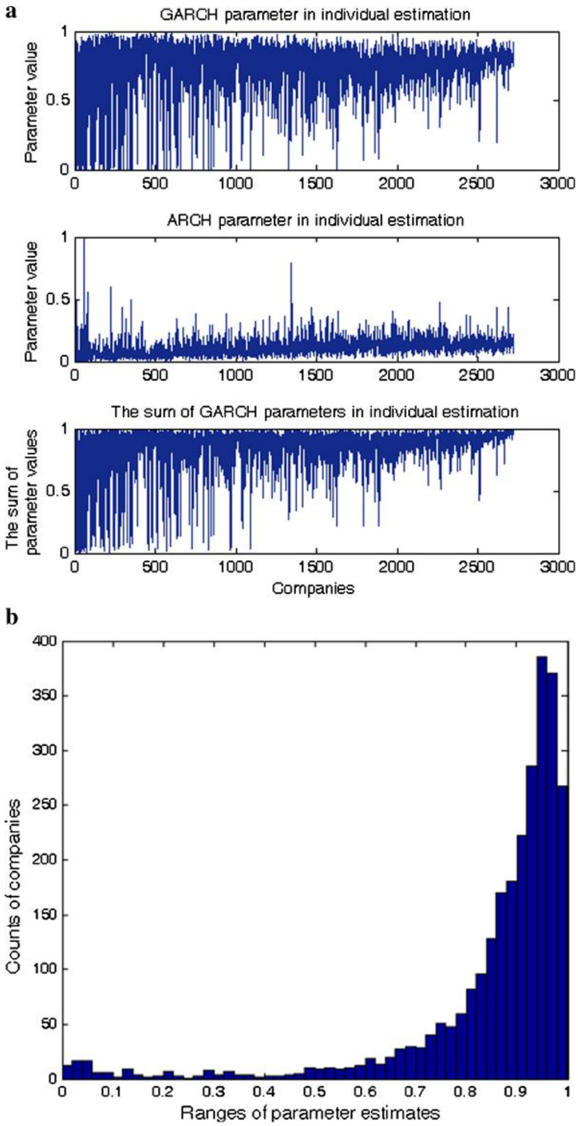

Figure 3 (a) Estimation results for the intraday GARCH models for 2721 separate companies. Sample period April to May 2000. For the purpose of this picture, companies were sorted by their average daily number of trades from low to high. Top panel: GARCH β from equation (7). Second panel: ARCH parameter α from equation (7). Third panel: persistence measure $(\alpha + \beta)$. (b) Histogram of the persistence measure $(\alpha + \beta)$ from intraday GARCH estimation for 2721 separate companies. Sample period April to May 2000. The horizontal axis denotes the value of the persistence parameter $(\alpha + \beta)$, and the vertical axis denotes the number of companies with the estimated persistence parameters falling into a corresponding bin.

magnitude of 10 standard deviations or so). This, however, seems to be a rather arbitrary procedure. When confronted with a big cross-section of companies, as in this paper, such arbitrary practice could prove very tedious and virtually impossible in real-time big scale implementation.

3.3.2 Grouped estimation results. As discussed in the previous section, Figure 3 indicates that for some companies, particularly the less liquid ones, mainly due to the widespread presence of influential or outlying observations, numerical problems with convergence are more likely. We seek to use the cross-section information to improve estimation results. We will judge the performance of particular models on the basis of their forecasting results, presented in sections following the present one.

The purpose is to group/pool companies and estimate a GARCH model for each group. Just as in the case of pooled OLS estimation, we append one series to the end of the previous one[1]. As will be shown in the forecasting section, pooled specifications outperform predictions made on the basis of stock-by-stock estimates.

An important question we need to answer is what a good criterion for grouping should be. Grouping similar companies increases the sample size and will improve accuracy. However grouping dissimilar companies will introduce bias. The way we group series could certainly influence the parameters of model (7). Although industry grouping is an obvious candidate, we have investigated a number of different admissible ways of sorting companies.

We have considered groups based on the exchange the stocks are traded on and if they are included in major indices. In particular, we obtained five groups: NYSE/NASDAQ exchange and S&P and non-S&P equities, with the remaining fifth category "Other stocks". This exercise was motivated by the finding reported by some authors (cf. Bennett and Wei 2006) documenting changing volatility levels for companies that have switched exchanges. Our five groups turned out to be very unbalanced in terms of size, and the forecast comparison seemed to be worse than the other grouping modes applied. Therefore, we do not report results of this exercise in this paper.

Liquidity is increasingly viewed as a risk factor and can be expected to influence volatility dynamics. Since there is general lack of agreement in the literature on what is the best measure of liquidity, we have examined several criteria. We have grouped companies according to their capitalization and intensity

[1]We would argue that the question of "initial values" for each pooled series does not have practical importance. There are so many observations that initial conditions do not matter. The alternative to our approach would be to treat initial values for $q_{t,i}$s as unknown parameters. This, however, implies 2721 initial values and 2721 additional parameters in the case of our biggest model. It would be very costly to estimate so many parameters. We have different numbers of parameters in each of the pooling problem. More importantly, we would not get consistent estimates because as the sample gets larger, there would be more and more parameters to estimate.

of trading measured as both the average number of trades per day and the percentage of zero returns. Capitalization grouping placed companies with visibly different volatility patterns into the same groups and it underperformed other measures in forecasting. In the rest of the paper, our favorite liquidity criterion will be the average number of trades per day. However, estimation and forecasting results were indistinguishable for the percentage of zero returns as a criterion for sorting.

In summary, we will investigate three different ways of sorting companies into groups. INDUST denotes GARCH estimation for companies grouped according to their primary industry classification. In LIQUID mode, we have grouped companies according to the average number of trades per day. The last mode (ONEBIG) involves estimation of a single large GARCH model, for all companies pooled together into one group.

In the INDUST mode, we group data into 54 industries and estimate 54, instead of 2721, intraday GARCH (1,1) models. Estimation results of this step are summarized in Table 2. We use the industry classification of the daily volatility forecast vendor. It turns out that four of six industries for which the LM(20) test is significant are technology/internet sectors. These are electrical equipment, semiconductors, computer software and the Internet. Our sample (April to June 2000) covered the aftermath of the Internet bubble. Attempts to estimate higher order GARCH models for these industries did not improve the LM(20) results considerably.

What is important, in INDUST mode we do not encounter any convergence problems as was the case for some stocks in individual estimation. The persistence parameter for most industries falls in the range of 0.86–0.96, and the minimum value is 0.761. Similarly to Table 1, the persistence values are lower than it is customary for daily GARCH models. This, however, does not contradict temporal aggregation results of Drost and Nijman (1993), since the daily volatility component is expected to account for a longer persistence.

Next, we estimate GARCH models for 50 groups of companies sorted according to the liquidity criterion. The number of groups was chosen to make results approximately comparable with the industry estimation. Table 3 and Figure 4 indicate an increase in persistence measure for more liquid groups. We also notice a further reduction in intraday GARCH parameter variation between groups. A somewhat disappointing result emerging from Table 3 is that most of the actively traded groups produce GARCH residuals that show statistically significant volatility clustering (as indicated by ARCH LM(1) and LM(20) tests). We think that it does not invalidate our emphasis on using GARCH(1,1) as a base model. As will be seen in Section 4.2 on forecasting results, liquidity-based grouping works best for illiquid stocks. And indeed for the illiquid groups (at the top of Table 3), the LM test does not appear to be significant at the 5% level.

The histogram in Figure 4 seemingly documents a reduction of the range of intraday persistence measure with one notable exception. The least liquid group comprising 55 companies has a persistence parameter equal to 0.575. This group

Table 2 Industry sorting estimation results—INDUST mode

Industry	Skewness coefficient	Kurtosis coefficient	No. of observations	$\alpha+\beta$	ω	ω *t*-stat	β	β *t*-stat	α	α *t*-stat	LM(1)	LM(20)
1	−0.17	19.6	92970	0.863	0.139	105.9	0.797	476.0	0.065	105.2	2.1	16.0
2	0.09	21.6	24100	0.761	0.244	62.0	0.663	130.6	0.098	56.7	0.0	4.5
3	−0.33	22.3	71292	0.861	0.142	131.4	0.787	532.0	0.074	92.3	0.5	10.7
4	0.12	18.4	134195	0.889	0.113	123.1	0.825	639.9	0.064	108.1	2.5	13.2
5	0.15	14.9	100579	0.896	0.106	113.8	0.829	634.0	0.067	95.2	0.4	13.0
6	−0.01	11.9	43641	0.906	0.095	50.0	0.849	325.3	0.058	51.4	0.4	4.6
7	−0.22	12.9	55380	0.892	0.111	63.0	0.795	288.6	0.097	60.3	0.9	38.7**
8	−0.07	16.8	102608	0.844	0.159	107.2	0.751	354.8	0.092	93.2	0.5	17.2
9	−0.04	30.3	23673	0.892	0.116	75.1	0.812	351.9	0.079	64.1	0.3	4.2
10	0.61	21.6	10647	0.858	0.150	48.0	0.757	179.4	0.101	36.2	0.4	6.2
11	−0.09	16.2	37322	0.836	0.167	66.1	0.743	218.8	0.093	78.3	0.1	7.1
12	−0.38	32.0	28236	0.766	0.235	60.6	0.680	133.9	0.086	44.6	0.0	3.7
13	−0.19	19.4	33540	0.819	0.185	53.2	0.708	132.8	0.111	42.9	0.0	2.2
14	−0.26	15.9	56705	0.909	0.093	77.5	0.839	452.9	0.071	73.9	1.1	9.8
15	−0.13	17.3	44927	0.911	0.091	88.4	0.818	461.0	0.093	78.9	4.1*	15.6
16	−0.01	13.1	37401	0.894	0.110	57.1	0.790	242.5	0.104	56.4	1.4	11.4
17	−0.27	16.5	90714	0.892	0.112	137.6	0.789	544.5	0.103	98.5	0.3	6.6
18	0.19	12.6	24959	0.874	0.129	54.3	0.770	216.1	0.104	49.7	0.6	13.2
19	0.21	22.1	84239	0.922	0.079	117.7	0.865	810.5	0.057	99.8	0.7	6.0
20	0.01	15.3	57330	0.928	0.075	78.5	0.863	568.6	0.065	72.9	1.9	19.4
21	−0.33	19.4	103229	0.909	0.094	142.1	0.832	807.2	0.076	110.4	0.4	13.7

(continued)

Table 2 (*continued*)

Industry	Skewness coefficient	Kurtosis coefficient	No. of observations	$\alpha + \beta$	ω	ω t-stat	β	β t-stat	α	α t-stat	LM(1)	LM(20)
22	0.21	12.1	28119	0.866	0.137	40.7	0.791	176.8	0.075	41.9	0.5	6.9
23	0.20	12.9	37401	0.797	0.208	49.2	0.679	124.4	0.119	59.2	0.7	11.1
24	−0.06	18.0	25583	0.929	0.073	65.4	0.867	516.7	0.061	55.2	0.5	9.7
25	−0.37	16.0	48163	0.911	0.090	86.8	0.839	529.8	0.072	69.5	3.5	14.2
26	0.22	7.7	10919	0.918	0.083	19.1	0.826	123.0	0.092	24.6	1.9	9.0
27	−0.11	14.0	42119	0.879	0.124	62.0	0.785	242.4	0.094	55.2	0.2	4.5
28	−0.33	16.5	17121	0.816	0.185	31.8	0.743	100.3	0.073	31.2	0.1	2.5
29	0.11	22.2	103542	0.910	0.093	173.4	0.845	1040.1	0.065	126.9	0.8	13.0
30	0.15	10.0	119105	0.865	0.136	121.4	0.786	517.5	0.079	90.0	8.2**	40.6**
31	0.18	13.8	58733	0.914	0.087	72.6	0.857	505.5	0.057	70.1	0.8	9.3
32	0.28	11.9	15600	0.912	0.093	25.2	0.807	136.5	0.106	30.5	0.0	9.0
33	0.14	19.3	31628	0.819	0.187	77.5	0.707	206.9	0.112	58.0	0.1	6.9
34	−0.11	20.5	44691	0.909	0.096	115.0	0.816	590.0	0.094	83.5	3.3	11.5
35	0.11	12.1	51400	0.914	0.090	60.8	0.809	350.2	0.105	75.5	12.5**	23.1
36	0.20	18.1	108809	0.904	0.101	121.8	0.808	629.2	0.095	122.4	4.1*	18.7
37	−0.19	16.7	214610	0.919	0.086	160.4	0.810	985.7	0.109	198.7	2.2	15.9
38	−0.03	13.3	222297	0.920	0.083	186.0	0.810	943.2	0.110	164.7	3.4	44.0**
39	0.02	10.5	148821	0.944	0.060	108.0	0.822	755.4	0.122	128.8	10.1**	69.1**
40	0.04	14.7	125260	0.918	0.084	178.5	0.817	871.1	0.100	130.7	0.5	7.4
41	−0.13	13.0	258050	0.933	0.072	210.5	0.813	1319.2	0.120	211.7	19.8**	82.7**

(*continued*)

Table 2 (*continued*)

Industry	Skewness coefficient	Kurtosis coefficient	No. of observations	$\alpha + \beta$	ω	ω t-stat	β	β t-stat	α	α t-stat	LM(1)	LM(20)
42	0.21	13.3	34240	0.903	0.100	54.5	0.814	285.0	0.089	53.5	3.8	12.4
43	−0.18	15.3	113017	0.907	0.095	124.0	0.818	622.6	0.089	105.5	0.6	14.0
44	−0.37	19.1	50504	0.889	0.115	120.6	0.779	464.7	0.110	84.7	0.0	3.7
45	−0.18	15.8	174678	0.922	0.082	178.0	0.820	937.9	0.102	152.7	3.0	25.7
46	−0.06	17.6	85601	0.919	0.083	114.8	0.845	765.2	0.074	116.6	5.0*	19.9
47	0.08	21.5	34865	0.896	0.109	87.2	0.802	428.8	0.094	84.2	5.7*	17.6
48	−0.05	20.1	96524	0.930	0.072	162.9	0.869	1462.3	0.060	123.0	12.0 **	22.1
49	−0.07	21.4	218205	0.905	0.097	176.9	0.845	1042.4	0.060	144.9	3.1	15.1
50	−0.15	19.4	48044	0.892	0.110	59.7	0.839	348.0	0.053	60.0	0.4	7.6
51	−0.19	15.2	57679	0.910	0.094	114.3	0.808	540.9	0.102	88.1	0.3	8.6
52	−0.39	18.9	87902	0.944	0.058	148.1	0.880	1358.0	0.064	127.4	4.9*	9.2
53	−0.06	11.2	79754	0.956	0.050	73.2	0.822	514.5	0.133	97.5	7.8 **	39.1 **
54	0.06	13.4	156155	0.906	0.096	94.1	0.850	607.8	0.056	101.1	0.9	15.5

This table presents estimation results for intraday GARCH(1,1) models for 54 industries. Sample period April to May 2000. Symbols α, β, and ω denote GARCH parameters from the variance equation (7). Persistence is measured as the sum of parameters $(\alpha + \beta)$. LM(1) and LM(20) statistics are calculated as the ARCH LM test, cf. Engle (1982), on the residuals from (8). Under the null of no ARCH effects at lag q, the statistic has a χ^2 distribution with q degrees of freedom, where $q = 1, 20$. * and ** denote significance at the 5% and 1% levels, respectively.

Table 3 Liquidity sorting estimation results LIQUID mode

Group	Skewness coefficient	Kurtosis coefficient	$\alpha+\beta$	ω	ω t-stat	β	β t-stat	α	α t-stat	LM(1)	LM(20)
1	−1.05	133.9	0.575	0.439	136.1	0.514	146.3	0.061	66.1	0.3	2.1
2	−0.48	66.6	0.916	0.085	191.6	0.886	1655.6	0.030	148.1	0.6	4.5
3	−0.45	41.1	0.926	0.075	124.7	0.892	1145.5	0.034	116.7	1.0	8.7
4	−0.07	28.9	0.925	0.076	100.3	0.893	934.5	0.032	99.4	5.5*	14.6
5	−0.16	29.2	0.917	0.085	113.7	0.869	866.7	0.047	110.5	1.0	4.7
6	0.07	24.8	0.887	0.115	118.1	0.826	654.5	0.061	124.7	2.0	8.1
7	−0.02	21.5	0.918	0.084	93.9	0.866	714.3	0.051	102.4	1.7	10.1
8	−0.06	20.2	0.928	0.074	103.6	0.876	889.0	0.052	114.1	4.7*	16.0
9	0.04	15.7	0.946	0.055	75.0	0.907	905.9	0.039	89.5	9.7**	25.6
10	−0.19	15.5	0.933	0.068	82.2	0.883	774.1	0.050	90.7	6.8**	22.8
11	0.08	15.6	0.918	0.083	81.9	0.864	624.6	0.054	87.8	5.2*	18.2
12	0.03	17.0	0.873	0.129	87.6	0.804	398.2	0.069	82.5	1.4	17.8
13	−0.11	16.6	0.909	0.093	83.8	0.830	511.9	0.079	110.5	0.6	11.3
14	0.09	14.0	0.928	0.074	94.4	0.859	760.1	0.069	98.0	16.7**	38.5**
15	−0.14	14.1	0.905	0.097	75.8	0.833	448.2	0.072	81.8	1.8	11.7
16	0.10	13.6	0.931	0.071	86.0	0.863	690.1	0.068	93.0	9.8**	25.9
17	−0.05	12.4	0.915	0.089	74.2	0.832	478.3	0.083	85.9	8.0**	30.8
18	−0.10	14.1	0.916	0.087	77.8	0.829	496.1	0.087	97.6	3.8	21.7
19	−0.10	15.1	0.928	0.075	107.0	0.832	651.4	0.096	107.6	3.1	20.2
20	−0.03	11.8	0.934	0.069	80.6	0.855	600.0	0.078	84.1	9.6**	31.7*

(continued)

Table 3 (*continued*)

Group	Skewness coefficient	Kurtosis coefficient	$\alpha + \beta$	ω	ω t-stat	β	β t-stat	α	α t-stat	LM(1)	LM(20)
21	−0.22	14.8	0.906	0.099	78.8	0.802	427.4	0.104	89.7	8.4**	28.1
22	−0.16	12.2	0.901	0.103	80.7	0.800	401.6	0.101	86.9	1.9	15.1
23	−0.01	10.8	0.903	0.101	71.6	0.804	376.0	0.099	82.0	9.0**	44.9**
24	0.09	11.2	0.898	0.105	76.6	0.798	388.9	0.100	81.0	13.6**	41.0**
25	−0.09	13.3	0.930	0.075	98.4	0.834	630.5	0.095	92.8	14.5**	38.0**
26	0.04	10.8	0.921	0.083	72.6	0.823	441.9	0.098	82.4	14.5**	41.9**
27	0.05	10.9	0.925	0.080	73.2	0.815	448.5	0.110	87.6	14.3**	48.5**
28	−0.09	13.4	0.909	0.098	78.1	0.785	379.7	0.123	93.0	2.4	25.4
29	0.08	10.4	0.908	0.095	67.5	0.799	352.1	0.110	79.0	5.2*	32.0*
30	−0.08	11.5	0.899	0.105	86.9	0.787	401.6	0.112	88.9	9.7**	31.5*
31	−0.11	11.0	0.922	0.083	73.6	0.811	436.3	0.111	91.6	14.5**	60.8**
32	−0.12	12.4	0.920	0.085	76.2	0.808	446.4	0.112	97.6	7.3**	33.7*
33	0.02	10.6	0.889	0.115	73.8	0.768	316.4	0.121	83.0	12.4**	47.9**
34	−0.11	9.5	0.894	0.111	72.9	0.761	302.3	0.133	79.5	4.7*	43.1**
35	0.01	9.7	0.904	0.100	69.8	0.786	312.7	0.118	73.0	7.9**	38.2**
36	−0.12	10.5	0.900	0.105	63.8	0.768	280.0	0.131	76.2	6.2*	34.4*
37	0.03	10.0	0.908	0.098	68.6	0.774	317.9	0.134	83.9	5.7*	30.6

(*continued*)

Table 3 (*continued*)

Group	Skewness coefficient	Kurtosis coefficient	$\alpha+\beta$	ω	ω t-stat	β	β t-stat	α	α t-stat	LM(1)	LM(20)
38	−0.04	9.7	0.915	0.090	77.5	0.791	425.2	0.124	91.5	6.4*	42.3**
39	0.01	8.5	0.908	0.097	62.3	0.782	307.0	0.126	74.7	11.6**	46.0**
40	−0.02	8.9	0.922	0.082	64.7	0.804	366.2	0.119	76.0	10.2**	52.1**
41	−0.07	9.4	0.924	0.081	73.6	0.797	393.2	0.126	81.9	5.4*	46.9**
42	−0.02	9.4	0.919	0.087	87.8	0.784	454.4	0.134	90.5	8.2**	59.8**
43	−0.07	8.9	0.931	0.072	61.2	0.816	390.1	0.116	78.3	20.8**	74.9**
44	−0.11	10.6	0.936	0.070	71.2	0.810	426.8	0.126	85.7	6.7**	40.3**
45	−0.05	9.3	0.934	0.070	60.0	0.808	380.1	0.126	86.4	10.8**	47.4**
46	−0.04	9.0	0.919	0.084	71.0	0.790	330.6	0.130	76.3	13.2**	49.0**
47	0.03	9.2	0.948	0.055	53.9	0.828	392.6	0.120	71.1	37.9**	86.8**
48	0.07	9.3	0.959	0.046	63.8	0.837	545.7	0.121	82.0	10.5**	44.9**
49	0.12	9.4	0.954	0.050	55.9	0.823	418.9	0.131	79.8	7.7**	31.1
50	0.27	9.9	0.964	0.040	66.9	0.802	389.7	0.162	86.1	1.6	36.9*

This table presents estimation results for intraday GARCH(1,1) models for 50 groups of liquidity sorted companies. Sample period April to May 2000. Symbols α , β, and ω denote GARCH parameters from the variance equation (7). Persistence is measured as the sum of parameters $(\alpha+\beta)$. LM(1) and LM(20) statistics are calculated as the ARCH LM test, cf. Engle (1982), on the residuals from equation (8). Under the null of no ARCH effects at lag q, the statistic has a χ^2 distribution with q degrees of freedom, where $q = 1, 20$. * and ** denote significance at the 5% and 1% levels, respectively.

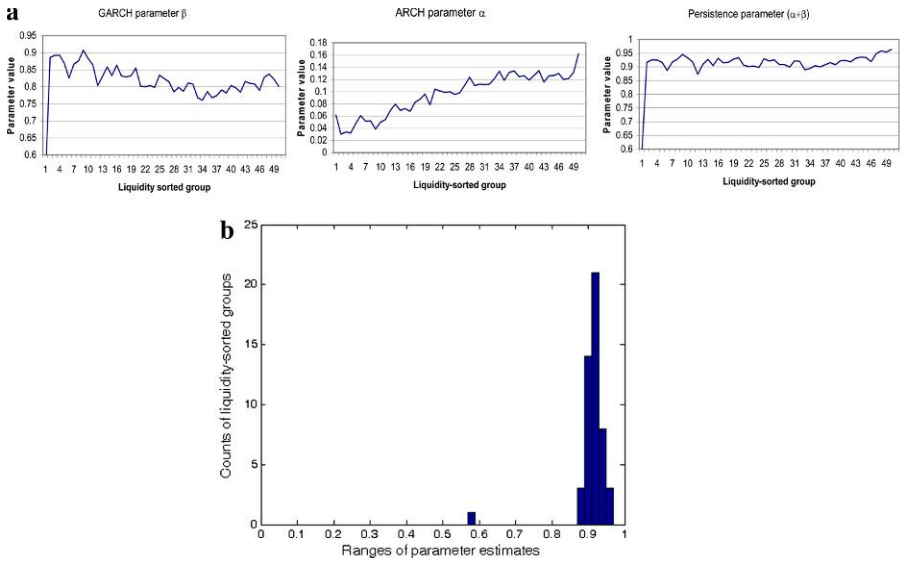

Figure 4 Estimation results for the intraday GARCH models for 50 liquidity-sorted groups. Trading intensity or liquidity increases from the left to the right side of each picture. Sample period April to May 2000. First panel: GARCH parameter β from equation (7). Second panel: ARCH parameter α from equation (7). Third panel: persistence measure $(\alpha + \beta)$. (b) Histogram of the persistence measure $(\alpha + \beta)$ from intraday GARCH estimation for 50 liquidity-sorted groups. Sample period April to May 2000. The horizontal axis denotes the value of the persistence parameter $(\alpha + \beta)$, and the vertical axis denotes the number of companies with the estimated persistence parameters falling into a corresponding bin.

is characterized by spectacular kurtosis and skewness coefficients. Figure 5 gives a snapshot of 10-minute returns on the least traded stocks and medium traded stocks. The bulk of observations in the least traded group are equal to zero, and many nonzero observations could be described as outlying or "influential" because there are several standard deviations from the mean.

Finally, the upper panel of Table 4 reports results for GARCH estimation for the intraday component for one giant pool of all the companies, comprising over 4.2 million observations. As in the industry case, this table indicates modest persistence of intraday volatility.

3.3.3 Asymmetric models. A number of studies have found superior forecasting performance of volatility models that incorporate the asymmetric influence of negative and positive lagged returns. In the following, we will also refer to this phenomenon as the "leverage effect". Early contributions included Nelson (1991), Engle and Ng (1993) and Glosten, Jagannathan, and Runkle (1993).

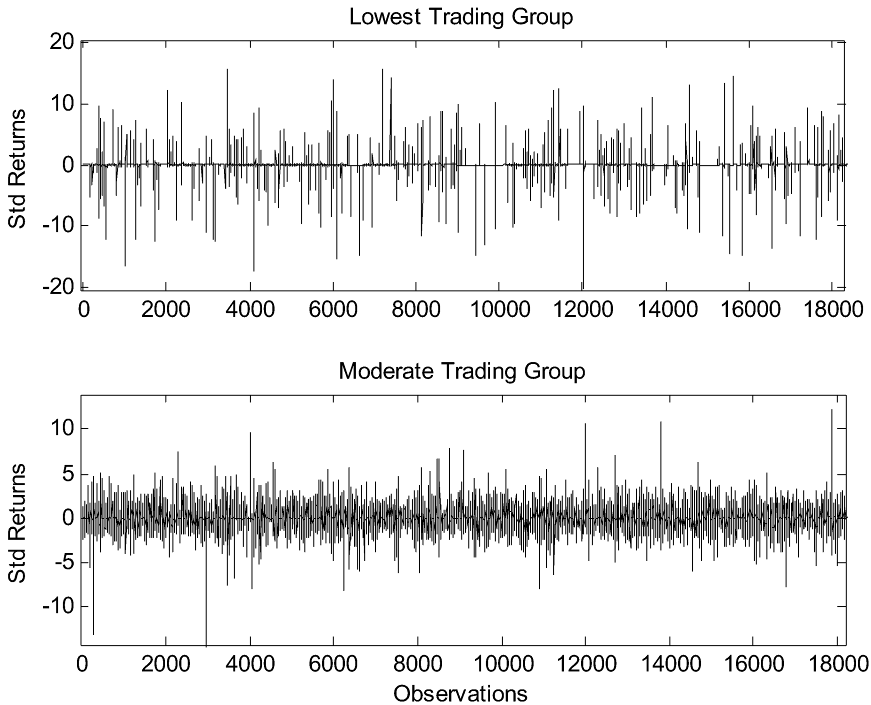

Figure 5 Examples of standardized logarithmic returns for two of the 50 liquidity-sorted groups. Horizontal axis denotes ith observation in each liquidity-sorted group and snapshots were chosen at random from the upper half of the groups.

Recently, several papers have turned attention to the importance of asymmetric models for intraday returns. Chen and Ghysels (2011) offer a framework for forecasting daily volatility using past intraday returns and report that asymmetric models outperform their symmetric counterparts for spot and futures S&P and Dow Jones indices. Chen, Ghysels, and Wang (2010, 2011) use multiplicative GARCH models to investigate different asymmetric and periodic structures at the intraday level and find that asymmetric models forecast volatility of stock index returns better than symmetric specifications.

Therefore, another specification for the intraday stochastic variance component that we examine is the asymmetric GJR model (Glosten, Jagannathan, and Runkle 1993)

$$q_{t,i} = \omega + \alpha z_{t,i-1}^2 + \gamma z_{t,i-1}^2 d_{t,i-1} + \beta\, q_{t,i-1}. \tag{14}$$

We define $d_{t,i}$ as a dummy variable which is 1 for bins with negative intraday returns. We choose equation (14) to investigate the leverage effect because we work with a large sample and the speed of estimation was of a great concern to us. If the leverage effect matters, the coefficient γ is expected to be positive and

Table 4 All sample estimation results- ONEBIG mode GARCH(1,1) and GJR models

Parameter	Value	Standard error	T statistic
GARCH(1,1)			
ω	0.096	0.000112	854.3
β	0.823	0.000180	4570.4
α	0.084	0.000115	728.6
GJR model			
ω	0.094	0.000109	857.46
β	0.826	0.000176	4703.58
α	0.073	0.000132	550.17
γ	0.0187	0.000178	104.96

This table presents estimation results for intraday GARCH(1,1) model (upper panel) and GJR model (lower panel) for one large group of companies pooled together. Sample period April to May 2000. Symbols α, β, and ω denote GARCH parameters from the variance equation (7). For GJR model, α, β, ω, and γ denote parameters from the variance equation (14).

statistically significant. This means that a recent negative return increases volatility more than a positive return of the same magnitude. Estimation results of model (14) for each stock separately reveal that the estimate of parameter γ had a wrong (negative) sign in the case of approximately 1000 companies. There are 1309 companies that have a statistically significant positive leverage coefficient—this is slightly less than half the total. We have also estimated leverage effects for pooled models. The lower panel of Table 4 displays results of model (14) estimated for all companies pooled together. The leverage parameter has the correct positive sign and is statistically significant. Overall, as will be seen in Section 4.2, forecasting results give a small advantage to GARCH(1,1) over GJR. It is only in the case of company-by-company individual estimation, however, that this forecasting advantage of GARCH(1,1) over GJR seems substantive.

4 FORECASTING RESULTS

4.1 Loss Functions and Design

We now turn to out of sample forecast accuracy. We use the parameter estimates for the period April to May 2000 and forecast one-step-ahead volatilities for each bin in June 2000. Forecasts are obtained in a sequential procedure on the basis of estimated parameters and the volatility forecast calculated at the previous bin, as well as actual returns from the previous bin. From the structure of the model, forecasts of the variance of returns are the product of the daily variance forecast, the diurnal variance, and the GARCH variance. In this analysis, the variance that is forecast is of the return deflated by the daily volatility and the diurnal volatility. In the forecast period, the daily volatility is taken from the same commercial source as for the estimation period.

It should be appreciated that forecasting volatility is connected with an additional complication since of course we do not observe the variable we want to forecast. In our forecasting evaluation, we will compare our forecasts with the squared return $z_{t,i}^2 = r_{t,i}^2/\hat{h}_t\hat{s}_i$. This return is a random variable drawn from a distribution with a variance we are trying to estimate. We expect that the squared return will be large only when the true variance is large; however, the squared return may be small even when the variance is large. As a consequence, it is not at all clear what a sensible loss function should be. For recent discussions of forecast accuracy measures, see Granger (2003) or Patton (2006).

In the following, we use two loss functions: the out-of-sample likelihood (LIK)

$$L_{1\{t,i\}} = \log q_{t,i} + \frac{z_{t,i}^2}{q_{t,i}^f}$$

and the mean squared error (MSE)

$$L_{2\{t,i\}} = (z_{t,i}^2 - q_{t,i}^f)^2$$

to evaluate forecast accuracy. The use of squared return or RV measure in place of the true volatility introduces biases in many popular loss functions. However, under MSE and LIK loss functions, optimal forecasts are unbiased (cf. Patton 2006; Hansen and Lunde 2006).

Although part of the literature on assessing forecasting performance of daily models (cf. Hansen and Lunde 2006) recommends using RV to evaluate forecast accuracy, this paper applies 10-minute squared returns. As demonstrated by Hansen and Lunde (2006) in the presence of microstructure noise, RV is a biased volatility measure and the bias increases with increasing frequency. On the other hand, loss functions MSE and LIK are consistent as long as the volatility proxy is unbiased.

Another reason for the choice of the squared return is that 10-minute interval does not allow a reliable measure of realized volatility to be estimated for many of the companies in our sample, particularly outside of the active trading periods. The more disaggregated a level we choose, the bigger percentage of zero returns we get. So for many companies and many 10-minute bins we would in fact be adding mainly zeros while calculating the RV measure.

We determine forecasts for each company separately, using parameters estimated in both separate and pooled estimations. Therefore for each time period, for each company, we obtain five different forecasts that will form the basis for a subsequent model evaluation and comparison.

4.2 Out-of-Sample Forecast Comparison

We perform five different estimations for companies pooled into groups in various ways described in Sections 3.2 and 3.3 and refer to these ways as modes. The first mode (NSTOCH) contains no stochastic component (7) at all. Mode number 2 (UNIQUE) involves no pooling, that is, we estimate unique GARCH models for separate companies. Mode number 3 (INDUST) denotes a GARCH estimation for companies grouped according to their primary industry classification. In Mode number 4 (LIQUID), we have grouped companies according to the average number of trades per day. The last mode (ONEBIG) involves estimation of a large GARCH model, for all companies pooled together to form one group. For each of these five separate estimations, we have calculated a series of forecast errors under the MSE and LIK loss functions. These forecast errors are used to calculate accuracy measurement criteria $L_j = \frac{1}{\tau N}\sum_{t=1}^{\tau}\sum_{i=1}^{N} L_{j\{t,i\}}$, where $j = 1, 2$; $\tau = 22$ and the length of the forecasting period is equal to $\tau \times N = 858$. We have calculated two forecast accuracy measures for each of the five estimations and for each company, which amounts to a total of 27,210 numbers (5 modes$\times$2 criteria$\times$2721 stocks).

Table 5 offers one way to compare the volatility forecast measures described above. For each company, we have compared performance of different modes of estimation pair by pair and calculated a percentage of times a forecast from a

Table 5 Comparison of one-period-ahead forecasts for estimation modes frequency with which the mode in a row outperforms the mode in a column

Modes	NSTOCH	UNIQUE	INDUST	LIQUID	ONEBIG
LIK loss function					
NSTOCH					
UNIQUE	0.795				
INDUST	0.849	0.595			
LIQUID	0.831	0.604	0.504		
ONEBIG	0.846	0.599	0.499	0.501	
Modes	NSTOCH	UNIQUE	INDUST	LIQUID	ONEBIG
MSE loss function					
NSTOCH					
UNIQUE	0.618				
INDUST	0.725	0.646			
LIQUID	0.694	0.596	0.428		
ONEBIG	0.738	0.661	0.555	0.623	

This table compares accuracy of one-step-ahead volatility forecasts obtained from five estimation modes. It contains relative frequency with which a forecast from an estimation described in a row outperforms a forecast from an estimation mode indicated in the column. Top panel: forecasts comparisons using LIK (out-of-sample likelihood) loss function. Second panel: forecasts comparison using MSE loss function.

given estimation outperforms each of the remaining forecasts. Hence, numbers contained in Table 5 give us the relative frequency with which the mode in the row outperforms the mode in the column for a particular loss function. We will first focus our attention on the upper panel of Table 5, which presents results of forecast comparison using the LIK loss function. For example, the third row second column compares results of NSTOCH versus separate GARCH estimations. Here, the number 0.795 means that the specification without component (7) yields worse forecasts than the individual company-by-company estimation 80% of times. The second column of the table informs us that NSTOCH estimation performs worse than all the other modes.

As we learn from column three, separate estimation gives worse forecasting results than INDUST, LIQUID, and ONEBIG modes but outperforms NSTOCH. Two remaining columns in the panel devoted to LIK evaluation criterion suggest that the aggregation methods give similar results. The above discussion carries over to the lower panel of Table 5, which presents results for the MSE loss function. One exception is that the MSE criterion seems to distinguish between methods of pooling companies and favors ONEBIG model over INDUST estimation and over LIQUID mode. Taking into account the criticism directed at the MSE loss function as being unduly influenced by a few big errors, we think that overall ONEBIG mode emerges as a winner of the forecasting comparison.

Table 6 reports the mean and median of the forecast accuracy measures calculated for each of the 2721 companies and five estimation modes. The smallest

Table 6 Mean and median of forecast accuracy measures for individual stocks

Loss function		NSTOCH	UNIQUE	INDUST	LIQUID	ONEBIG
Mean of forecasts accuracy measures						
LIK	Mean	1.0001	0.9485	0.9306	0.9271	0.9313
	Rank	5	4	2	1	3
MSE	Mean	3.4930	3.4835	3.4668	3.4661	3.4666
	Rank	5	4	3	1	2
Median of forecasts accuracy measures						
LIK	Median	1.0002	0.9543	0.9430	0.9432	0.9410
	Rank	5	4	2	3	1
MSE	Median	2.9765	2.9571	2.9495	2.9498	2.9478
	Rank	5	4	2	3	1

Top panel contains sample means of the forecast accuracy measures calculated for five estimation modes for each company separately for both loss functions. Second panel contains sample medians of the forecast accuracy measures. Rows labeled "rank" indicates the ranking of models on the basis of their mean (or median) errors. Models with smallest errors are ranked as 1, and the worst models are assigned the rank number 5.

number in a row denotes the smallest mean and median error, respectively. This table also contains the ordering that LIK and MSE criteria assign to the five estimation modes. Rank 1 denotes the best model, with the smallest error, model numbered as fifth performs the worst. Starting from the top panel, the liquidity-sorted model appears to give the smallest mean errors for both loss functions. Please note that UNIQUE GARCH outperforms the model with no stochastic intraday component. We test the differences in the value of mean estimates of forecast errors using the Diebold and Mariano (1995) tests reported in Table 7. This table presents t-values for the null hypothesis that the difference in forecast errors is zero. Note that if the model in the row forecasts worse than the model in the column, the t-ratio is negative. Column (or Row) 5 indicate that, according to the LIK loss function, liquidity sorting gives significantly better mean forecasts than the other models. The lower panel of Table 6 contains medians of the forecast accuracy measures. Here, ONEBIG model performs best, similarly to what we have concluded from Table 5.

Table 7 Forecast accuracy Diebold-Mariano test, t-values

Modes	NSTOCH	UNIQUE	INDUST	LIQUID
LIK loss function				
NSTOCH				
UNIQUE	5.8297*!			
INDUST	35.772*!	2.074*		
LIQUID	40.947*!	2.459*	4.964*!	
ONEBIG	34.688*!	1.998*	−1.367	−5.351*!
MSE	NSTOCH	UNIQUE	INDUST	LIQUID
MSE loss function				
NSTOCH				
UNIQUE	3.598*!			
INDUST	22.176*!	7.404*!		
LIQUID	23.956*!	7.481*!	1.377	
ONEBIG	20.394*!	7.322*!	0.334	−0.624

This table presents t-values for the null hypothesis that the difference in forecast errors between estimation modes is not significantly different from zero. If the model in the row forecasts worse than the model in the column, the t-ratio is negative. *denotes significance at the 5% level, ! denotes significance at the 0.5% level (critical t-value = 2.81). Remarks on multiple comparisons: we are aware that due to multiple comparisons, the stated level of significance may not be equal to the actual level of significance for the Diebold and Mariano test. Hansen, Lunde, and Nason (2005) propose a simulation technique to obtain critical values for the test. Due to a large number of observations, we thought that the application of their technique is unfeasible in our case. Such procedures seem to be particularly suitable, as designed and applied by their authors, for comparison of large numbers of models. We only have five basic models to compare. We think that even when a very conservative technique like Bonferroni adjustment is applied to the results in the Table 7, the conclusions of the paper remain unaltered. Therefore, we quote significance at the 0.5% significance level (together with the nominal 5% level).

Tables 5 and 6 give somewhat conflicting answers to the question—which method of company grouping should be adopted. However, they agree that grouping is very desirable compared with separate estimation. We investigate the supposed disagreement looking at liquidity issues. We limit our attention to two separate samples of 550 most liquid and most illiquid companies. Table 8A reports forecast accuracy measures for the subsample of least liquid stocks. Please note that means of error functions are bigger for least liquid stocks than for most liquid stocks (Table 8B). Nowhere is the difference between both tails so visible as for the MSE measure. The numbers differ by an order of a magnitude, and this illustrates the sensitivity of the MSE criterion to outlying observations, more frequently haunting illiquid stocks. Similar to conclusions from Table 6, Table 8A recommends the liquidity-sorted GARCH model as a preferred forecasting tool for illiquid stocks.

The picture changes when we look at the results assembled in Table 8B, which concern the most liquid stocks. Here, the ONEBIG GARCH solution, followed closely by industry grouping, outperform both liquidity-sorted models and separate estimation. These conclusions closely resemble results reported in Table 5.

Table 9 contains relative frequency with which forecasts from GARCH (1,1) outperform GJR forecasts for the corresponding mode of company grouping.

Table 8 Forecast accuracy comparison for most and least liquid stocks

A. Least liquid stocks

	Forecast accuracy measures	LIK	Rank	MSE	Rank	Average rank
NSTOCH	No stochastic intraday component	1.0002	5	5.8425	2	3.5
UNIQUE	Separate GARCH estimation	0.9779	4	5.8871	5	4.5
INDUST	Industry GARCH estimation	0.9548	2	5.8481	3	2.5
LIQUID	Liquidity-sorted GARCH estimation	0.9405	1	5.8287	1	1
ONEBIG	One large GARCH estimation	0.9629	3	5.8541	4	3.5

B. Most liquid stocks

	Forecast accuracy measures	LIK	Rank	MSE	Rank	Average rank
NSTOCH	No stochastic intraday component	0.9999	5	2.7439	5	5
UNIQUE	Separate GARCH estimation	0.9679	4	2.7309	4	4
INDUST	Industry GARCH estimation	0.9293	2	2.7084	2	2
LIQUID	Liquidity-sorted GARCH estimation	0.9339	3	2.7172	3	3
ONEBIG	One large GARCH estimation	0.9274	1	2.7044	1	1

This table reports forecast accuracy measures for the subsample of least liquid stocks (top panel) and most liquid stocks (second panel). Rank denotes ordering from best (1) to worst (4). Average rank is calculated as the mean of ordering measures in each row. LIK and MSE are two loss functions used.

Table 9 Comparison of one-period-ahead forecasts for GARCH (1,1) and GJR estimation.

Modes	LIK	MSE
UNIQUE	0.6108	0.5825
INDUST	0.5454	0.5156
LIQUID	0.5344	0.513
ONEBIG	0.5035	0.563

This table compares accuracy of one-step-ahead volatility forecasts obtained from four estimation modes using LIK and MSE loss functions. It contains relative frequency with which a forecast from GARCH (1,1) outperforms a GJR forecast.

GARCH (1,1) performs better than GJR model for all modes; however, the difference seems to be substantial for separate company-by-company estimation.

To sum up, we have seen that inclusion of a stochastic intraday component (7) improves forecasting results in comparison with the model with diurnal and daily components only. We also observe better forecasting performance using cross-section information and applying different methods of pooling. The exact way we chose to group companies heavily relies on company characteristics, liquidity in this case. Liquid stocks seem to benefit from the widest pooling possible, that is, using all available cross-section information. Illiquid stocks apparently exhibit intraday dynamics which are idiosyncratic to their particular liquidity-determined group, and consequently, we may obtain better forecasts when we group these stocks together.

5 CONCLUSIONS

This paper proposes a new intraday volatility forecasting model. Conditional variance of asset returns is expressed as a product of daily, diurnal, and stochastic intraday volatility components. This model is applied to a comprehensive sample of 10-minute returns on more than 2500 US equities. The addition of a new stochastic intraday component gives better volatility forecasts than the benchmark models.

Empirical analysis of such a broad sample allows us to draw a number of interesting conclusions. We build models for separate companies, pool data into industries and consider various criteria for grouping returns. Both the estimation and forecasting results show clear benefits of aggregation. It turns out that results for the pooled regressions seem to be more stable. The forecasts from the pooled specifications outperform the corresponding forecasts from company-by-company estimation. In general, the broadest cross-section information produces superior forecasts. For less liquid stocks, however, we obtain better forecasts when we group less frequently traded companies together.

The model and the empirical results are of substantial practical importance for the finance industry since recent years have seen greatly increased volume of computerized trading which depends on volatility forecasts. The availability of reliable intradaily forecasts can be expected to play a central role in risk assessment, as well as evaluation of trading and market making strategies.

REFERENCES

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. DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer-Run Dependencies. *Journal of Finance* 53: 219–265.

Bennett, P., and L. Wei. 2006. Market Structure, Fragmentation and Market Quality —Evidence from Recent Listing Switches. *Journal of Financial Markets* 9: 49–78.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Bollerslev, T., R. Chou, and K. Kroner. 1992. ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics* 52: 5–59.

Chen, X., and E. Ghysels. 2011. News—Good or Bad—and Its Impact on Volatility Predictions Over Multiple Horizons. *Review of Financial Studies* 24: 46–81.

Chen, X., E. Ghysels, and F. Wang. 2010. "HYBRID-GARCH: A Generic Class of Models for Volatility Predictions Using High Frequency Data." Unpublished manuscript.

Chen, X., E. Ghysels, and F. Wang. 2011. HYBRID GARCH Models and Intra-daily Return Periodicity. *Journal of Time Series Econometrics* 3(1).

Deo, R., C. Hurvich, and Y. Lu. 2005. Forecasting Realized Volatility Using A Long Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment. *Journal of Econometrics* 131: 29–58.

Diebold, F., and R. Mariano. 1995. Comparing Forecasting Accuracy. *Journal of Business and Economic Statistics* 13: 253–63.

Drost, F. C., and T. E. Nijman. 1993. Temporal Aggregation of GARCH Processes. *Econometrica* 61: 909–927.

Ellul, A., C. W. Holden, P. Jain, and R. Jennings. 2007. Order Dynamics: Recent Evidence from the NYSE. *Journal of Empirical Finance* 14: 636–661.

Engle, R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50: 987–1007.

Engle, R. F., and R. Ferstenberg. 2007. Execution Risk. *The Journal of Portfolio Management* 33: 34–44.

Engle, R.F., and G.P. Gallo. 2006. A Multiple Indicators Model for Volatility Using Intra-daily Data. *Journal of Econometrics* 131: 3–27.

Engle, R.F., and V. Ng. 1993. Measuring and Testing the Impact of News on Volatility. *Journal of Finance* 48: 1749–1778.

Fabozzi, F., F. Jones, and R. Vardharaj. 2002. "Multi-Factor Risk Models." In F. Fabozzi and H. Markowitz (eds.), *The Theory and Practice of Investment Management*. Hoboken, NJ: John Wiley and Sons.

Ghose, D., and K. Kroner. 1996. "Components of Volatility in Foreign Exchange Markets: An Empirical Analysis of High Frequency Data." Unpublished manuscript, Department of Economics, University of Arizona.

Giot, P., 2005. Market Risk Models for Intraday Data. *European Journal of Finance* 11: 309–324.

Glosten, L., R. Jagannathan, and D. Runkle. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* 48: 1779–1801

Granger, C., 2003. *Empirical Modeling in Economics. Specification and Evaluation*. Cambridge: Cambridge University Press.

Griffiths, M., B. Smith, D. Turnbull, and R. White. 2000. The Costs and the Determinants of Order Aggressiveness. *Journal of Financial Economics* 56: 65–88.

Hansen, L.P. 1982. Large Sample Properties of Generalized Methods of Moments Estimators. *Econometrica* 50: 1029–1054.

Hansen, P. R., and A. Lunde. 2006. Consistent Ranking of Volatility Models. *Journal of Econometrics* 131: 97–121.

Hansen, P. R., A. Lunde, and J. Nason. 2005. "Model Confidence Sets for Forecasting Models." Working Paper 2005–07. Federal Reserve Bank of Atlanta.

Lee, S.-W., and B. Hansen. 1994. Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory* 10: 29–53.

Lumsdaine, R. 1996. Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models. *Econometrica* 64: 575–596.

Mueller, H., R. Sen, and U. Stadtmuller. 2007. "Functional Data Analysis for Volatility". Unpublished manuscript.

Nelson, D., 1990. ARCH Models as Diffusion Approximations. *Journal of Econometrics* 45: 7–38.

Nelson, D., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59: 347–370.

Newey, W.K., and D. McFadden. 1994. "Large Sample estimation and Hypothesis Testing." In D. McFadden and R. Engle (eds.) *Handbook of Econometrics*, vol. 4. Amsterdam: Elsevier.

Patton, A. 2006. "Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies." Working Paper. London School of Economics; forthcoming in *Journal of Econometrics*.

Taylor S. J., and X. Xu. 1997. The Incremental Volatility Information in Million Foreign Exchange Quotations. *Journal of Empirical Finance* 4: 317–340.