High Frequency Multiplicative Component GARCH[♣*]

Robert F. Engle[*], Magdalena E. Sokalska[**] and Ananda Chanda[***]

August 2, 2005

Abstract

This paper proposes a new way of modeling and forecasting intraday returns. We decompose the volatility of high frequency asset returns into components that 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. Apart from building a new model, we obtain several interesting forecasting results. In particular, it turns out that forecasts obtained from the pooled cross section of companies seem to outperform the corresponding forecasts from company-by-company estimation.

JEL Classifications: C22, C51, C53, G15
Keywords: Volatility, ARCH, Intra-day Returns.

[♣*] 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. We thank the participants of the Stern School of Business Finance seminar, Morgan Stanley Equity Market Research Conference, the "Frontiers in Time Series Analysis" conference in Olbia, the "Changing Structures in International and Financial Markets' conference in Venice and the SCE2005 meeting, for helpful comments and suggestions.

[*] Department of Finance, Stern School of Business, New York University, rengle@stern.nyu.edu.

[**] Department of Finance, Stern School of Business, New York University, msokalsk@stern.nyu.edu and Warsaw School of Economics msokal@sgh.waw.pl.

[***] Morgan Stanley, Ananda.Chanda@morganstanley.com.

1. INTRODUCTION

This paper proposes a new way of modeling and forecasting intraday returns. We decompose 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 build 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.

Conventional GARCH 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. In response see Ghose and Kroner (1996). Alternatively, Andersen and Bollerslev (1997, 1998) propose models for 5-minute returns on Deutschemark-dollar exchange rate and the S&P500 index. In the first paper, Andersen and Bollerslev build a multiplicative model of daily and diurnal volatility, and in the second they add an additional component which takes account of macroeconomic announcements. We think that adding a collection of dummy variables representing macroeconomic news announcements to model volatility of US stocks is not very practical. Most important macroeconomic announcements happen before the stock market opens. Secondly, it is difficult to imagine that the strength and pattern of market's response to news would be the same, even for the same type of events across time, disregarding the fact how much the news has been anticipated. Third, idiosyncratic announcements can be expected to be particularly important for equities, and the timing of the majority of them is very difficult to forecast. For most of Andersen and Bollerslev's models, the intra-daily volatility components are deterministic. In contrast, the intra-daily components in our model are both deterministic (the diurnal) and stochastic (a separate intra-daily ARCH).

Overall, in distinction to the huge volume of literature on daily volatility models, the research on intraday volatility has been by far less studied. 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. Giot (2005) examines a number of market risk models for intraday 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 the risk of slow trading (Engle, 2005) 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.* 2003, Griffith *et al.*, 2000). These strategies are implemented on intraday basis. Ranaldo (2004) documents that high volatility increases the probability of submission of limit orders, whereas Ellul *et al*. (2003) find an increased order submission probability for all order types in volatile periods. According to Lo *et al.* (2002), limit orders execute more quickly when markets become more volatile (the authors use the logarithm of the number of trades in the previous hour as a proxy for volatility). Hasbrouck and Saar (2002) investigate the effects of volatility on the cross-section of companies traded on Island ECN. They find that increased volatility reduces the expected time to execution, which they label as a mechanistic volatility effect. These results are of paramount practical importance for the finance industry, since recent years have seen an unprecedented surge in automated trading. In sum, we think that far more attention should be paid to modeling intraday volatility, and we propose an important contribution to the existing literature.

The paper is organized as follows. Section 2 presents the model. Section 3 describes the data and gives results of estimation. This is followed by a forecasting section and conclusions.

2. THE MODEL

2.1. Notation

We use the following notation. Days in the sample are indexed by t (t =1,…, T). Each day is divided into 10 minute intervals referred to as bins and indexed by i (i =0,…, 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 modeled as:

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

The overnight return in bin zero is deleted leading to a total number of return observations, M=TN.

2.2. Model

We propose a new 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 mean one, and

$\varepsilon_{\{t,i\}}$ is an error term.

The daily variance component could be specified in a number of ways. Andersen and Bollerslev (1997, 1998), estimate this component from a daily GARCH model for a longer sample, going back a number of months or years. It could also be estimated based on daily realized variance as proposed by Engle and Gallo(2005). We adopt a different route, however, and utilize commercially available volatility forecasts produced daily for each company in our sample. This eliminates the need for longer series for the daily model than for the intra-daily model. With the turnover of corporate ownership, it is difficult to get consistent long series for a big universe of stocks.

The diurnal component is calculated as the variance of returns in each bin after deflating by the daily volatility. To see this consider the variance of these returns:

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

and

$$E\left(\frac{r_{t,i}^2}{h_t}\right) = s_i E\left(q_{t,i}\right) = s_i \tag{3}$$

Practically, we estimate the model in two stages. First we normalize returns by daily and diurnal volatility components, and then model the residual volatility as a unit GARCH(1,1) process:

$$y_{\{t,i\}} = r_{\{t,i\}} / \sqrt{\hat{h}_t} = \sqrt{q_{\{t,i\}}} \varepsilon_{\{t,i\}} \tag{4}$$

$$q_{\{t,i\}} = \omega + \alpha (r_{\{t,i-1\}} / \sqrt{\hat{h}_t \hat{s}_{i-1}})^2 + \beta\, q_{\{t,i-1\}} \tag{5}$$

The GARCH specification can be rewritten as:

$$\begin{aligned} & z_{\{t,i\}} \big| F_{\{t,i-1\}} \sim N\left(0, q_{\{t,i\}}\right) \\ & q_{\{t,i\}} = \omega + \alpha z_{\{t,i-1\}}^2 + \beta\, q_{\{t,i-1\}} \\ & z_{\{t,i\}} = r_{\{t,i\}} / \sqrt{h_t \hat{s}_i} \end{aligned} \tag{6}$$

The unit GARCH might enforce the constraint $\omega = 1 - \alpha - \beta$ although in the empirical work this has not been done.

3. ECONOMETRIC ISSUES

In this section we will discuss statistical properties of the 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. Following equation (3) we estimate the diurnal component for each bin as the variance of $y_{\{t,i\}}$ in this bin. That is:

$$\hat{s}_i = \frac{1}{T} \sum_{t=1}^{T} y_{\{t,i\}}^2, \ \forall i = 1, \ldots, N \tag{7}$$

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 (6). Such a multi-step 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

$$\hat{\psi} = \begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} = \arg\min g_M ' W g_M = \arg\min g_M ' g_M \tag{8}$$

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) (c.f. 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:

$$\sqrt{M}\begin{pmatrix} \hat{\phi} - \phi_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{d} N\left(0, G^{-1}\Omega G^{-1}{'}\right) \tag{9}$$

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

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:

$$g_{1,M}\left(\phi\right) = g_{1,M} = \begin{pmatrix} 1/T\sum\limits_{\{t,1\}}^{T}(y_{\{t,1\}}{}^{2} - s_{1}) \\ \vdots \\ 1/T\sum\limits_{\{t,N\}}^{T}(y_{\{t,N\}}{}^{2} - s_{N}) \end{pmatrix} \tag{10}$$

$$g_{2,M}\left(\hat{\phi},\theta\right) = g_{2,M} = 1/TN\sum_{t=1}^{T}\sum_{i=1}^{N}\nabla_{\theta}\left(\log(q_{\{t,i\}}) + \left(y_{\{t,i\}}{}^{2}/\hat{s}_{i}q_{\{t,i\}}\right)\right) \tag{11}$$

$$G = \frac{1}{M}\sum\begin{pmatrix} \nabla_{\phi}g_{1} & 0 \\ \nabla_{\phi}g_{2} & \nabla_{\theta}g_{2} \end{pmatrix} \text{ and } \frac{1}{M}\sum\begin{pmatrix} g_{1,t}^{2} & g_{1,t}g_{2,t} \\ g_{1,t}g_{2,t} & g_{2,t}^{2} \end{pmatrix} \xrightarrow{p} \Omega \tag{12}$$

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 (7). 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, Hansen and Lee (1994) or Lumsdaine (1996). In sum, the consistency and asymptotic normality of the estimator (11) is a corollary to Theorem 6.1 (p. 2178) in Newey and McFadden (1994). The above results could, in principle, be generalized to a multi-step estimation.

4. EMPIRICAL RESULTS

4.1. DATA

Our sample consists of price data on 2721 companies obtained from the TAQ database. 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. Data spans a three-month period in April-June 2000.

4.2. RESULTS FOR A SINGLE STOCK

Some results will be presented using returns on a single randomly chosen NASDAQ-listed stock – Semitool Inc (SMTL). This company produces equipment for semiconductor industry. We divide 10-minute returns by their respective daily commercially available volatility forecasts. What can be observed for these data, however, is a 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.

[INSERT FIGURES 1,2 AND 3 ABOUT HERE]

The sample variance for each bin will be our estimate of diurnal variance component s_i. Hence in the second step, returns are normalized by their respective diurnal standard deviations. In order to take account of the remaining intraday dynamics, we fit a GARCH (1,1) model into returns standardized in that way. Figure 2 superimposes the three volatility components described above. For clarity, we have chosen to show an approximately three-week period at the very beginning of the sample (3-25 April 2000). The bold blue line shows the daily volatility forecast, which is the same for all bins on a given day. The green thin line represents the regular diurnal pattern, and the stochastic intraday component appears in red with dotted marks. We may appreciate that this component is able to modify the regular deterministic diurnal pattern.

Figure 3 consists of five panels. Top panel shows logarithmic returns normalized by the unconditional standard deviation of the series. This is followed by the square roots of the estimated variance components: daily, diurnal and intraday. These are followed by the square root of a composite variance component, being the product of the preceding three variance components.

4.3. RESULTS FOR A SAMPLE OF 2721 STOCKS

4.3.1 SEPARATE ESTIMATION RESULTS

Model (6) 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. Any remaining serial correlation is eliminated by fitting an ARMA(1,1). Estimation is performed for the period April-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 of a manageable size, we report results of this procedure resorting to graphical methods. Figure 4a 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 (5). The top and middle panels of Figure 4a depict β and α parameters, respectively. The bottom panel plots the sum of both parameters, thus informing us how persistent the volatility is. Figure 4b 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.

[INSERT TABLE 1 AND FIGURE 4 ABOUT HERE]

Companies at the left of Figure 4a are very actively traded, and at the very right- trade seldom. It can be observed that estimates' variability 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, Figures 4a and 4b give us a rationale for grouping companies for the purpose of estimation. It turns out that for some companies, especially the least trading 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 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.

4.3.2 GROUPED ESTIMATION RESULTS

As discussed in previous section, Figure 4a 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. However, we must normalize each company returns by its standard deviation to prevent the switching point from being a structural break.

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 (6). Although industry grouping is an obvious candidate, we have investigated a number of different admissible ways of sorting companies.

We have attempted to classify groups based on the exchange the stocks are traded on and if they are included in major indices. In particular we have obtained 5 groups: NYSE/NASDAQ exchange and S&P and non-S&P equities, with the remaining 5th category “Other stocks”. This exercise was motivated by the finding reported by some authors (c.f. Bennett and Wei, 2003) 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.

Another approach to pooling stocks is to sort them by time series characteristics. As previously mentioned, Figure 4a suggests a liquidity criterion for pooling companies into groups. We have tried several categories: we have grouped companies according to their capitalization and intensity 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 a 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.

[INSERT TABLES 2-4 AND FIGURES 5-7 ABOUT HERE]

In the INDUST mode we group data into 54 industries and estimate 54, instead of 2721 intraday GARCH (1,1) models. Each return series has been divided by its standard deviation in order to render returns comparable across stocks. Estimation results of this step are summarized in Table 2 and parameters plotted in Figure 5. We do not encounter any convergence problems as was the case for some companies 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. Please note that 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 we have previously removed the daily volatility component, responsible for a longer persistence.

Next we estimate GARCH models for 50 groups of companies sorted according to the liquidity criterion. Table 3 gives results and Figure 6 plots parameters. 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). It appears, however, that this sorting mode works well for less liquid stocks, a finding that will be reinforced by forecasting comparisons. Histogram in Figure 6b seemingly documents a reduction in intraday GARCH parameter variation with one notable exception. The least liquid group comprising 55 companies has a persistence parameter equal to 0.575. This group is characterized by spectacular kurtosis and skewness coefficients. Figure 7 gives a snapshot of 10-minute returns on the least trading stocks and modestly trading stocks. The bulk of observations in least trading group are equal to zero, and many nonzero observations could be described as outlying or “influential” because they equal to several standard deviations from the mean.

Finally, Table 4 reports results for ARCH estimation for the intraday component for one giant pool of all the companies, comprising over 4.2 million observations. Similar to the industry case, this table also indicates modest persistence of intraday volatility.

5. FORECASTING RESULTS

5.1. LOSS FUNCTIONS AND DESIGN

We now turn to out of sample forecast accuracy. We use the parameter estimates for the period April-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 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 times 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 ${z_{\{t,i\}}}^2 = r^2_{\{t,i\}} / \hat{h}_t \hat{s}_i$. 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 (2004). In the following, we use two loss functions:

L1	LIK	Out-of-sample likelihood	$L_{1t} = \log q_{\{t,i\}} + \frac{{z_{\{t,i\}}}^2}{q^f_{\{t,i\}}}$
L2	MSE	Mean Squared Error	$L_{2t} = \left({z_{\{t,i\}}}^2 - q^f_{\{t,i\}}\right)^2$

The use of squared return in place of the true volatility introduces biases in many popular loss functions problematic (so does the RV measure, c.f. Patton, 2004). However, under MSE and LIK loss functions optimal forecasts are unbiased (c.f. Patton, 2004 and Hansen and Lunde, 2005).

Although part of the literature on assessing forecasting performance of daily models (c.f. Hansen and Lunde, 2004) recommends using RV as a proxy, this paper uses squared returns because 10 minute interval does not allow a reliable measure of realized volatility to be estimated. For many of the companies in our sample and outside of the active trading periods, the small numbers of trades in 10-minute bins raised concern about the precision of RV estimates.

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 5 different forecasts that will form the basis for a subsequent model evaluation and comparison.

5.2. OUT-OF-SAMPLE FORECAST COMPARISON

We have performed five different estimations for companies pooled into groups in various ways and will refer to these ways as modes. The first mode (NSTOCH) contains no stochastic component (5) at all. Mode No. 2 (UNIQUE) involves no pooling, i.e. we estimate unique GARCH models for separate companies. Mode No. 3 (INDUST) denotes a GARCH estimation for companies grouped according to their primary industry classification. In Mode No. 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 5 separate estimations we have calculated a series of forecast errors. These forecast errors are used to calculate accuracy measurement criteria using loss functions L_1 and L_2, $L_j = \frac{1}{\tau}\sum_{t=1}^{\tau} L_{jt}$, where j=1,2; τ= 858 and τ denotes the length of the forecasting period.

Table 5 compares accuracy of volatility forecasts obtained from the above estimation modes. We have calculated two forecast accuracy measures for each of the 5 estimations and for each company, which amounts to a total of (5 modes*2 criteria * 2721 stocks =) 27210 numbers. Then 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 given estimation outperforms each of the remaining forecasts. Hence, numbers contained in Table 5 give us the 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 LIK loss function. For example, the third row second column compares results of NSTOCH vs. separate GARCH estimations. Here the number 0.618 means that the specification without component (5) yields worse forecasts than the individual company-by-company estimation 62% 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. The fifth column establishes forecasting inferiority of LIQUID mode in comparison to both INDUST and ONEBIG modes. Finally the sixth column offers ONEBIG mode as a winner of this competition. The above discussion carries over to the lower panel of Table 5, which presents results for LIK loss function. One exception is that the MSE criterion marginally favours INDUST estimation over ONEBIG model. 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.

[INSERT TABLES 5-7 ABOUT HERE]

Table 6 pertains to the same set of forecasting results as Table 5. It reports the mean and median of the forecast accuracy measures calculated for each of the 2721 companies and five estimation modes. The smallest number in each row denotes the smallest mean or median error. 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 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.

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 8.A 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 8.B.). 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 8.A recommends liquidity sorted GARCH model as a preferred forecasting tool for illiquid stocks.

[INSERT TABLE 8 ABOUT HERE]

The picture suddenly changes when we look at the results assembled in Table 8.B, which concern the most liquid stocks. Here 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.

To sum up, we have seen that inclusion of stochastic intraday component (5) 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. Most liquid stocks seem to benefit from the widest pooling possible, i.e. using all available cross section information. Most 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.

6. CONCLUSION

This paper proposes a new way of modeling and forecasting intraday returns. We decompose the volatility of high frequency asset returns into components that 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.

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.

Bekaert, G, C. Harvey and C. Lundblad (2005), Liquidity and Expected Returns: Lessons from Emerging Markets, unpublished manuscript.

Bennett P. and L. Wei (2003), 'Market Structure, Fragmentation and Market Quality - Evidence from Recent Listing Switches", unpublished manuscript.

Deo, R., C. Hurvich and Y. Lu (2005), "Forecasting realized volatility using a long memory stochastic volatility model: Estimation, prediction and seasonal adjustment" Forthcoming in the *Journal of Econometrics*.

Diebold F. and R. Mariano (1995), "Comparing forecasting accuracy", Journal of Business and Economic Statistics, 13(3), 253-63.

Drost, F.C. and Nijman, T.E. (1993), "Temporal Aggregation of GARCH Processes," Econometrica, 61, 909-927.

Ellul A. , C. W. Holden, P. Jain and R. Jennings (2003), "Determinants of Order Choice on the New York Stock Exchange", manuscript, Indiana University.

Engle, R. F. (2005), Integrating Investment Risk and Trading Risk, manuscript, Department of Finance, NYU.

Engle, R.F. (2002), "New Frontiers for ARCH Models", Journal of Applied Econometrics, 17, 425-446.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica 50(4), 987-1007.

Engle, R.F. and G.P. Gallo (2005), A Multiple Indicators Model for Volatility Using Intra-Daily Data.

Ghose D. and K. Kroner (1996), "Components of Volatility in Foreign Exchange Markets: An Empirical Analysis of High Frequency Data", mimeo Dept. of Economics, Univ. of Arizona.

Giot, P. (2003), Market Risk Models for Intraday Data, European Journal of Finance, forthcoming

Granger, C. (2003),Empirical modeling in economics.Specification and Evaluation, CUP.

Griffiths, M., Smith, B., Turnbull, D., White, R.W. (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: 1029–54.

Hansen, P. R. and A. Lunde, (2005), 'Consistent ranking of volatility models', Journal of Econometrics, Forthcoming.

Hasbrouck J. and G. Saar (2002) "Limit Orders and Volatility in a Hybrid Market: The Island ECN", Stern School of Business Working Paper, July 2002.

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

Lo, A. W., MacKinlay, A.C., and Zhang, J. (2002), Econometric models of limit order execution, Journal of Financial Economics 65, 31-71

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.

Newey W.K., McFadden, D. (1994) - Large Sample Estimation and Hypothesis Testing, in Handbook of Econometrics, vol. 4, 2113-2245.

Patton, A (2004), Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies, manuscript, London School of Economics.

Ranaldo, A. (2004), "Order aggressiveness in limit order book markets", Journal of Financial Markets 7, 53–74.

Taylor S.J., and X. Xu (1997), "The incremental volatility information in million foreign exchange quotations" Journal of Empirical Finance, 4, 317-340.

Table 1. SMTL Intraday GARCH Results

Parameter	Value	Standard Error	T Statistic
C	0.0013	0.0177	0.0726
ω	0.1896	0.0208	9.1076
β	0.6814	0.0293	23.2738
α	0.1295	0.0130	9.9463

Notes: This table presents estimation results for intraday GARCH(1,1) model for Semitools Inc. Sample period April-May 2000. Symbols α , β and ω denote GARCH parameters from the variance equation (5). C denotes a constant in the mean equation.

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	AIC	BIC	LM(1)	LM(20)
1	-0.17	19.6	92970	**0.863**	0.139	105.9	0.797	476.0	0.065	105.2	-258309.7	-258338.1	2.1	16.0
2	0.09	21.6	24100	**0.761**	0.244	62.0	0.663	130.6	0.098	56.7	-67122.8	-67147.1	0.0	4.5
3	-0.33	22.3	71292	**0.861**	0.142	131.4	0.787	532.0	0.074	92.3	-198004.8	-198032.3	0.5	10.7
4	0.12	18.4	134195	**0.889**	0.113	123.1	0.825	639.9	0.064	108.1	-372851.6	-372881.1	2.5	13.2
5	0.15	14.9	100579	**0.896**	0.106	113.8	0.829	634.0	0.067	95.2	-279073.6	-279102.2	0.4	13.0
6	-0.01	11.9	43641	**0.906**	0.095	50.0	0.849	325.3	0.058	51.4	-121803.9	-121830.0	0.4	4.6
7	-0.22	12.9	55380	**0.892**	0.111	63.0	0.795	288.6	0.097	60.3	-153314.6	-153341.4	0.9	38.7 **
8	-0.07	16.8	102608	**0.844**	0.159	107.2	0.751	354.8	0.092	93.2	-284211.1	-284239.7	0.5	17.2
9	-0.04	30.3	23673	**0.892**	0.116	75.1	0.812	351.9	0.079	64.1	-65365.8	-65390.0	0.3	4.2
10	0.61	21.6	10647	**0.858**	0.150	48.0	0.757	179.4	0.101	36.2	-29420.7	-29442.5	0.4	6.2
11	-0.09	16.2	37322	**0.836**	0.167	66.1	0.743	218.8	0.093	78.3	-103739.8	-103765.4	0.1	7.1
12	-0.38	32.0	28236	**0.766**	0.235	60.6	0.680	133.9	0.086	44.6	-78267.4	-78292.2	0.0	3.7
13	-0.19	19.4	33540	**0.819**	0.185	53.2	0.708	132.8	0.111	42.9	-92707.1	-92732.3	0.0	2.2
14	-0.26	15.9	56705	**0.909**	0.093	77.5	0.839	452.9	0.071	73.9	-156693.2	-156720.1	1.1	9.8
15	-0.13	17.3	44927	**0.911**	0.091	88.4	0.818	461.0	0.093	78.9	-122184.8	-122210.9	4.1 *	15.6
16	-0.01	13.1	37401	**0.894**	0.110	57.1	0.790	242.5	0.104	56.4	-102781.0	-102806.6	1.4	11.4
17	-0.27	16.5	90714	**0.892**	0.112	137.6	0.789	544.5	0.103	98.5	-247939.1	-247967.3	0.3	6.6
18	0.19	12.6	24959	**0.874**	0.129	54.3	0.770	216.1	0.104	49.7	-69013.1	-69037.5	0.6	13.2
19	0.21	22.1	84239	**0.922**	0.079	117.7	0.865	810.5	0.057	99.8	-232447.2	-232475.2	0.7	6.0
20	0.01	15.3	57330	**0.928**	0.075	78.5	0.863	568.6	0.065	72.9	-158422.2	-158449.1	1.9	19.4
21	-0.33	19.4	103229	**0.909**	0.094	142.1	0.832	807.2	0.076	110.4	-283229.7	-283258.4	0.4	13.7
22	0.21	12.1	28119	**0.866**	0.137	40.7	0.791	176.8	0.075	41.9	-78377.3	-78402.0	0.5	6.9
23	0.20	12.9	37401	**0.797**	0.208	49.2	0.679	124.4	0.119	59.2	-103768.6	-103794.2	0.7	11.1
24	-0.06	18.0	25583	**0.929**	0.073	65.4	0.867	516.7	0.061	55.2	-70375.5	-70399.9	0.5	9.7
25	-0.37	16.0	48163	**0.911**	0.090	86.8	0.839	529.8	0.072	69.5	-132640.9	-132667.2	3.5	14.2
26	0.22	7.7	10919	**0.918**	0.083	19.1	0.826	123.0	0.092	24.6	-30003.2	-30025.1	1.9	9.0
27	-0.11	14.0	42119	**0.879**	0.124	62.0	0.785	242.4	0.094	55.2	-115949.9	-115975.8	0.2	4.5
28	-0.33	16.5	17121	**0.816**	0.185	31.8	0.743	100.3	0.073	31.2	-47691.1	-47714.3	0.1	2.5
29	0.11	22.2	103542	**0.910**	0.093	173.4	0.845	1040.1	0.065	126.9	-286380.4	-286409.0	0.8	13.0
30	0.15	10.0	119105	**0.865**	0.136	121.4	0.786	517.5	0.079	90.0	-332176.0	-332205.0	8.2 **	40.6 **
31	0.18	13.8	58733	**0.914**	0.087	72.6	0.857	505.5	0.057	70.1	-163754.1	-163781.0	0.8	9.3
32	0.28	11.9	15600	**0.912**	0.093	25.2	0.807	136.5	0.106	30.5	-42909.4	-42932.4	0.0	9.0
33	0.14	19.3	31628	**0.819**	0.187	77.5	0.707	206.9	0.112	58.0	-87301.5	-87326.6	0.1	6.9
34	-0.11	20.5	44691	**0.909**	0.096	115.0	0.816	590.0	0.094	83.5	-121879.7	-121905.8	3.3	11.5
35	0.11	12.1	51400	**0.914**	0.090	60.8	0.809	350.2	0.105	75.5	-140405.5	-140432.1	12.5 **	23.1
36	0.20	18.1	108809	**0.904**	0.101	121.8	0.808	629.2	0.095	122.4	-298386.7	-298415.5	4.1 *	18.7
37	-0.19	16.7	214610	**0.919**	0.086	160.4	0.810	985.7	0.109	198.7	-582635.7	-582666.6	2.2	15.9
38	-0.03	13.3	222297	**0.920**	0.083	186.0	0.810	943.2	0.110	164.7	-599522.4	-599553.3	3.4	44.0 **
39	0.02	10.5	148821	**0.944**	0.060	108.0	0.822	755.4	0.122	128.8	-396821.0	-396850.7	10.1 **	69.1 **
40	0.04	14.7	125260	**0.918**	0.084	178.5	0.817	871.1	0.100	130.7	-338694.3	-338723.5	0.5	7.4
41	-0.13	13.0	258050	**0.933**	0.072	210.5	0.813	1319.2	0.120	211.7	-690593.4	-690624.7	19.8 **	82.7 **
42	0.21	13.3	34240	**0.903**	0.100	54.5	0.814	285.0	0.089	53.5	-94093.9	-94119.2	3.8	12.4
43	-0.18	15.3	113017	**0.907**	0.095	124.0	0.818	622.6	0.089	105.5	-308729.2	-308758.1	0.6	14.0
44	-0.37	19.1	50504	**0.889**	0.115	120.6	0.779	464.7	0.110	84.7	-137111.2	-137137.7	0.0	3.7
45	-0.18	15.8	174678	**0.922**	0.082	178.0	0.820	937.9	0.102	152.7	-473090.9	-473121.1	3.0	25.7
46	-0.06	17.6	85601	**0.919**	0.083	114.8	0.845	765.2	0.074	116.6	-234855.2	-234883.2	5.0 *	19.9
47	0.08	21.5	34865	**0.896**	0.109	87.2	0.802	428.8	0.094	84.2	-95357.6	-95382.9	5.7 *	17.6
48	-0.05	20.1	96524	**0.930**	0.072	162.9	0.869	1462.3	0.060	123.0	-266470.5	-266498.9	12.0 **	22.1
49	-0.07	21.4	218205	**0.905**	0.097	176.9	0.845	1042.4	0.060	144.9	-605493.6	-605524.5	3.1	15.1
50	-0.15	19.4	48044	**0.892**	0.110	59.7	0.839	348.0	0.053	60.0	-134115.7	-134142.0	0.4	7.6
51	-0.19	15.2	57679	**0.910**	0.094	114.3	0.808	540.9	0.102	88.1	-156792.1	-156819.0	0.3	8.6
52	-0.39	18.9	87902	**0.944**	0.058	148.1	0.880	1358.0	0.064	127.4	-240292.6	-240320.8	4.9 *	9.2
53	-0.06	11.2	79754	**0.956**	0.050	73.2	0.822	514.5	0.133	97.5	-209371.4	-209399.2	7.8 **	39.1 **
54	0.06	13.4	156155	**0.906**	0.096	94.1	0.850	607.8	0.056	101.1	-436825.6	-436855.5	0.9	15.5

Notes: This table presents estimation results for intraday GARCH(1,1) models for 54 industries. Sample period April-May 2000. Symbols α, β and ω denote GARCH parameters from the variance equation (5). Persistence is measured as the sum of parameters ($\alpha + \beta$). AIC and BIC denote Akaike and Schwartz Information Criteria, respectively. LM(1) and LM(20) statistics are calculated as the ARCH LM test, cf. Engle (1982), on the residuals from (6). Under the null of no ARCH effects at lag q, the statistic has a chi2 distribution with q degrees of freedom, where q= 1, 20. * , ** 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	AIC	BIC	LM(1)	LM(20)
1	-1.05	133.9	**0.575**	0.439	136.1	0.514	146.3	0.061	66.1	-166244.5	-166271.4	0.3	2.1
2	-0.48	66.6	**0.916**	0.085	191.6	0.886	1655.6	0.030	148.1	-223214.4	-223242.3	0.6	4.5
3	-0.45	41.1	**0.926**	0.075	124.7	0.892	1145.5	0.034	116.7	-234652.3	-234680.3	1.0	8.7
4	-0.07	28.9	**0.925**	0.076	100.3	0.893	934.5	0.032	99.4	-238318.5	-238346.5	5.5 *	14.6
5	-0.16	29.2	**0.917**	0.085	113.7	0.869	866.7	0.047	110.5	-237045.4	-237073.5	1.0	4.7
6	0.07	24.8	**0.887**	0.115	118.1	0.826	654.5	0.061	124.7	-237362.3	-237390.3	2.0	8.1
7	-0.02	21.5	**0.918**	0.084	93.9	0.866	714.3	0.051	102.4	-238251.4	-238279.5	1.7	10.1
8	-0.06	20.2	**0.928**	0.074	103.6	0.876	889.0	0.052	114.1	-237731.6	-237759.6	4.7 *	16.0
9	0.04	15.7	**0.946**	0.055	75.0	0.907	905.9	0.039	89.5	-238954.2	-238982.2	9.7 **	25.6
10	-0.19	15.5	**0.933**	0.068	82.2	0.883	774.1	0.050	90.7	-238538.4	-238566.5	6.8 **	22.8
11	0.08	15.6	**0.918**	0.083	81.9	0.864	624.6	0.054	87.8	-238514.3	-238542.4	5.2 *	18.2
12	0.03	17.0	**0.873**	0.129	87.6	0.804	398.2	0.069	82.5	-238273.7	-238301.8	1.4	17.8
13	-0.11	16.6	**0.909**	0.093	83.8	0.830	511.9	0.079	110.5	-236173.6	-236201.7	0.6	11.3
14	0.09	14.0	**0.928**	0.074	94.4	0.859	760.1	0.069	98.0	-236305.5	-236333.6	16.7 **	38.5 **
15	-0.14	14.1	**0.905**	0.097	75.8	0.833	448.2	0.072	81.8	-237521.2	-237549.3	1.8	11.7
16	0.10	13.6	**0.931**	0.071	86.0	0.863	690.1	0.068	93.0	-236377.8	-236405.9	9.8 **	25.9
17	-0.05	12.4	**0.915**	0.089	74.2	0.832	478.3	0.083	85.9	-236897.4	-236925.5	8.0 **	30.8
18	-0.10	14.1	**0.916**	0.087	77.8	0.829	496.1	0.087	97.6	-235707.8	-235735.9	3.8	21.7
19	-0.10	15.1	**0.928**	0.075	107.0	0.832	651.4	0.096	107.6	-232606.3	-232634.3	3.1	20.2
20	-0.03	11.8	**0.934**	0.069	80.6	0.855	600.0	0.078	84.1	-235792.1	-235820.2	9.6 **	31.7 *
21	-0.22	14.8	**0.906**	0.099	78.8	0.802	427.4	0.104	89.7	-234870.7	-234898.8	8.4 **	28.1
22	-0.16	12.2	**0.901**	0.103	80.7	0.800	401.6	0.101	86.9	-230967.3	-230995.3	1.9	15.1
23	-0.01	10.8	**0.903**	0.101	71.6	0.804	376.0	0.099	82.0	-231523.6	-231551.6	9.0 **	44.9 **
24	0.09	11.2	**0.898**	0.105	76.6	0.798	388.9	0.100	81.0	-231429.2	-231457.3	13.6 **	41.0 **
25	-0.09	13.3	**0.930**	0.075	98.4	0.834	630.5	0.095	92.8	-230710.7	-230738.7	14.5 **	38.0 **
26	0.04	10.8	**0.921**	0.083	72.6	0.823	441.9	0.098	82.4	-230872.9	-230900.9	14.5 **	41.9 **
27	0.05	10.9	**0.925**	0.080	73.2	0.815	448.5	0.110	87.6	-229528.8	-229556.8	14.3 **	48.5 **
28	-0.09	13.4	**0.909**	0.098	78.1	0.785	379.7	0.123	93.0	-229022.1	-229050.1	2.4	25.4
29	0.08	10.4	**0.908**	0.095	67.5	0.799	352.1	0.110	79.0	-229024.3	-229052.3	5.2 *	32.0 *
30	-0.08	11.5	**0.899**	0.105	86.9	0.787	401.6	0.112	88.9	-230512.6	-230540.6	9.7 **	31.5 *
31	-0.11	11.0	**0.922**	0.083	73.6	0.811	436.3	0.111	91.6	-241774.3	-241802.5	14.5 **	60.8 **
32	-0.12	12.4	**0.920**	0.085	76.2	0.808	446.4	0.112	97.6	-238271.6	-238299.8	7.3 **	33.7 *
33	0.02	10.6	**0.889**	0.115	73.8	0.768	316.4	0.121	83.0	-235022.9	-235051.0	12.4 **	47.9 **
34	-0.11	9.5	**0.894**	0.111	72.9	0.761	302.3	0.133	79.5	-233621.2	-233649.3	4.7 *	43.1 **
35	0.01	9.7	**0.904**	0.100	69.8	0.786	312.7	0.118	73.0	-230334.7	-230362.7	7.9 **	38.2 **
36	-0.12	10.5	**0.900**	0.105	63.8	0.768	280.0	0.131	76.2	-232441.0	-232469.1	6.2 *	34.4 *
37	0.03	10.0	**0.908**	0.098	68.6	0.774	317.9	0.134	83.9	-232382.9	-232411.0	5.7 *	30.6
38	-0.04	9.7	**0.915**	0.090	77.5	0.791	425.2	0.124	91.5	-232209.1	-232237.1	6.4 *	42.3 **
39	0.01	8.5	**0.908**	0.097	62.3	0.782	307.0	0.126	74.7	-230539.3	-230567.4	11.6 **	46.0 **
40	-0.02	8.9	**0.922**	0.082	64.7	0.804	366.2	0.119	76.0	-234015.9	-234044.0	10.2 **	52.1 **
41	-0.07	9.4	**0.924**	0.081	73.6	0.797	393.2	0.126	81.9	-227993.4	-228021.4	5.4 *	46.9 **
42	-0.02	9.4	**0.919**	0.087	87.8	0.784	454.4	0.134	90.5	-228162.5	-228190.5	8.2 **	59.8 **
43	-0.07	8.9	**0.931**	0.072	61.2	0.816	390.1	0.116	78.3	-227845.4	-227873.4	20.8 **	74.9 **
44	-0.11	10.6	**0.936**	0.070	71.2	0.810	426.8	0.126	85.7	-231733.2	-231761.3	6.7 **	40.3 **
45	-0.05	9.3	**0.934**	0.070	60.0	0.808	380.1	0.126	86.4	-231166.0	-231194.0	10.8 **	47.4 **
46	-0.04	9.0	**0.919**	0.084	71.0	0.790	330.6	0.130	76.3	-228379.7	-228407.7	13.2 **	49.0 **
47	0.03	9.2	**0.948**	0.055	53.9	0.828	392.6	0.120	71.1	-229369.1	-229397.2	37.9 **	86.8 **
48	0.07	9.3	**0.959**	0.046	63.8	0.837	545.7	0.121	82.0	-231876.6	-231904.7	10.5 **	44.9 **
49	0.12	9.4	**0.954**	0.050	55.9	0.823	418.9	0.131	79.8	-222887.3	-222915.3	7.7 **	31.1
50	0.27	9.9	**0.964**	0.040	66.9	0.802	389.7	0.162	86.1	-215098.5	-215126.5	1.6	36.9 *

Notes: This table presents estimation results for intraday GARCH(1,1) models for 50 groups of liquidity sorted companies. Sample period April-May 2000. Symbols α , β and ω denote GARCH parameters from the variance equation (5). Persistence is measured as the sum of parameters ($\alpha + \beta$). AIC and BIC denote Akaike and Schwartz Information Criteria, respectively. LM(1) and LM(20) statistics are calculated as the ARCH LM test, cf. Engle (1982), on the residuals from (6). Under the null of no ARCH effects at lag q, the statistic has a chi2 distribution with q degrees of freedom, where q= 1, 20. * , ** denote significance at the 5% and 1% levels, respectively.

Table 4. All Sample Estimation Results- ONEBIG Mode

Parameter	Value	Standard Error	T Statistic
ω	0.096	0.000112	854.3
β	0.823	0.000180	4570.4
α	0.084	0.000115	728.6

Notes: This table presents estimation results for intraday GARCH(1,1) models for one large group of companies pooled together. Sample period April-May 2000. Symbols α , β and ω denote GARCH parameters from the variance equation (5).

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		0.382	0.275	0.245	0.262
UNIQUE	0.618		0.354	0.404	0.339
INDUST	0.725	0.646		0.572	0.445
LIQUID	0.755	0.596	0.428		0.377
ONEBIG	0.738	0.661	0.555	0.623	
Modes	**NSTOCH**	**UNIQUE**	**INDUST**	**LIQUID**	**ONEBIG**
MSE loss function					
NSTOCH		0.304	0.203	0.307	0.222
UNIQUE	0.696		0.419	0.465	0.432
INDUST	0.798	0.581		0.615	0.520
LIQUID	0.694	0.535	0.385		0.411
ONEBIG	0.778	0.568	0.480	0.590	

Notes: This table compares accuracy of one-step-ahead volatility forecasts obtained from five estimation modes. It contains 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.

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, the worst models are assigned the rank number 5.

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

Modes	**NSTOCH**	**UNIQUE**	**INDUST**	**LIQUID**	**ONEBIG**
LIK loss function					
NSTOCH		-5.8297 *	-35.772 *	-40.947 *	-34.688 *
UNIQUE	5.8297 *		-2.074 *	-2.459 *	-1.998 *
INDUST	35.772 *	2.074 *		-4.964 *	1.367
LIQUID	40.947 *	2.459 *	4.964 *		5.351 *
ONEBIG	34.688 *	1.998 *	-1.367	-5.351 *	
MSE	**NSTOCH**	**UNIQUE**	**INDUST**	**LIQUID**	**ONEBIG**
MSE loss function					
NSTOCH		-3.5983 *	-22.176 *	-23.956 *	-20.394 *
UNIQUE	3.598 *	*	-7.404 *	-7.481 *	-7.322 *
INDUST	22.176 *	7.404 *		-1.377	-0.335
LIQUID	23.956 *	7.481 *	1.377		-0.624
ONEBIG	20.394 *	7.322 *	0.334	-0.624	

Notes: 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..

Table 8. Forecast accuracy comparison for most and least liquid stocks

A. Least Liquid Stocks

	Forecast accuracy measures	**LIK**	**Rank**	**MSE**	**Rank**	**Av. 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**	**Av. 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**

Notes: 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.

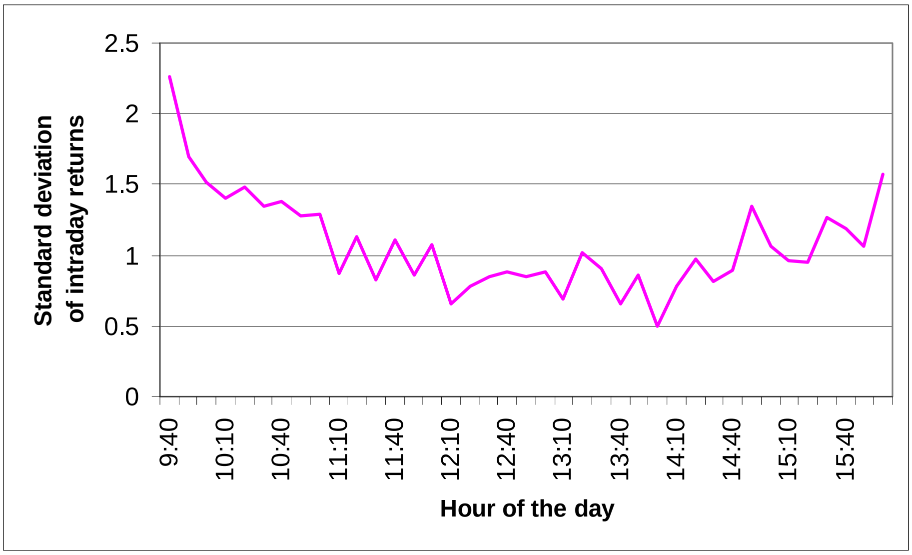


Figure 1. Standard deviation of returns across bins for SMTL stock. The horizontal axis labels denote hours during a trading day. Values depicted in this graph are calculated as a standard deviation of 10-min returns in each bin. Returns have been previously divided by daily volatility components.

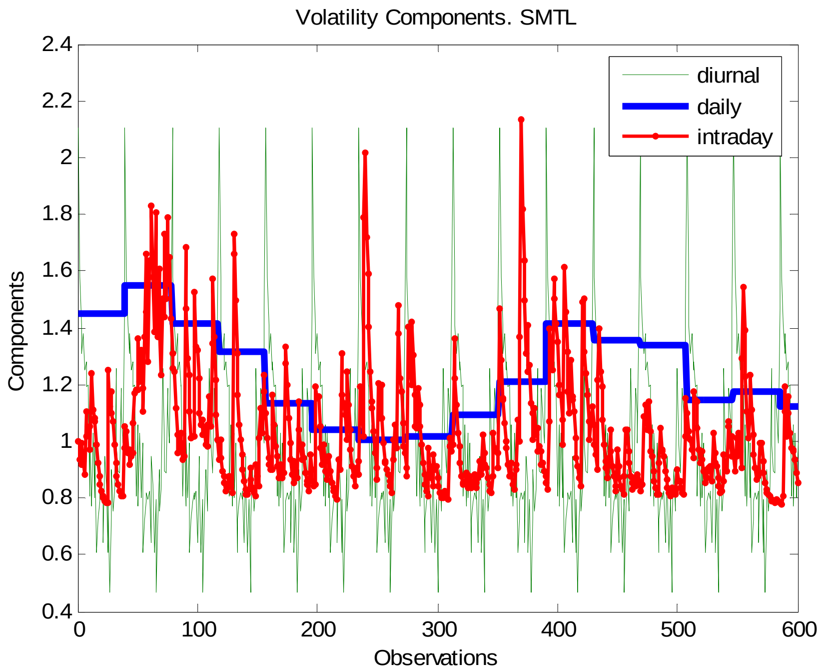


Figure 2. Volatility Components for Semitools Inc. This figure superimposes square roots of variance components estimated for SMTL, Semitools Inc. For clarity this picture offers a snapshot for the period 3-25 April 2000. The bold line shows the daily volatility forecast, which is the same for all bins on a given day. The green thin line represents the regular diurnal pattern, and the stochastic intraday component appears in red with dotted marks.

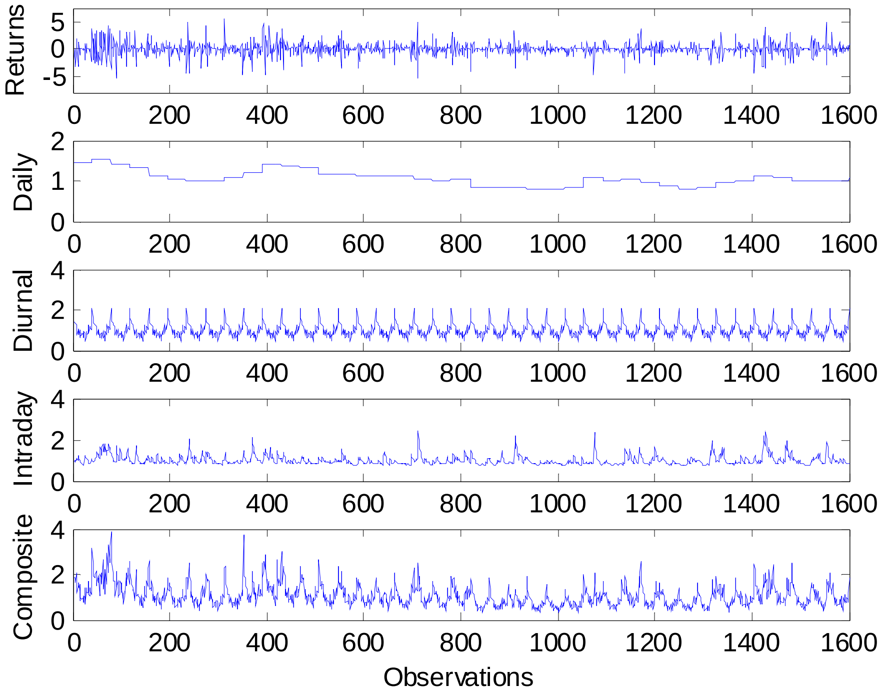


Figure 3. Volatility Components for Semitools Inc. Estimation period April-May 2000. *Top panel:* Logarithmic intraday returns on SMTL stock normalized by their unconditional standard deviation. *Second panel:* The square root of the daily variance component. *Third panel:* The square root of the diurnal 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.

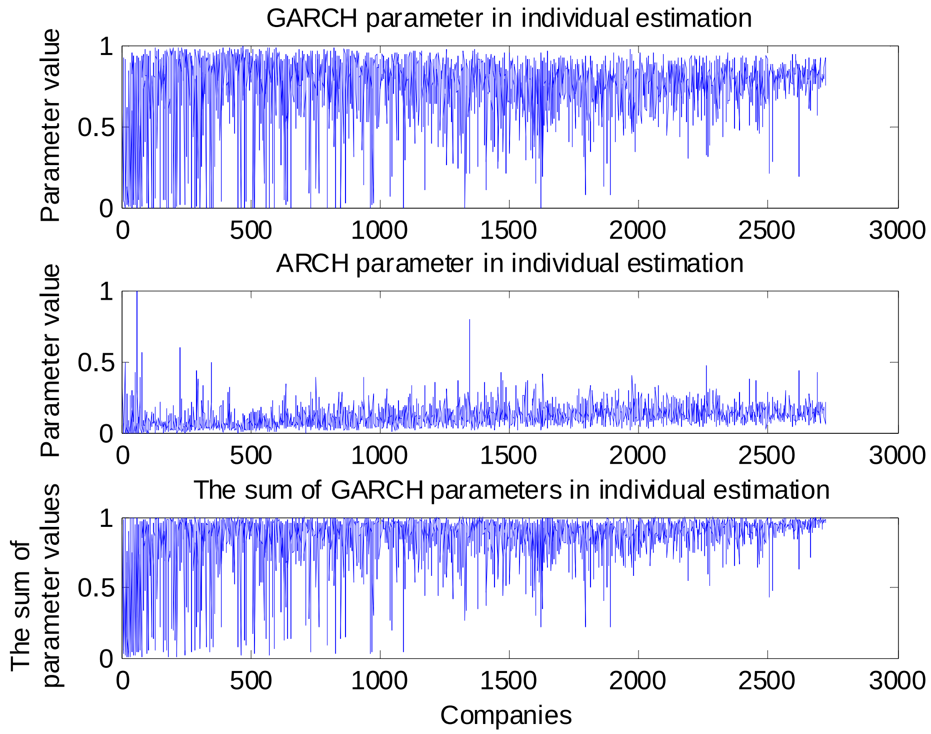


Figure 4a. Estimation results for the intraday GARCH models for 2721 for separate companies. Sample period April-May 2000. For the purpose of this picture companies were sorted by their average daily number of trades. *Top panel:* GARCH β from equation (5). *Second panel:* ARCH parameter α from equation (5). *Third panel:* Persistence measure $(\alpha + \beta)$.

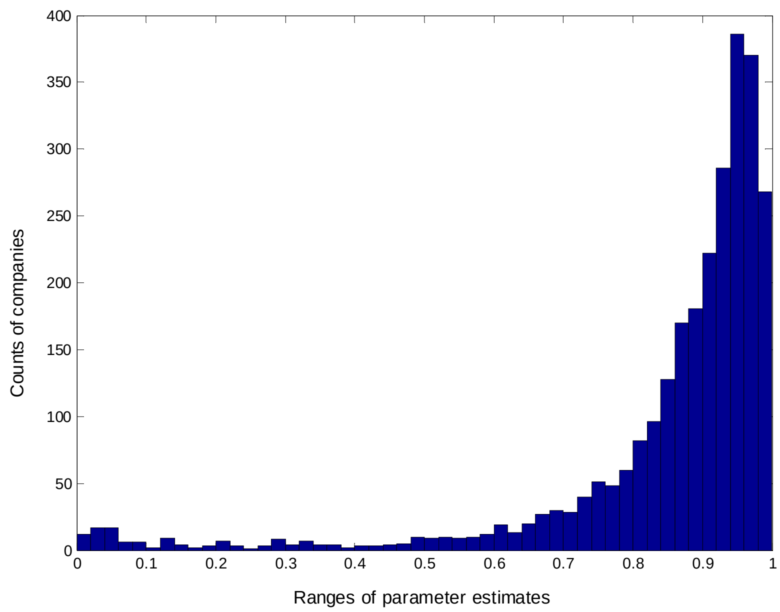


Figure 4b. Histogram of the persistence measure $(\alpha + \beta)$ from intraday GARCH estimation for 2721 Separate Models. Sample period April-May 2000. The horizontal axis denotes the value of the persistence parameter $(\alpha + \beta)$ and the vertical axis denotes the number of companies with the estimated persistence parameters falling into a corresponding bin.

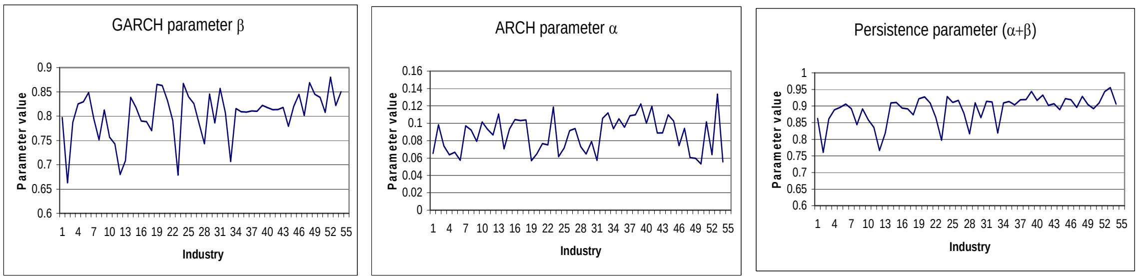


Figure 5a. Estimation results for the intraday GARCH models for 54 industries. Sample period April-May 2000. *First panel:* GARCH parameter β from equation (5). *Second panel:* ARCH parameter α from equation (5). *Third panel:* Persistence measure (α+β).

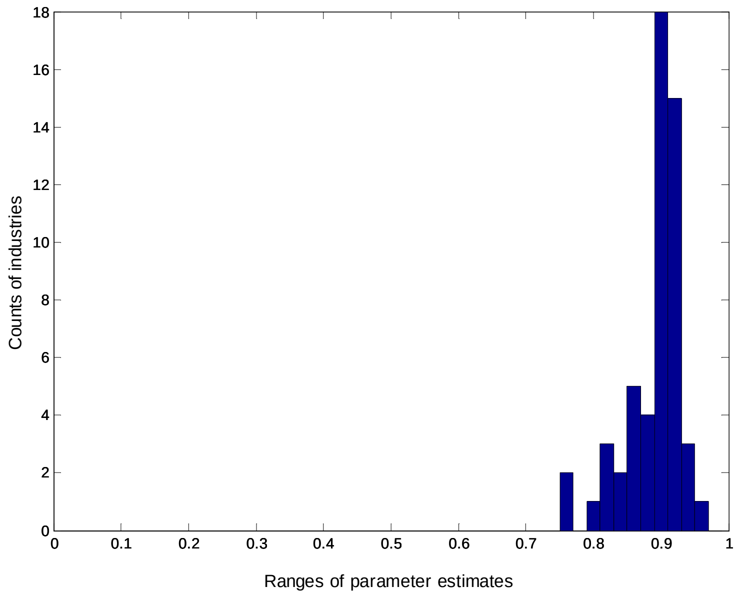


Figure 5b. Histogram of the persistence measure (α + β) from intraday GARCH estimation for 54 industries. Sample period April-May 2000. The horizontal axis denotes the value of the persistence parameter (α + β) and the vertical axis denotes the number of companies with the estimated persistence parameters falling into a corresponding bin.

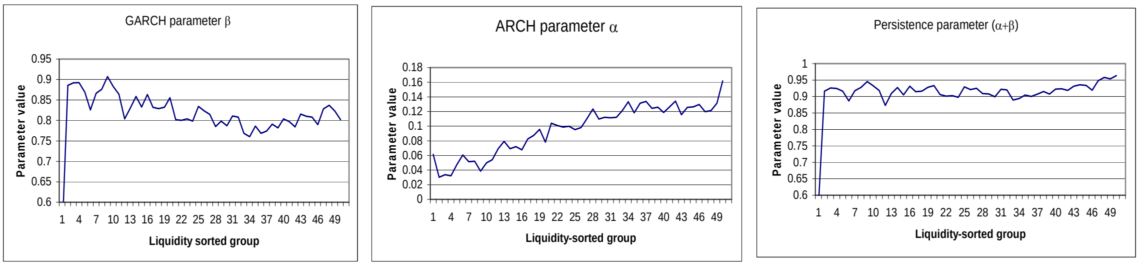


Figure 6a. Estimation results for the intraday GARCH models for 50 liquidity-sorted groups. Trading intensity or liquidity increases from the left to the right side of each picture. Sample period April-May 2000. *First panel:* GARCH parameter β from equation (5). *Second panel:* ARCH parameter α from equation (5). *Third panel:* Persistence measure (α+β).

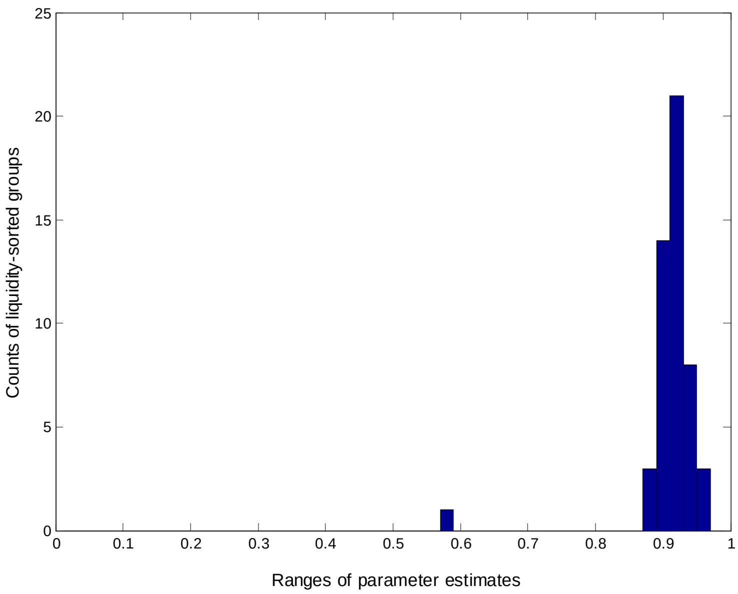
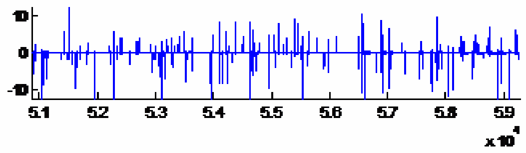


Figure 6b. Histogram of the persistence measure (α + β) from intraday GARCH estimation for 50 liquidity-sorted groups. Sample period April-May 2000. The horizontal axis denotes the value of the persistence parameter (α + β) and the vertical axis denotes the number of companies with the estimated persistence parameters falling into a corresponding bin.

A. Least Trading Group (1/50)



B. Modestly Trading Group (15/50)

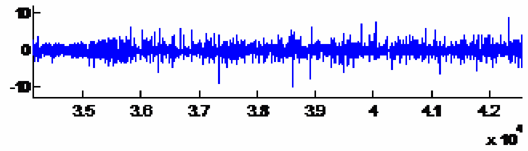


Figure 7. Examples of standardized logarithmic returns for two of the 50 Liquidity-Sorted Groups. Horizontal axis denotes *i*th observation in each liquidity sorted group and snapshots were chosen at random from the upper half of the groups.