When Is Noise Not Noise – A Microstructure Estimate of Realized Volatility*

by
Robert Engle[1] and Zheng Sun[2]

Abstract

This paper studies the joint distribution of tick by tick returns and durations between trades. We build an econometric model for estimating and forecasting the volatility of stock returns using high-frequency data, correcting for the bias incurred by microstructure noise. Three features of the model are worth mentioning: first the conditional volatility adapts a structure which incorporates past days as well as recent trades' information; second the volatility of returns is a nonlinear function of its contemporaneous duration; third the assumption of microstructure noise is general enough to encompass most of the properties implied from the theoretical literature. We apply the above model to frequently traded NYSE stock transactions data. It appears that contemporaneous duration has little effect on the volatility per trade after conditioning on the past, which means average per second volatility is inversely related to the duration between trades. Microstructure noise is found to be informative about the unobserved efficient price, and the informational component explains 45% of the total variation of the microstructure noise.

* Previously titled "Forecasting Volatility Using Tick by Tick Data". For comments and suggestions, we are greatly indebted to Joel Hasbrouk, Charles Jones, Lasse Pedersen, Gideon Sarr, Albert Menkveld and senimar participants at NBER microstructure group meeting, NYU Quantitative Financial Econometric meeting and European Financial Association annual conference at Moscow.

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

2 Department of Finance, Stern School of Business, New York University, zsun@stern.nyu.edu.

I. Introduction

Ultra-high frequency data consist of records of all transactions and their characteristics. Compared with the lower frequency wall-clock data, tick by tick data have several advantages in estimating return volatility. First, under ideal circumstances, the realized volatility is consistent for integrated volatility (Andersen and Bollerslev (1998)). As the sample frequency increases, the estimation error diminishes. Therefore the realized volatility computed from the highest frequency data ought to provide the best possible estimate for the integrated volatility. Second, when transformed to calendar time data, price at time t does not necessarily have a corresponding observation at that time, which will induce the so called "non-synchronous trading effect" [Andrew Lo and MacKinlay (1990)]. Third, and perhaps most importantly, tick by tick data preserves one important feature of the financial markets – duration between transactions. If information causes trading, then non-trading and price are jointly determined by the amount of news in the market[3], hence are correlated. Therefore, forecasts of moments of returns conditional on duration will be more efficient than those without. This motivates us to build an econometric model that incorporates the duration information in forecasting the volatility of returns.

The autoregressive conditional duration model (ACD) proposed by Engle and Russell (1998), which focuses on the time elapsed between the occurrences of trading events, forms a basis for incorporating trading duration information into the analysis of irregularly-spaced high frequency data. Basically, they model the duration process as a conditional exponential process where past information only affects waiting time through the conditional mean. Combined with the ACD model, Engle (2000) builds an UHF-GARCH model of volatility per unit of calendar time, and finds that both returns and variances are negatively influenced by long durations. Following these works, we model the joint density of the marked point process of durations and tick by tick

[3] Diamond and Verrecchia (1987), Admati and Pfleiderer (1988) and Easley and O'Hara (1987) have written theoretical models that have implications on how news and transaction frequency are related.

returns extending the ACD/UHF-GARCH framework. We first model the duration variable as an ACD process that could potentially depend on past returns. We then model conditional volatility of the efficient price change as a function of previous-day and recent trade information as well as the duration since the last trade. Once the distribution of duration conditional on past information (return and duration) and the distribution of return conditional on current duration and past information are specified, the joint distribution of return and duration is obtained. Therefore, we could forecast return volatility during any arbitrary length of time using simulation.

Standard Market Microstructure theory implies a difficulty in estimating volatility using ultra-high frequency data: the unobservability of the efficient price. For example, the existence of bid-ask spread makes the observed price not the efficient price, but the efficient price plus some noise, which adds additional volatility to returns. This problem is most serious in the high-frequency data, since the volatility from bid-ask spread usually doesn't shrink with the time interval but the volatility from efficient price does [Yacine Ait-Sahalia (2003)]. Therefore, in practice only moderate-frequency data are used which results in an inefficient measure of volatility. In order to filter out the bias from microstructure noise, and regain efficiency from the high frequency data, several novel nonparametric methods have been designed. Zhang, Mykland and Ait-Sahalia (2003) develop a 'two scales estimator', which yields a consistent estimator of realized volatility. However, their method assumes that microstructure noise is an i.i.d. process, which is too special compared with the implications from large parts of microstructure theory. Bearing this limitation in mind, Hansen and Lunde incorporate autocorrelation as well as price-noise dependence in the microstructure noise, and propose an auto-covariance correction of the realized variance. However, the model assumes finite moving average for the noise process, which rules out infinite moving average process such as AR or ARMA ones. Moreover, as the authors pointed out themselves, their realized variance estimator does not guarantee positive-ness, yet choosing some other bandwidth such as in Newey-West guarantees positiveness, but this may incur some biasedness.

Our assumption for the microstructure noise is general enough to encompass most of the sources that have been studied in the literature [Stoll (1989), Huang and Stoll (1997)]. In particular, we include two parts of microstructure noise: the first part is the fixed noise potentially due to order processing cost [Roll (1984)] or inventory control by dealers [(Amihud and Mendelson (1986)]; and the second part is time-varying noise correlated with the efficient price change, which may come from asymmetric information or stale prices. Both of the noise processes are allowed to be infinite moving averages. The model is estimated using Kalman Filtering. We specify different structures of the variances on the informational and non-informational innovation to help to identify the model.

In terms of econometric modeling, the paper that is mostly closely related to ours is the one by Frijns and Schotman (2005), where the authors study price discovery in tick time. They also use state-space models for incorporating microstructure noise into the price processes and treat the volatility of return as a nonlinear function of duration multiplying the volatility of the innovation. However, there are several distinctions between the two papers. First, the two papers try to answer different questions, Frijns and Schotman study the information share from different markets, so quote data are used, while our paper models volatility for transaction prices. Second, volatility was not a focus of the other paper, so it is assumed constant other than the effect from duration. Our paper builds a more elaborate model of volatility which incorporates long term and short term persistence of volatility as well as the time of day effect. Third, our specification for microstructure noise is more general than Frijns and Schotman's paper, so it is able to encompass a wide range of theoretical microstructure models. And finally, the stocks that are studied in our paper are traded on NYSE, while they use NASDAQ stocks; since the two markets have very different trading mechanisms as well as investor composition, the price-trading intensity relationship could be different.

There are two main findings: first, volatility per trade increases less than linearly in duration, consistent with the “no news no trade” prediction by Easley and O’Hara (1987). Moreover, for most of the less frequently traded stocks, duration has little effect on tick by tick volatility, implying a shorter memory in tick time rather than "wall clock" time. Each trade bears the same amount of information, and the total amount of information determines the number of transactions. Second, the realized volatility for high frequency data tends to be an upward biased estimate of the volatility of the efficient price change, because the observed price is contaminated by the microstructure noise. The microstructure noise tends to be time dependent and is informative about the unobserved efficient price change. The informational component accounts for 45% of the total variation of the microstructure noise on average.

The rest of the paper is organized as follows: Section two summarizes the properties of microstructure noise implied by the theoretical literature; Section three lays out the econometric models of the joint distribution of returns and durations and microstructure noise. Model estimation is described in Section four and Section five applies the model to a sample of NYSE stocks. Section six discusses implications of the results on the realized volatility estimation, and Section seven concludes.

II. Microstructure Noise

Our paper assumes a typical price generating mechanism: auction-dealer mechanism where market specialists quote bid and ask prices that they are willing to trade, and the orders are executed at either bid or ask. When quoting the prices, the market maker has a belief about the fundamental value of the stock, conditional on all the public information up to the moment. However, she is not willing to transact at a price level equal to this belief due to various market frictions documented in the literature. Instead, she will ask for a price concession for both incoming buy and sell orders, which makes the observed transaction price different from her belief of the fundamental. Hereafter, we name the market maker's belief of value of the stock as the efficient price (denoted by m_t) since it is the price available for the investor when the market is free of frictions. Efficient market hypothesis implies that m_t is a martingale process. We define the microstructure noises as the difference between the observed price and the efficient price. In particular, let p_t be the observed price and u_t be the microstructure noise, then we have $p_t = m_t + u_t$. Usually the observed price cannot deviate too much from the fundamental price, so u_t is generally assumed stationary. Various sources of friction will impose different structures on u_t. Empirically, which sources are observable and whether there is a dominant one is not clear. However, if one wants to test the existence of these sources and later filter them out, he would need to make general enough assumptions about the structure of the noise so that it incorporates most of the existing literature. This section summarizes various types of frictions suggested in the literature, which shed light on properties of the microstructure noise in our model.

1. Roll Model (1984)

The observed price process is the efficient price adjusted by order processing cost.

$$p_t = m_t + cq_t$$

Where q_t is the trading direction. q_t=1 if a trade is initiated by a buy order and q_t=-1 if a trade comes from a sell order. q_t is an i.i.d. process, which is commonly assumed in the high-frequency volatility estimating models.

2. Stale Prices

In this model, due to slow operational systems, trades actually occur relative to a stale price.

$$p_t = m_{t-1} + cq_t$$

Again q_t is i.i.d. noise as in the Roll model. It follows that

$$p_t = m_t + (cq_t - \Delta m_t)$$

In this case the microstructure noise is still i.i.d., but is (negatively) correlated with the innovation of efficient price.

3. Lagged Adjustment

The beliefs about the efficient prices are given by a martingale. But transaction prices adjust to beliefs gradually.

$$p_t = p_{t-1} + \alpha(m_t - p_{t-1}) = \alpha m_t + (1-\alpha)p_{t-1}$$

It is easy to show that $p_t = m_t + \frac{(\alpha - 1)}{1-(1-\alpha)L}(m_t - m_{t-1})$, therefore the microstructure noise follows an AR (1) process and is perfectly dependent with the efficient price change.

4. Inventory Control

The market makers' objective is to maintain an $I*$ inventory through the adjustment of bid and ask quotes; therefore their inventory level will follow a mean-reverting process. Let I_t be inventory at time t, and suppose some fraction of inventory imbalance can be liquidated each period, then $I_t - I* = \gamma(I_{t-1} - I*) + \varepsilon_t$, where γ is between 0 and 1. I_t can be viewed as the accumulation of trading directions of market orders q_t, i.e. $(1-L)I_t = q_t$.The observed price is again $p_t = m_t + cq_t$. It then follows that

$$p_t = m_t + c\left(\frac{1-L}{1-\gamma L}\right)\varepsilon_t$$

The above model implies that the noise is a noninvertible ARMA process with no permanent effect on prices.

5. Asymmetric Information

This model follows the intuition in Glosten and Harris (1988). The evolution of the efficient price is given by $m_t = m_{t-1} + \omega_t$. The increments to the efficient prices are driven by (i) new public information which are not associated with trading (denoted by ε_t), and (ii) order flow q_t which are partly generated by the informed traders. In particular, $\omega_t = \lambda q_t + \varepsilon_t$, where q_t and ε_t are uncorrelated. The actual price is $p_t = m_t + cq_t$. It can be shown that in this model the price process will be given as

$$p_t = m_t + c(a\omega_t + e_t)$$

Where $a = \frac{\lambda \operatorname{var}(q_t)}{\lambda^2 \operatorname{var}(q_t) + \operatorname{var}(\varepsilon_t)}, e_t = \frac{(1-a\lambda)\omega_t - \varepsilon_t}{\lambda}$ and e_t and ω_t are uncorrelated. In this model, the microstructure noise is time independent but is correlated with the increment to the efficient prices.

6. Asymmetric Information and Autocorrelated Order Flow

Madhavan, Richardson and Roomans (1997) build a structured model for both asymmetric information and auto-correlated order flow. m_t denotes the post-trade expected value of the stock conditional upon public information and the trade initiation variable q_t.[4] The revision in beliefs ω_t is the sum of the change in beliefs due to new public information and order flow innovations, so that $m_t = m_{t-1} + \theta(q_t - E(q_t \mid q_{t-1})) + \varepsilon_t$. The transaction price is expressed as $p_t = m_t + cq_t$, where c captures the temporary effect of order flow on prices. Note we do not include the additional source of noise of p_t capturing the effect of stochastic rounding errors induced by price discreteness or possibly time-varying returns. q_t follows a Markov process where $q_t = \rho q_{t-1} + \xi_t$. The model can be reorganized as

$$p_t = m_t + c\left(\frac{a\omega_t + e_t}{1-\rho L}\right)$$

Where $a = \dfrac{\lambda \operatorname{var}(\xi_t)}{\lambda^2 \operatorname{var}(\xi_t) + \operatorname{var}(\varepsilon_t)}, e_t = \dfrac{(1-a\theta)\omega_t - \varepsilon_t}{\theta}$. It can be shown that e_t and ω_t are uncorrelated.

Therefore, the model implies both time dependence of the noise and price noise correlation.

III. Econometric Model

Suppose we want to use tick by tick data to forecast the volatility of returns over the next certain period of time T. Let r_i be the ith return over the period, and t_i be the time of the ith trade. The duration for return r_i is $d_i = t_i - t_{i-1}$. Then the T-period return is simply $\Sigma_{i=1}^{n-1} r_i$ where n is the

[4] Notations used here are different from the original paper.

stopping time such that the cumulative duration is bigger than T for the first time. If we assume that r_i is i.i.d. and independent of n [Ross (1996)], then we can simply apply Wald's theorem to obtain the forecasted volatility of T-period return given all past information F_0, $\mathrm{var}(\Sigma_{i=1}^{n-1} r_i \mid F_0) = E(n-1 \mid F_0) Var(r_1 \mid F_0)$.

However, the characteristics of the financial markets complicate the problem in at least two levels. First, the observed high frequency returns will be serially correlated for the reason implied by the microstructure theory. Second, news arrival process and trading frequency could be inter-dependent, thus the stopping time n and returns may not be independent. Therefore the forecasting problem depends on the joint distribution of durations and tick by tick returns.

Following Engle (2000), we specify the joint distribution in two steps: first we model the distribution of current duration conditional on information about past returns and durations under the ACD framework; then we model the distribution of current trade return conditional on past information as well as its contemporaneous duration.

1. ACD model for duration

We use ACD model proposed by Engle and Russell (1998) for the conditional distribution of duration, in particular

$$d_{t,i} = E(d_{t,i} \mid F_{t,i-1}) \xi_i \qquad (3.1)$$

where ξ_i ~iid with E(ξ)=1. The expected duration has both deterministic and stochastic components. One important deterministic component is time of day effect, which can be formulated as a multiplicative function to the stochastic part.

The stochastic component of the conditional distribution adapts a GARCH process which could potentially depend on past returns.

$$\Psi_{t,i}(\hat{d}_{t,i-1}) = \alpha_d + \Sigma_{j=1} p_j d_{t,i-j} + \Sigma_{j=1} q_j \psi_{i-j} + \Sigma_{j=1} \gamma_j \, | r_{i-j} | \tag{3.2}$$

where $\hat{d}$ is the seasonally adjusted duration.

We use Generalized Gamma as the distribution of innovation term ξ_i , i.e.

$$f_\xi(\xi_i) = \frac{a\lambda^{am} \xi_i^{am-1} \exp\left[-\left(\lambda \xi_i\right)^a\right]}{\Gamma(m)} \tag{3.3}$$

The generalized Gamma reduces to Weibull when m=1, to the two-parameter Gamma distribution when α=1, and to the Exponential model when α=m=1.

2. Return distribution conditional on current duration and past information

In this section, we propose a parametric model for the distribution of returns conditional on its contemporaneous duration and past information. There are two issues mainly considered: first, how to extract information about the unobserved underlying efficient price process from the observed trading prices. Second, how duration should enter the conditional density.

In our paper, we model return as a continuous variable, but in reality, prices change by tick size, so return should be a discrete variable. The discreteness of return is most significant before 1997 when tick size is 1/8 of a dollar and most of the price changes are just one or two tick sizes. However, since January 2001, the tick size has been reduced to a penny, so modeling return as a continuous variable should be less of a problem.

Following the convention in the literature, we model the observed log price $p_{t,i}$ for the ith trade at date t as the sum of the efficient price $m_{t,i}$ and a microstructure noise $u_{t,i}$. In particular,

$$p_{t,i} = m_{t,i} + u_{t,i} \tag{3.4}$$

The efficient price follows a martingale

$$m_{t,i} = m_{t,i-1} + \sigma_{\omega(t,i)}\omega_i \tag{3.5}$$

where ω_i follows an i.i.d standard normal distribution.

$\sigma_{\omega(t,i)}\omega_i$ reflects new information incorporated in the efficient price from the ith trade on date t. The volatility of the efficient price innovation $\sigma_{\omega(t,i)}$ can be time-varying and is captured by 4 components in our model as following:

$$\sigma_{\omega(t,i)} = \sqrt{h_t s_{t,i} g_{t,i} d_{t,i}{}^{\delta}} \tag{3.6}$$

h_t is the forecasted daily volatility from information up to date t-1, which captures a relatively long term effect (past several days' information); $s_{t,i}$ is the time of day effect of ith trade at date t; $g_{t,i}$ is the forecasted volatility for the ith return conditional on information up to (i-1)th trade of date t, which captures the short term effect (past several trades) on volatility; $d_{t,i}$ is the duration from (i-1)th to ith transaction measured in the fraction of a day, and finally δ is the parameter governing the speed of information arrivals. δ is bigger than/equal to/smaller than 1 if information is incorporated faster/equal/slower than linearly in time.

There is tremendous flexibility in modeling the first 3 components of volatility. For example, for the daily volatility h, one can use the implied daily volatility from the options market, or one can use a GARCH type of volatility; in modeling the time of day effect s, one can use a spline function or a step function of time. In our paper, we use GARCH processes for both daily volatility and tick volatility, and an exponential spline function for daily seasonal effect. The detailed specifications are the following[5]:

$$h_t = c_1 + c_2 h_{t-1} + c_3 r_{t-1}^2 \quad (3.7)$$

$$g_{t,i} = \rho + \beta g_{t,i-1} + \alpha \frac{{r_{t,i-1}}^2}{h_t s_{t,i-1} {d_{t,i-1}}^{\delta}} \quad (3.8)$$

$$s_{t,i} = cs_0 \cdot \exp(\sum_{j=1}^{6} cs_j (\tau_{t,i} - \tau_j^0)^+) \quad (3.9)$$

where r_{t-1} denotes the daily return for date t-1.

We next move on to the model for microstructure noise $u_{t,i}$. The previous section has shed light on some characteristics of the noise. First, it should be stationary since the observed price is most likely to be co-integrated with the underlying efficient price. Second, it should be allowed to correlate with the efficient price change due to asymmetric information or lagged price adjustments. Third, it could be auto-correlated but it may not have a finite moving average representation. Based on these arguments, we make the following assumption for the microstructure noise. There are two components of the noise variable. One is correlated with the innovation of efficient price. We call it "informational component" since it carries some information about the underlying price process. The other part, which we call "non-informational" component, is mostly due to transaction cost or inventory control and hence is independent of the efficient price. We allow time-dependency in both components. In particular,

[5] To make the model identifiable, we impose several parameter restrictions: $E(h_t) = \text{var}(r_t^2)$, $E(s_{t,i}(d_{t,i})^{\delta}) = 1$.

we model them as two ARMA processes and use the Akaike Information Criterion to pick the orders for the processes.

$$u_{t,i} = \frac{\theta_0 + \theta_1 L + ... + \theta_{q_1} L^{q_1}}{1 - \phi_1 L - ... - \phi_{p_1} L^{p_1}} \sigma_{\omega(t,i)} \omega_i + \frac{1 + B_1 L + ... + B_{q_2} L^{q_2}}{1 - A_1 L - ... - A_{p_2} L^{p_2}} \eta_i \tag{3.10}$$

where η_i ~i.i.d. normal (0, Ω), and E ($\eta_i \omega_i$) =0

To insure identification of the model, we assume that there are no common roots between $\theta_0 + \theta_1 z + ... + \theta_{q_1} z^{q_1} = 0$ and $1 - \phi_1 z - ... - \phi_{p_1} z^{p_1} = 0$, same is true for $1 - A_1 z - ... - A_{p_2} z^{p_2} = 0$ and $1 + B_1 z + ... + B_{q_2} z^{q_2} = 0$.

Lastly, since we are interested in the distribution of returns, we take the first difference of both sides of equation (3.4) to express everything in terms of returns. Let $r_{t,i}$ be the observed return for transaction i at date t, then

$$\begin{aligned} r_{t,i} &= p_{t,i} - p_{t,i-1} = \sigma_{\omega(t,i)} \omega_i + (1-L) u_{t,i} \\ &= \frac{\tilde{\theta}_0 + \tilde{\theta}_1 L + ... + \tilde{\theta}_{r_1 - 1} L^{r_1 - 1}}{1 - \phi_1 L - ... - \phi_{p_1} L^{p_1}} \sigma_{\omega(t,i)} \omega_i + \frac{1 + \tilde{B}_1 L + ... + \tilde{B}_{q_2+1} L^{q_2+1}}{1 - A_1 L - ... - A_{p_2} L^{p_2}} \eta_i \end{aligned} \tag{3.11}$$

where $r_1 = \max(p_1 + 1, q_1 + 2)$, $\tilde{\theta}_0 = 1 + \theta_0$, $\tilde{\theta}_1 = \theta_1 - \theta_0 - \phi_1$, $\tilde{\theta}_2 = \theta_2 - \theta_1 - \phi_2 ...$ and $\tilde{B}_1 = (B_1 - 1)$, $\tilde{B}_2 = (B_2 - B_1)...$

IV. Estimation -- Kalman Filter Specification

Model (3.4)-(3.11) can be estimated using the Kalman Filter technique. Appendix A shows the equivalency of the structured model and the following state space model.

Let the state equation be:

$$\begin{bmatrix} \chi_{i+1} \\ {\scriptstyle (\kappa_1 \times 1)} \\ \zeta_{i+1} \\ {\scriptstyle (\kappa_2 \times 1)} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{\kappa_1} & & & & \\ 1 & 0 & \cdots & 0 & & 0 & & \\ \vdots & \ddots & \ddots & \vdots & & (\kappa_1 \times \kappa_2) & & \\ 0 & \cdots & 1 & 0 & & & & \\ & & & & A_1 & A_2 & \cdots & A_{\kappa_2} \\ & 0 & & & 1 & 0 & \cdots & 0 \\ & (\kappa_2 \times \kappa_1) & & & \vdots & \ddots & \ddots & 0 \\ & & & & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \chi_i \\ \zeta_i \end{bmatrix} + \begin{bmatrix} \sigma_{\omega(t,i+1)} \omega_{i+1} \\ 0 \\ \vdots \\ 0 \\ \eta_{i+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4.1}$$

and the observation equation be:

$$r_{t,i} = \begin{bmatrix} \tilde{\theta}_0 & \tilde{\theta}_1 & \cdots & \tilde{\theta}_{\kappa_1 - 1} & 1 & \tilde{B}_1 & \cdots & \tilde{B}_{\kappa_2 - 1} \end{bmatrix} \begin{bmatrix} \chi_i \\ \zeta_i \end{bmatrix} \tag{4.2}$$

where $\kappa_1 = \max(p_1 + 1, q_1 + 2)$, $\phi_{\kappa_1} = 0$ for $\kappa_1 > p_1$, $\tilde{\theta}_{\kappa_1} = 0$ for $\kappa_1 > (q_1 + 1)$ and $\kappa_1 > p_1$, $\kappa_2 = \max(p_2, q_2 + 2)$, $A_{\kappa_2} = 0$ for $\kappa_2 > p_2$, $\tilde{B}_{\kappa_2} = 0$ for $\kappa_2 > (q_2 + 1)$.

Assuming normality of ω_i and η_i , we obtain the distribution of $r_{t,i}$ conditional on $\left\{ \tilde{r}_{t,i-1}, d_{t,i}, \tilde{d}_{t,i-1} \right\}$, and therefore the parameters can be estimated using maximum likelihood method as in Hamilton (1994).

The specification of the model for microstructure noise U is general enough to incorporate a majority of the market microstructure theories; however it comes at a cost of large number of parameters to be estimated. In particular, U is a combination of 2 independent unobserved ARMA processes, which may raise concerns that they cannot be identified from a single time series of returns r. We follow the idea of Hamilton (1985) to prove that the model is identified based on the implicit restrictions that are imposed on the ARMA representation of r_i. The key reason why the 2 processes can be disentangled in our model is that the informational component of U has time varying volatility and the non-informational component's volatility is constant over time. This assumption is not only crucial for the identification of the model, but also economically sensible. The proof of identification is in appendix B.

V. Application: Estimating Volatility Using Tick by Tick Data

1. Data

This section applies the model to transaction data from TAQ database. We randomly pick 10 stocks traded on NYSE with different trading frequencies. The sample period is from Jan, 2003 to May, 2003. All trades before 9:30 AM or after 4:00pm are discarded. To take out the overnight duration effect, the first trade after 9:30 for each day is excluded. For transactions that happen at the same time, we take the transaction size weighted price as the price for that time and remove all zero durations. To filter out data errors, we exclude observations where the difference between price and mid-quote is larger than 1/3 of mid-quote since extremely large magnitude returns for a single trade are very unlikely. Finally, we only include observations whose correction indicator variable has the value of 0 or 1. Returns are calculated by the first difference of logged prices. Returns are measured in units of basis points and duration in seconds. We use the daily holding period return from CRSP to estimate the daily GARCH process h_t.

Table1 reports the summary statistics of the datasets. The first column of the table gives the number of observations during the sample period for each stock. Our sample ranges from relatively illiquid stocks to fairly liquid ones. The most illiquid stock in our sample is Cedar Fair LP(FUN), with 110 trades per day on average; the most liquid stock is IBM, which trades more than 4000 times a day. The mean of the durations is always less than its standard deviation, suggesting over-dispersion relative to an exponential distribution; thus a Weibull or a generalized gamma distribution might give better fit for the data. Lastly, the means of returns of the stocks are all very close to zero relative to their standard deviations, so we force all the returns to have 0 means in our empirical estimation.

2. Parameter Estimation

The measures that we are mostly interested in this paper are the dependence of tick volatility on duration δ, and how much microstructure noise contaminates the efficient price. Table 4 summarizes these estimates for each dataset. These results shed light on existing microstructure theory and time series modeling for high frequency data, which we will discuss in more detail in the next section.

Although the other parameters are equally indispensable for the model, they either have been studied in great detail in other papers, such as parameters in the ACD model, or they are mainly statistical instruments for better fitting the data, among them are the time of day effect and daily volatility h. Therefore, we will only report the full set of estimation results for one of the companies—ASL to discuss the general properties of the model.

First, the seasonality patterns of duration and volatility are plotted in Figures 1 and 2. Two spline functions are used to adjust for the daily seasonality of duration and volatility respectively. We apply a linear spline function to adjust for the time of day effect for duration, and an exponential linear spline for tick by tick volatility, with nodes set on each hour. Figure 1 is the nonparametric estimate of the daily pattern for duration, which shows a clear inverted "U" shape similar to Engle and Russell (1998), suggesting that the trading frequency is higher at the beginning and toward the end of each day. Figure 2 shows the daily pattern of volatility of return. Tick by tick volatility also shows a "U" shape, suggesting the stock tends to be more volatile at the beginning and toward the end of the day even in tick time.

Parameter estimates for the ACD model are presented in table 2, while parameter estimates for the Kalman Filter model are in table 3. First, we find that ACD (3, 1) fits the duration data satisfactorily: the residual from the model has a mean insignificantly different from 1. (P-value =0.9888), and the Ljung-Box statistics show that the autocorrelation and partial autocorrelation until the 15th lag are all insignificant. Both estimators for a and m in (3.3) are significantly different from 1, suggesting generalized gamma is a better fit than exponential distribution in order to capture the over-dispersion in durations. Figure 3 graphically tests the goodness of fit of the generalized gamma model for duration. The probability plot of standardized duration falls narrowly along the line, suggesting generalized gamma is a reasonable distributional assumption for duration.

For the Kalman Filter model of returns, we use AIC and Likelihood ratio tests to determine how many ARMA terms should be included for the two components of microstructure noise. Both informational and non-informational components have significant loadings on microstructure noise, suggesting multiple sources for bid-ask spread. The model chosen is an ARMA (1, 1) process for the informational component and an AR (1) process for the non-informational component. On the one hand, microstructure noise is correlated with the efficient price innovation for the reason of asymmetric information or lagged price adjustment; on the other hand, transaction cost and inventory control by dealer bring an independent component to the bid-ask spread.

VI. Implication for Realized Volatility Estimation

1. Information from Duration

Our model examines the dependence of tick by tick volatility of the efficient price innovation on duration between trades, which is summarized by parameter δ. Table 4 summarizes the estimated δ for each dataset. δ with a value of 1 suggests that news is incorporated into price linearly in time, which translates to volatility over a fixed interval is independent of number of trades. This is consistent with the standard assumption that price follows an unobserved continuous process with a Brownian motion innovation, and the observed price is just a random sample of this process at discrete times. If this is the case, then forecasting of the volatility over the next trade is just the forecasted duration till the next trade times the expected per-second volatility. Appealing as it looks, however, our estimation results do not support the above assumption. All the data give estimates of δ significantly less than 1. Therefore it is important to model the joint dependence between the instantaneous volatility and the duration if one wants to forecast tick by tick volatility from a continuous time model.

5 out of 10 estimates are insignificantly different from 0. This suggests that tick-time stationarity is a better description for high frequency financial data than stationarity in wall-clock time. In other words, tick by tick volatility might have a shorter memory than volatility over a fixed time interval, therefore it is more appropriate to build a parsimonious model on tick by tick data. In order to forecast volatility over a certain period, one can use tick by tick data to forecast the number of trades in that period. If tick by tick volatility is relatively stable, the higher is the number of trades, the higher the total volatility will be. Note this will give an additional reason for volatility to be time varying, which traditional, equally spaced time series models, tend to ignore. Therefore, taking into account trading frequency will render a more efficient forecast of volatility, especially when the forecasting horizon is not tremendously larger than the average duration between trades.

δ with a value bigger/smaller than or equal to 0 can also shed light on microstructure theory about the trading behavior in the market. Trading intensity and trading volume are two sides of the same coin. Suppose informed traders receive private information about a stock. They can either trade a few large blocks of securities or divide the large blocks into smaller sizes and quickly trade out their position. Easley and O'Hara (1987) model this trade size decision by informed traders. In their model, equilibrium can be either a separating or a pooling one. In the latter case, the informed follow a strategy of sometimes breaking up their trades so that they will reduce price impacts when trading with dealers. We argue that the trading behavior in pooling equilibrium provides one possible explanation of why δ is zero in most of our datasets. In such a case, the amount of information incorporated into prices in each trade is likely to stay stable given the constant trading volume. Therefore, the total amount of news that informed investors have will determine how many smaller trades they need to submit; in other words, trading intensity will be higher/lower when the total information is higher/lower. A separating equilibrium tends to be reached only when uninformed traders are willing to trade large quantities and when a market has a low probability of informed trading. These conditions can be hard to satisfy by small firms since their operations are more opaque and receive less attention from equity analysts. These firms are also traded less frequently. Table 4 shows that all firms with number of trades less than 200,000 have δ equal to either zero or a small negative number, while larger firms such as IBM and Boeing Airline have δ significantly bigger than zero.

The finding that δ is negative for some firms is surprising and may be subject to specific sample period. However, we provide one possible economic explanation: there may be uninformed positive feedback investors trying to follow the footsteps of informed traders, so for example, whenever good news comes, both the informed and uninformed investors submit buy orders, so that information incorporation will be accelerated. If informed traders are aware of such trend chasing behavior by the uninformed, they would like to trade ahead of the uninformed. The bigger the news, the more aggressively the informed traders will submit their orders and more feedback traders will follow, and hence more news will be incorporated into price for a single trade. Therefore the duration between trades and volatility of tick returns will be negatively correlated. This is more likely to happen when the securities are believed to have a lot of private information, where uninformed traders are more likely to benefit from following the informed traders. Note the overall trading frequencies will be low for such firms since many uninformed traders will simply stay away from them. This is consistent with our findings that the two negative δ companies have relatively low number of trades.

In contrast, more frequently traded securities render higher estimates of δ. In our sample, the most heavily traded stock--IBM also has the highest estimate of δ. δ with a value larger than 0 suggests that the trading intensity increases less than linearly with the amount of news, so that per trade volatility will be higher for longer durations. This tends to happen to large firms where uninformed investors dominate the informed. Investors often trade large firms' stocks for reasons other than the firms' own news; for example, S & P component stock prices tend to move with S & P index closely, therefore investor trading intensity is less sensitive to the firm specific news. Note the above discussions are based on casual conjecture; more elaborate theoretical models capturing both the random occurrence of news events and investors' decision on trading intensity need to be built in order to make any final conclusion. But our empirical finding might provide some insights on how to build such a model.

Finally, the finding that tick by tick volatility is homogenous of degree 0 to duration is coherent with several other empirical findings. First, using equally spaced data, Jones, Kaul and Lipson (1994) finds that "the positive volatility-volume relation actually reflects the positive relation between volatility and the number of transactions. Thus, it is the occurrence of transactions, per se, and not their size, that generates volatility". This is consistent with our conjecture that informed investors tend to break down large block of trades into a sequence of smaller ones, and then information is absorbed into prices little by little at a roughly constant magnitude per trade. Engle and Russell (1998) find that low inter-trade duration is associated with high average volatility per second, and our results suggest that the two quantities tend to be inversely related. The negative association between instantaneous volatility and duration is also found in Renault and Werker (2005), and the magnitude of the negative coefficient is bigger for less liquid stocks, which corresponds to our findings that illiquid stocks tend to have smaller δ s.

2. Implication of the Noise Structure on Nonparametric Estimators

One interesting question is how much the observed prices are contaminated by the microstructure noise if one wants to use tick by tick data to estimate volatility. Table 5 summarizes the ratio of average conditional volatility for efficient price innovation per trade and the unconditional volatility of observed returns per trade $\frac{E(\sigma^2_{\omega(t,i)})}{\text{var}(r_{t,i})}$. We find a lot of variation in the ratio. Overall our estimates of the volatility of the efficient price innovation $\sigma^2_{\omega(t,i)}$ tends to be smaller than the realized volatility, suggesting that using realized volatility for high frequency data might overestimate the true underlying price variation. The degree of such contamination can be as high as 47% as in some of the stocks, so for these stocks it is a must that microstructure noise be filtered out first if one wants to use high frequency data.

There have been several other methods trying to estimate the efficient price volatility out of the observed prices, therefore it will be interesting to see how our estimate compares to the existing ones. Since all of previous measures are based on equally spaced data, we aggregate the tick by tick volatility over a day for comparison.. Table 6 calculates the ratio between daily volatility of the efficient price change and the daily realized volatility. This ratio essentially provides the same information as the ratio between tick volatility in the last column of table 5. However, one interesting observation is that the former measures tend to be less than the latter, suggesting a positive correlation between informativeness of price (ratio of daily volatilities) and the number of trades during a day. In comparison, we also calculated 4 other realized volatilities. The third column is the ratio when the Roll model is assumed. The column "30 minutes" computes the ratio between daily realized volatility using data sampled at every 30 minutes and every tick. The column "HL" adapts the method introduced by Hansen and Lunde (2004) to compute the one day bias free realized volatility, and the last column is the first-best realized volatility measure in Zhang Mykland and Ait-Sahalia (2004). Our measure is shown to be highly correlated with the others, which are reported in the last row of Table 6. The 5 ratios vary because of different assumptions. On average, 30 minutes return gives the highest ratio and ZMA renders the lowest, and the ratio from our paper comes in between. Note, our measure conforms most closely with the Hansen and Lunde measure in terms of magnitude, which may be due to the fact that our assumptions on the microstructure noise are the closest to theirs. In their measure, although they do not allow for infinite moving average in the correlation structure for the noises, it seems their measure is not hurt much by this simplifying assumption even if we identify an AR structure for most of the stocks.

Since we have identified the whole structure of the microstructure noise, we are able to study its properties in more detail than the previous literature. We are particularly interested in two questions: first whether the noise is correlated with the information about the efficient price; and second how important is this informational component? In answering the first question, the first column in Table 5 shows the model of microstructure noise picked by AIC for each dataset: all 10 stocks require both the informational components and non-informational to explain the noise. The relative importance of the informational components is given in the last column in Table 7. On average, the informational components explain 44.8% of the total variation of the microstructure noise. Our results conform with Stoll (1989) which uses a totally different approach based on serial correlation of transaction and quote prices. He analyzes NASDAQ stocks and concludes that 43% of the spread is due to adverse selection, 10% is due to inventory cost and 47% is due to order processing costs. However, Huang and Stoll (1997) use a more general model on 19 actively traded NYSE stocks to conclude that 10% of the spread is due to adverse selection, 29% is due to inventory cost and 62% is due to order processing costs. The magnitude difference between our measures and their measures are worth further investigation, but one thing to note is that their sample firms are more liquid than most of the firms in our sample.

VII. Conclusion

The paper proposes an econometric model for the joint distribution of tick by tick return and duration, with microstructure noise explicitly filtered out. We can easily forecast volatility of returns over any arbitrary time interval through simulation using all the observations available. We take into account the dependency of returns on duration when forming the forecast, therefore avoiding the unnecessary efficiency loss from transforming the data into equally spaced ones. Interestingly, we find that for most of the data, tick by tick volatility is homogeneous of degree zero to duration, suggesting stationarity in transaction time. This has implications for both empirical modeling of high frequency data and theory on market microstructure.

The specification for microstructure noise in our model is general enough to encompass most of the models from microstructure theory, and the estimation results suggest that both asymmetric information and fixed transaction cost are important resources for bid-ask spread. Moreover, transaction prices can be contaminated by the noise to a great extent, and the degree of such contamination varies from stock to stock.

One point we want to make clear is that modeling return conditional on duration does not mean that duration is an exogenous process set before price. In reality, trading frequency and volatility should be contemporaneously determined, our modeling of return as conditional distribution is only a strategy to obtain joint distributions. Equally interesting, one could also go from price process first, and model duration conditional on its contemporaneous return. And finally, a bi-variate state space model for both return and duration could make the source of dependency between the two variables more specific. It would be interesting to compare the results from the three models, and see to what extent can duration and price be isolated.

References

Admati, Anat R., and Paul Pfleiderer, 1988, A Theory of Intraday Patterns: Volume and Price Variability, *The Review of Financial Studies* 1, 3-40.

Amihud, Yakov, and Haim Mendelson, 1980, Dealership market: Market making with inventory, *Journal of Financial Economics* 8, 31-53.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, Journal of Financial Economics 17, 223-249.

Anderson, Brian D.O. , and John B. Moore. 1979, *Optimal Filtering*. (Englewood Cliffs, N.J.:Prentice-Hall).

Andrew Lo, and A. Craig MacKinlay, 1990, An econometric analysis of nonsynchronous trading, *Journal of Econometrics* 45, 181-221.

Diamond, D., and R. Verrecchia, 1987, Constraints on short-selling and asset price adjustments to private information, *Journal of Financial Economics* 18 277-311.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph Paperman, 1996, Liquidity, information and infrequently traded stocks, *Journal of Finance* 51, 1405-36

Easley, David, and Maureen O'Hara, 1987, Price, Trade Size and Information in Securities Markets, *Journal of Financial Economics* 19, 69-90.

Engle, Robert F., 2000, The Econometrics of Ultra-High Frequency Data, *Econometrica* 68, 1-22.

Engle, Robert F., and Jeffery R. Russell, 1998, Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, *Econometrica* 66, 1127 - 1162.

Frijns, Bart and Schotman, Peter 2005, Price discovery in Tick Time, *CEPR Discussion paper*.

Ghysels, Eric, and Joanna Jasiak, 1998, GARCH for irregularly-spaced financial data: the ACD-GARCH model, *Studies in Nonlinear Dynamics and Econometrics* 2, 133-50.

Glosten, L.R. 1989. Insider Trading, Liquidity, and the Role of the Monopolist Specialist. *Journal of Business* 62, no. 2: 211-35.

Glosten, Lawrence R., 1994, Is the electronic open limit order book inevitable?, *Journal of Finance* 49, 1127-61.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Grammig, Joachim, and Marc Wellner, 2002, Modeling the interdependence of volatility an inter-transaction duration processes, *Journal of Econometrics* 106, 369-400.

Hamilton, James D., 1985. Uncovering Financial Market Expectations of Inflation, *Journal of Political Economy* 93, 1224-1241

Hamilton, James D., 1994. *Time Series Analysis* (Princeton University Press).

Hansen, P. R. & Lunde, A. 2004, An unbiased Measure of Realized Variance, Brown University Working Paper.
Hasbrouk, Joel, 1991, Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.

Hasbrouck, Joel, 1999, Trading Fast and Slow: Security Market Events in Real Time, *Working Paper*, New York University.

Ho, Thomas, and Hans Stoll R., 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47-73.

Ho, Thomas, and Hans Stoll R., 1983, The dynamics of dealer markets under competition, *Journal of Finance* 38, 1053-1074.

Huang, R and H Stoll, 1997, The components of the bid-ask spread: a general approach, *Review of Financial Studies* 10, 995-1034.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-36.

Lancaster, Tony, 1992. *The Econometric Analysis of Transition Data* (Cambridge University Press).

Madhavan, Richardson and Roomans, 1997, Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks, *The Review of Financial Studies* 10, No. 4, 1035-1064.

O'Hara, Maureen, 1995. *Market Microstructure Theory* (Blackwell Publishers Inc.).

Renault, Eric and Werker, Bas, 2005, Causality effect in return volatility measures with random tims, *Working Paper,* University of North Carolina and Tilburg University.

Roll, Richard, 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* 39, 1127-1139.

Ross, Sheldon, 1996. *Stochastic Process* (John Wiley & Sons, Inc.).

Scholes, and Williams, 1977, Estimating Beta From Nonsynchronous Data, *Journal of Financial Economics* 5, 309-327.

Stoll, Hans, 1989, Inferring the components of the bid-ask spread: Theory and empirical tests, *Journal of Finance* 44, 115-134.

Wall, Kent D, 1987, Identification Theory for Varying Coefficient Regression Models, *Journal of Time Series Analysis* 8, 359-371.

Yacine Ait-Sahalia, Per A.Mykland, 2003, How often to sample a continuous-time process in the presence of market microstructure noise, *NBER Working Paper Series*.

Zhang, L. P. A. Mykland, and Y. Ait-Sahalia, 2003, A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data, *Working Paper w10111*, NBER

Appendix

Appendix A: Transform the Structured Model for Return to State Space Model
Prove that equation (4.1)-(4.2) can be written in the form(3.11):
Let $\chi_{j,i}$ be the jth entry of the vector χ_i ,from (4.1) we have

$$\chi_{1,i+1} = \phi_1\chi_{1,i} + \phi_2\chi_{2,i} + \ldots + \phi_{r_1}\chi_{r_1,i} + \sigma_{\omega(t,i+1)}\omega_{i+1} \tag{A.1}$$

$$\chi_{2,i+1} = \chi_{1,i} = L\chi_{1,i+1}$$

$$\chi_{3,i+1} = \chi_{2,i} = L^2\chi_{2,i+1}$$

$$\vdots$$

$$\chi_{r_1,i+1} = L^{r_1-1}\chi_{1,i+1}$$

Plug $\chi_{2,i+1}, \chi_{3,i+1}, \cdots, \chi_{r_1,i+1}$ into (A.1) and rearrange the equation

$$\chi_{1,i+1} = \frac{\sigma_{\omega(t,i+1)}\omega_{i+1}}{1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_{r_1} L^{r_1}} \tag{A.2}$$

Similarly, let $\zeta_{j,i}$ be the jth entry of the vector ζ_i , then from (4.1) we have

$$\zeta_{1,i+1} = \frac{\eta_i}{1 - A_1 L - \ldots - A_{r_2} L^{r_2}} \tag{A.3}$$

Expanding Equation (4.2)

$$\begin{aligned} r_{t,i} &= (\tilde{\theta}_0\chi_{1,i} + \tilde{\theta}_1\chi_{2,i} + \ldots + \tilde{\theta}_{r-1}\chi_{r_1-1,i}) + (\zeta_{1,i} + \tilde{B}_1\zeta_{2,i} + \ldots + \tilde{B}_{r_2-1}\zeta_{r_2-1,i}) \\ &= (\tilde{\theta}_0 + \tilde{\theta}_1 L + \ldots + \tilde{\theta}_{r-1}L^{r_1-1})\chi_{1,i} + (1 + \tilde{B}_1 L + \ldots + \tilde{B}_{r_2-1}L^{r_2-1})\zeta_{1,i} \end{aligned} \tag{A.4}$$

Finally, Equation (3.11) can be obtained by plugging (A.2) and (A.3) into (A.4).

Appendix B: Identification of Parameters
According to equation(3.10), the vector of parameters to be estimated in U is

$$(\phi_1, \ldots, \phi_{p_1}, \theta_0, \theta_1, \ldots, \theta_{q_1}, A_1, \ldots, A_{p_2}, B_1, \ldots, B_{q_2}, E(\sigma^2_{\omega(t,i)}), \sigma^2_\eta$$

a total of ($p_1+q_1+p_2+q_2+2$) parameters. Therefore we need at least ($p_1+q_1+p_2+q_2+2$) linearly independent equations to identify the model. Define

$y_i = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_{p_1} L^{p_1})(1 - A_1 L - \ldots - A_{p_2} L^{p_2}) r_i$, which is the moving average component of r_i, then from equation (3.11), we have

$$
\begin{aligned}
y_i &= (1-\phi_1 L - \ldots - \phi_{p_1} L^{p_1})(1 - A_1 L - \ldots - A_{p_2} L^{p_2})\sigma_i \omega_i \\
&+(1 - A_1 L - \ldots - A_{p_2} L^{p_2})(1-L)(\theta_0 + \theta_1 L + \ldots + \theta_{q_1} L^{q_1})\sigma_i \omega_i \\
&+(1-\phi_1 L - \ldots - \phi_{p_1} L^{p_1})(1-L)(1 + B_1 L + \ldots + B_{q_2} L^{q_2})\eta_i
\end{aligned}
$$

Since the model is estimated based on the distribution assumption of the return series, we would be able to identify the parameters if there are no 2 sets of parameter values yielding the same moments of returns. From theorem E of Hannan (1971) as interpreted by Preston and Wall (1975), the auto-regressive coefficients of the observed return r_i can be estimated separately from the moving average terms if there are no common roots between the autoregressive and the 2 component of the moving average determinant polynomials (C1):

$$
\begin{aligned}
&(1-\phi_1 z - \ldots - \phi_{p_1} z^{p_1})(1 - A_1 z - \ldots - A_{p_2} z^{p_2}) = 0 \\
&\left[(1-\phi_1 z - \ldots - \phi_{p_1} z^{p_1})(1 - A_1 z - \ldots - A_{p_2} z^{p_2}) + (1 - A_1 z - \ldots - A_{p_2} z^{p_2})(1-z)(\theta_0 + \theta_1 z + \ldots + \theta_{q_1} z^{q_1})\right] E(\sigma_i^2) = 0 \\
&(1-\phi_1 z - \ldots - \phi_{p_1} z^{p_1})(1-z)(1 + B_1 z + \ldots + B_{q_2} z^{q_2})\sigma_\eta^2 = 0
\end{aligned}
$$

Since we assume that there is no common roots between $\theta_0 + \theta_1 z + \ldots + \theta_{q_1} z^{q_1} = 0$ and $1-\phi_1 z - \ldots - \phi_{p_1} z^{p_1} = 0$, and between $1 + B_1 z + \ldots + B_{q_2} z^{q_2} = 0$ and $1 - A_1 z - \ldots - A_{p_2} z^{p_2} = 0$, and all the roots are outside unit circle, the only situation that condition C1 is satisfied is that $1-\phi_1 z - \ldots - \phi_{p_1} z^{p_1} = 0$ and $1 - A_1 z - \ldots - A_{p_2} z^{p_2} = 0$ has common roots. In this case, the model can be simplified until there are no common roots.

The rest of the parameters can be identified by the variance and covariances of $(y_i, y_{i-1}, \ldots, y_0)$. Since ω_i and η_i are independent, the conditional covariance between

y_i and y_{i+j} at trade (i-1) is of the form $E_{i-1}(y_i y_{i+j}) = C_j \frac{\sigma_i^2}{E(\sigma_i^2)} + D_j$,where C_j is a constant determined by the parameters $\left[\phi_1,...,\phi_{p_1}, A_1,..., A_{p_2}, \theta_0,...\theta_{q_1}, E(\sigma_i^2)\right]$ and D_j is a constant determined by $(\phi_1,...\phi_{p_1}, B_1,...B_{q_2}, \sigma_\eta^2)$. The critical assumption that the efficient price innovation has time-varying volatility while the uninformational component of the microstructure noise has a constant volatility helps to identify the model. To see this, $E_{i-1}(y_i y_{i+j})$ will not be the same unless both C_j and D_j are the same. Hence there are max(p1+p2,p2+q1+1)+1 parameter restrictions for $[\phi_1,...,\phi_{p_1}, \theta_0,...,\theta_{q_1}, A_1,...A_{p_2}, E(\sigma^2_{\omega(t,i)})]$ coming from $[C_j]$ and (p1+q2+2) restrictions for $\left[\phi_1,...\phi_{p_1}, B_1,..., B_{q_2}, \sigma_\eta^2\right]$. Therefore, we are guaranteed with at least same number of equations as the number of parameters to be estimated. Construct f: $R^{q_1+2} \to R^{\max(p_1+p_2, p_2+q_1+1)+1}$ mapping $\left[\theta_0,...,\theta_{q_1}, E(\sigma^2_{\omega(t,i)})\right]$ into $\left[C_0, C_1.., C_{\max(p_1+p_2, p_2+q_1+1)+1}\right]$ and g: $R^{q_2+1} \to R^{p_1+q_2+2}$ mapping $(B_1,..., B_{q_2}, \sigma_\eta^2)$ into $\left[D_0, D_1,..., D_{p_1+q_2+2}\right]$. The Jacobian of f(.) and g(.) have full rank unless by some coincidence the different elements of parameters happen to obey particular exact numerical relations to one another. Therefore the parameters are generally identified.

For instance, suppose the informational and non-informational components of the microstructure noise admits two independent MA(1) processes, $u_i = (\theta_0 + \theta L)\sigma_i \omega_i + (1 + BL)\eta_i$, then the parameters are uniquely determined as

$$\theta_0 = -\frac{C_1 + C_2}{C_0 + C_1 + C_2},\ \theta = -\frac{C_2}{C_0 + C_1 + C_2},\ E(\sigma_t^2) = \frac{(C_0 + C_1 + C_2)^2}{C_0},\ \sigma_\eta^2 = D_0,\ B = -\frac{D_2}{D_0},$$

$$D_0 + D_1 + D_2 = 0.$$

In a word, having $\sigma_{\omega(t,i)}$ to be time-varying is a critical assumption to make the model identifiable. From its definition, the time-variability of $\sigma_{\omega(t,i)}$ are due to 4 sources: $h_t, s_{t,i}, g_{t,i}, d_{t,i}^{\delta}$. Therefore, the model is identifiable as long as at least one of the sources preserves the time-varying property, which is very likely to hold and can be easily tested.

Ticker	Number of Obs.	Mean # of trades/day	Average dur.	Std. Dev. dur.	Average ret.	Std.Dev. ret.
FUN	11280	110.59	213.6248	286.3584	0.1424	13.7352
RGR	13456	131.92	179.0769	265.6867	-0.0093	25.7334
CDI	14478	141.94	166.4223	267.1497	-0.0199	13.3400
WSO	15581	152.75	154.6483	232.5889	-0.0280	16.4393
OMM	16301	159.81	147.8363	283.8729	0.1958	24.7882
ASL	26973	233.65	89.3434	181.1853	0.0983	24.085
LUK	30038	294.49	80.2315	123.7295	0.0068	7.6016
CTX	218997	2126.18	11.0046	15.9803	0.0198	3.7429
BA	279900	2744.12	7.8862	10.5704	-0.0035	5.1970
IBM	419994	4117.59	5.7372	5.9587	0.0026	2.9096

Table 1. Summary Statistics for durations and returns for 10 randomly sampled stocks in TAQ database. Returns are measured in basis points and durations in seconds. The Sample period is from Jan, 2003 to May, 2003. All trades before 9:30 AM or after 4:00pm are discarded. The first trade after 9:30 for each day is also excluded.

Parameter	Estimate	Std.Error	Prob.
α_d	2.8376	0.3626	0.0000
p_1	0.0215	0.0104	0.0384
p_2	0.9104	0.0077	0.0000
p_3	0.1167	0.0128	0.0000
q_1	0.0575	0.0072	0.0000
γ_1	-0.1147	0.0110	0.0000
a	8.9291	1.4046	0.0000
m	0.2073	0.0166	0.0000
Statistics for Residual ξ_i			
Mean		1.0002	
Std. Dev		1.7319	
Ljung-Box		16.2839	
Prob.		0.3634	

Table 2: Parameter estimates of ACD model as in (3.1)-(3.3) for stock ASL.

Parameter	Estimate	Std. Error	Prob.
δ	0.0120	0.0172	0.4854
ρ	2.91E-04	9.82E-06	0.0000
α	0.0746	0.0075	0.0000
β	0.8544	0.0018	0.0000
ϕ_1	-0.5209	0.2901	0.0726
θ_0	0.0861	0.0135	0.0000
θ_1	0.0600	0.0174	0.0006
A_1	0.4862	0.2569	0.0584
Ω	40.5506	5.8507	0.0000
L	-1.2100e+005		
AIC	2.42E+05		

Table 3: Parameter Estimates for model (3.6)-(3.10)

$$\sigma_{\omega(i,t)} = \sqrt{h_t s_{t,i} g_{t,i} (d_{t,i})^{\delta}}$$

$$g_{t,i} = \rho + \alpha \frac{r_{t,i-1}{}^2}{h_t s_{t,i-1} (d_{i,t-1})^{\delta}} + \beta g_{t,i-1}$$

$$u_{t,i} = \frac{\theta_0 + \theta_1 L + \ldots + \theta_{q_1} L^{q_1}}{1 - \phi_1 L - \ldots - \phi_{p_1} L^{p_1}} \sigma_{\omega(i,t)} \omega_i + \frac{1 + B_1 L + \ldots + B_{q_2} L^{q_2}}{1 - A_1 L - \ldots - A_{p_2} L^{p_2}} \eta_i$$

Stock	# of obs	δ	ρ	β	α	Ω
FUN	11280	**-0.0293**	2.00E-03	0.7778	0.0614	48.3598
		[0.0050]	[1.970e-06]	[0.0364]	[0.0136]	[4.8181]
RGR	13456	-0.0352	-8.27E-05	0.8764	0.0598	0.2885
		[0.0219]	[7.316e-06]	[0.0150]	[0.0080]	[0.0165]
CDI	14478	0.0134	3.44E-04	0.9047	0.056	0.6097
		[0.0139]	[1.1429e-06]	[0.0086]	[0.0083]	[0.1281]
WSO	15580	**-0.0534**	7.33E-05	0.9193	0.0535	0.2187
		[0.0252]	[5.1400e-07]	[0.0151]	[0.0099]	[1.2332]
OMM	16301	0.0065	1.99E-04	0.9087	0.0656	34.6167
		[0.0444]	[5.0165e-05]	[0.0205]	[0.0300]	[0.5230]
ASL	23832	0.0120	2.91E-04	0.8544	0.0746	40.5506
		[0.017]	[9.8200e-06]	[0.0018]	[0.0049]	[5.8507]
LUK	30038	-0.0281	7.87E-06	0.9207	0.0414	10.5479
		[0.0240]	[4.025e-06]	[0.0081]	[0.0060]	[0.3251]
CTX	196920	**0.0912**	5.18E-06	0.9288	0.032	3.5028
		[0.0139]	[4.606e-06]	[0.0031]	[0.0021]	[0.2810]
BA	279900	**0.0718**	-1.09E-06	0.9916	0.0093	2.6005
		[0.0331]	[5.13e-07]	[0.0019]	[0.0018]	[0.3637]
IBM	419994	**0.325**	2.99E-05	0.7359	0.1017	0.4776
		[0.1164]	[7.1371e-08]	[0.0030]	[0.0146]	[0.0371]

Table 4. Estimates for selected parameters from the Kalman Filter model. δ is the volatility dependency on duration, ρ, α and β are parameters for tick by tick GARCH g. Ω is the variance for η. Robust standard errors are included in the squared brackets. Only the first 20000 observations are used in the estimation for companies BA, CTX and IBM because of the large sample size of these data.

Ticker	Model Picked for $U_{t,i}$	$E(\sigma^2_{\omega(t,i)})$	$Var(r_{t,i})$	$\frac{E(\sigma^2_{\omega(t,i)})}{\mathrm{var}(r_{t,i})}$	Upward bias
FUN	$MA_\omega(1)+MA_\eta(1)$	129.0961	158.9854	0.8120	0.2315
RGR	$MA_\omega(1)+ARMA_\eta(1,1)$	397.5943	587.5488	0.6767	0.4778
CDI	$AR_\omega(1)+AR_\eta(1)$	332.1447	169.4789	1.9598	-0.4897
WSO	$WN_\omega+ARMA_\eta(1,1)$	222.9469	254.6219	0.8756	0.1421
OMM	$ARMA_\omega(1,1)+WN_\eta$	517.1564	575.8980	0.8980	0.1136
ASL	$ARMA_\omega(1,1)+MA_\eta(1)$	475.7309	580.0889	0.8201	0.2194
LUK	$ARMA_\omega(1,1)+AR_\eta(1)$	67.2119	56.8966	1.1813	-0.1535
CTX	$ARMA_\omega(1,1)+ARMA_\eta(1,2)$	15.8280	12.6526	1.2510	-0.2006
BA	$ARMA_\omega(1,1)+ARMA_\eta(2,1)$	17.5371	19.6759	0.8913	0.1220
IBM	$ARMA_\omega(1,1)+AR_\eta(1)$	5.4576	7.7085	0.7080	0.4124

Table 5. Microstructure noises. AIC and Likelihood ratio test are used to choose the best model. $AR_\omega(1)$ stands for AR(1) process for the information component in $U_{t,i}$, $ARMA_\omega(1,1)$ means ARMA(1,1) model is picked for non-informational component. $E(\sigma^2_{\omega(t,i)})$ is the mean of conditional volatility for efficient price change. $Var(r_{t,i})$ is the unconditional variance for tick returns. Upward bias is calculated as $\frac{Var(r_{t,i})}{E(\sigma^2_{\omega(t,i)})}-1$.

Stock	$\frac{E(\Sigma_{i=1}^{n}\sigma^2_{\omega(t,i)})}{E(\Sigma_{i=1}^{n}r^2_{t,i})}$	**Roll**	**30 minutes**	**HL**	**ZMA**
FUN	0.7417	0.8201	0.6861	0.7019	0.3863
RGR	0.6734	0.7735	0.5856	0.5861	0.3846
CDI	1.8186	1.1194	1.5233	1.4632	1.0377
WSO	0.8029	0.8289	0.8583	0.8698	0.5789
OMM	0.8006	0.8384	0.8243	0.8045	0.5493
ASL	0.7005	0.8047	0.7434	0.7777	0.4411
LUK	1.0406	0.9698	0.9612	1.0461	0.7682
CTX	1.1912	0.9527	1.3914	1.2263	0.9749
BA	0.8813	0.8188	0.4556	0.3623	0.6938
IBM	0.6873	0.776	0.8506	0.8192	0.5798
Median	0.8017	0.8245	0.8375	0.8118	0.5794
Correlation with our model	.	0.9677	0.8416	0.8118	0.8869

Table 6. Efficient volatility vs. realized volatility. The ratio between one day volatility of efficient price change and one day realized volatility calculated using all the tick data. The second column is the ratio from our model, the third column is the ratio calculated from Roll(1984) model. The column "30 minutes" computes the ratio between one day realized volatility using data sampled at every 30 minutes and every tick. The column "HL" adapts the method introduced by Hansen and Lunde (2004) to compute the one day biased free realized volatility, where the sample frequency is every 1 minutes and the autocorrelation adjustment is up to 10 minutes. The last column computes the first-best realized volatility as in Zhang Mykland and Ait-Sahalia(2004), where sparse sampling is every 10 ticks. The last row reports the Pearson correlation between our measures and the other measures.

Stock	Var_Info	Var_NonInfo	Var_Info/(Var_Info+Var_nonInfo)
FUN	3.5738	23.1357	0.1338
RGR	55.0043	24.5987	0.6910
CDI	86.9459	0.0042	0.9999
WSO	1.7223	1.0748	0.6157
OMM	1.361	4.3896	0.2367
ASL	2.9086	5.0227	0.3667
LUK	4.4878	31.5427	0.1246
CTX	1.0597	0.9135	0.5370
BA	2.1149	4.8124	0.3053
IBM	0.0916	0.0812	0.5301
Median			0.4484

Table 7. Informational vs. Non-informational Components.

Figure 1 Daily Seasonality for Duration

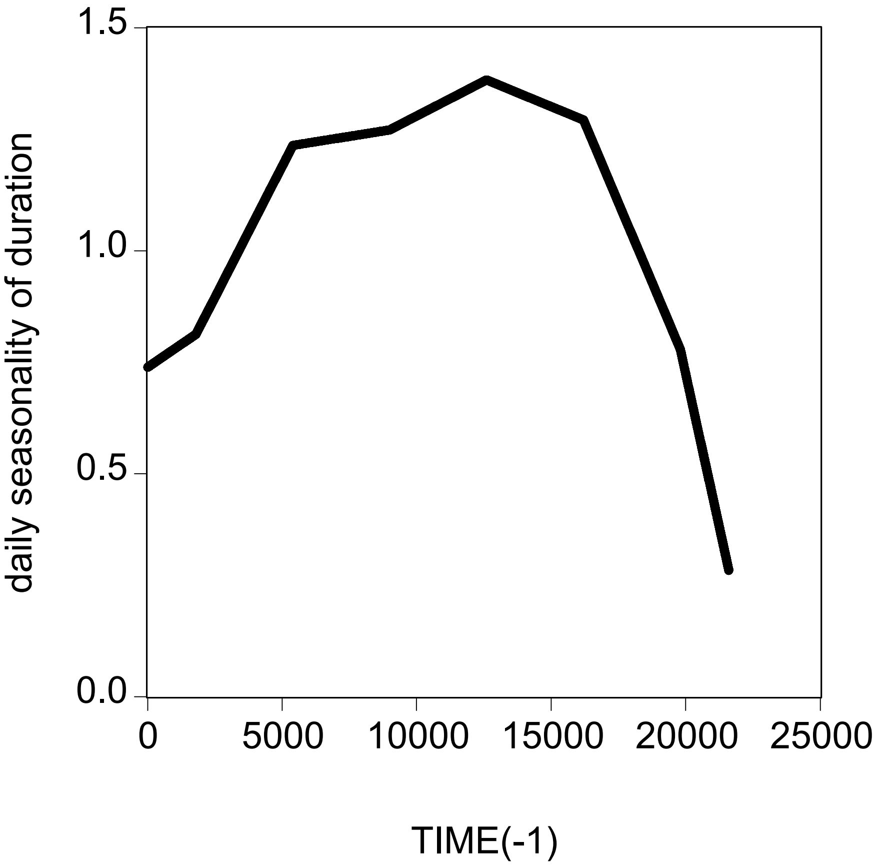

Figure 2 Daily Seasonality for Tick Volatility

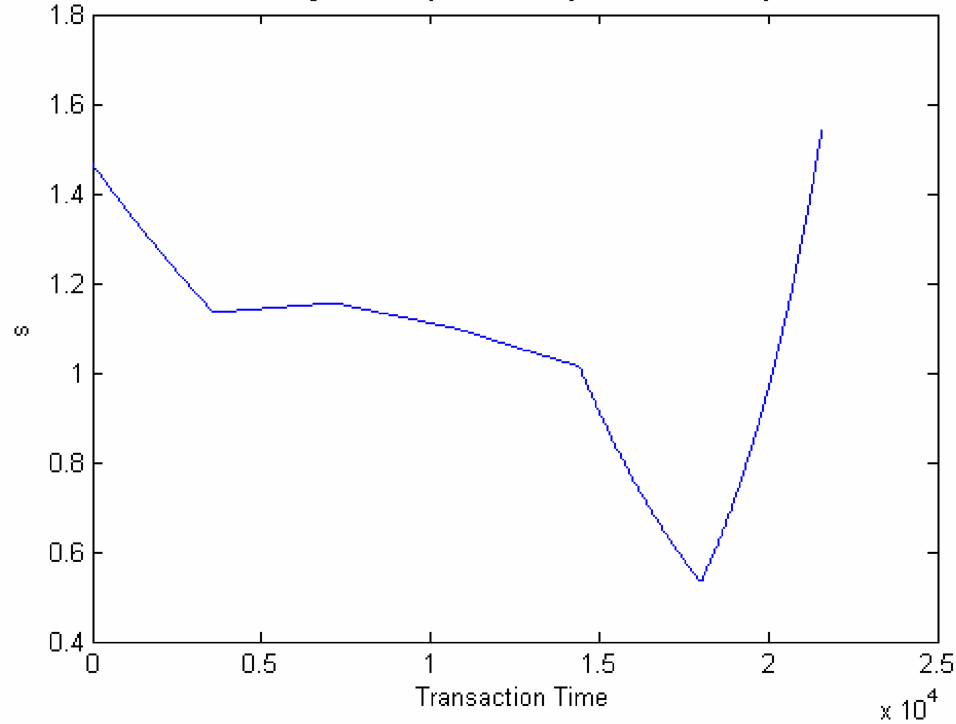

Figure 3: Generalized Gamma Probability Plot

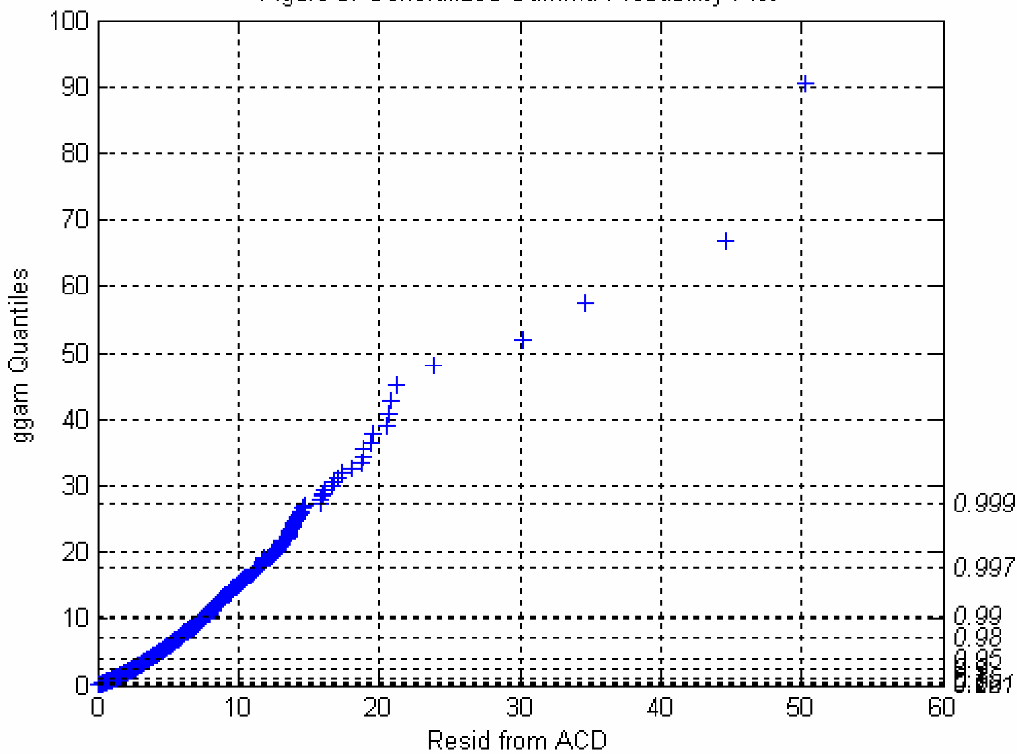

Figure 4: Time Series Plot of m(i) and p(i)

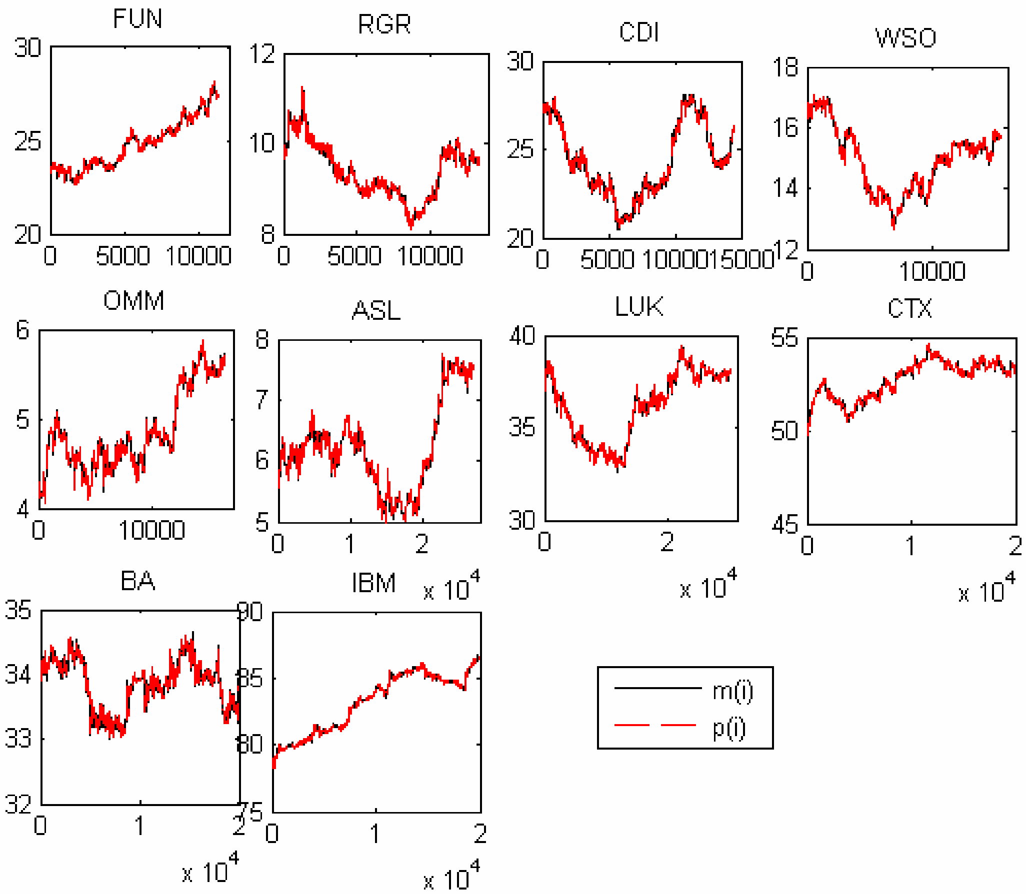

Figure 5:Time Series Plot of U(i)

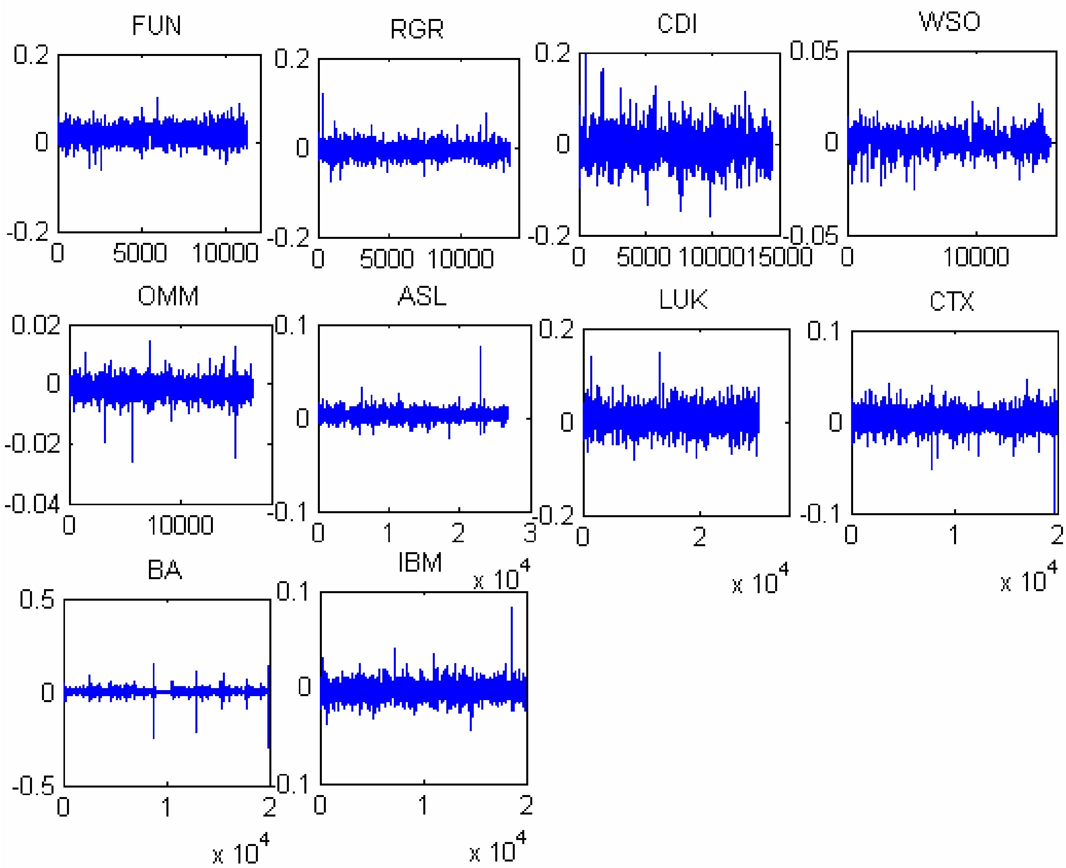

Figure 6: Daily Volatility