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

Robert Engle and Zheng Sun[1]

Abstract

This paper studies the joint distribution of tick by tick returns and durations between trades. Returns are decomposed into changes in full information prices and microstructure noise, but the noise is modeled in accordance with various models of market microstructure allowing rich correlation structures both with the efficient price and over time. The full information price has time varying volatility which depends upon the arrival time of trades. The paper aims at three contributions: First, the noise is modeled to allow asymmetric information, inventory and order processing costs, and delayed quote setting. Second, the response to the trade arrival times allows trade durations to be informative on future volatility. Third, the estimated state space models can act as a laboratory to examine various non-parametric approaches to realized volatility estimation. Both simulated and actual data can be compared across methods and the accuracy and efficiency assessed as long as the parameteric model is viewed as a sufficiently accurate representation. We apply the above model to 10 NYSE stock transactions data series with varying transaction rates. 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".

[1] Department of Finance, Stern School of Business, New York University

I. Introduction

High frequency data on asset prices offers the prospect of more accurate measures and forecasts of volatility. In theory, (add citation)[2], realized volatility computed from the highest possible frequency data should provide both a consistent and efficient estimator for integrated volatility. However, standard market microstructure theories suggest that theoretically ideal circumstances are not likely to be satisfied in the real world. For Instance, the transaction price is likely "contaminated" by so called "microstructure noise". Thus a volatility estimator that does not account for this noise may be biased. This problem is most serious in high-frequency data since the volatility of true price usually shrinks with the time interval, while the volatility of noise components such as the bid-ask spread usually does not (Yacine Ait-Sahalia (2003)). Thus in practice, researchers use moderate-frequency data to estimate volatility. This practice, although inefficient, is at least unbiased. The goal of this paper is to apply standard microstructure theory to understand the properties of the "noise" and then build a parametric model to obtain a noise-free volatility estimator using tick by tick data. This estimator will be much more efficient, if it is correctly specified, and can serve as a benchmark to evaluate the wide range of non-parametric methods that have been developed.

Existing market microstructure models usually distinguish between the transaction price and the fair market price based on perfect information. The difference between the two is often given the name of "microstructure noise". However, such noise may not just be useless noise. It may reveal something about the underlying fundamental. For example, under the classical microstructure asymmetric information model (eg. Glosten and Milgrom(1985) and Kyle (1985)), the market makers post regret-free prices. Investors trade at prices that are fair market value based on public information. In this model then, "noise" is not pure noise. Rather it reflects the private information that is not yet priced. The noise is gradually incorporated into the transaction price as

[2] Andersen and Bollerslev (1998). Merton (1980)

more and more informed trading enters the market. Therefore "noise" conveys some information about the underlying efficient price. Furthermore, noise can be auto-correlated if there is sequential informed trading; this again suggests that it may contain useful information.

In other microstructure models such as Amihud and Mendelsohn(1986), transaction prices differ from the full information price because market makers accumulate excess or depleted inventories which temporarily depress or raise prices. In other models such as Roll(1984), order processing costs drive a wedge between transaction prices and full information prices.

Whether the microstructure noise conveys information and exhibits correlation properties is ultimately an empirical question. We write down an empirical model of microstructure noise. Our empirical specification is general enough to encompass most of the sources of noise that have been studied in the literature (Stoll (1989), Huang and Stoll (1997)). In particular, we include two components of microstructure noise. The first component is fixed noise potentially due to order processing cost or inventory control by dealers. The second component is time-varying noise that is correlated with change of the efficient price, which may come from asymmetric information or stale quotes. The two noise processes are allowed to be self auto-correlated. We use Kalman Filtering techniques to estimate the model. To help identification, we specify different variance structures for the informational and the non-informational innovations.

Another reason why using high frequency data to estimate volatility may improve efficiency is that tick by tick data preserve duration information between transactions. . If information causes trading, then non-trading and price should be jointly determined by the amount of news in the market[3]. Hence they are probably correlated. This observation motivates us to model volatility of efficient price innovation as a function of duration. The autoregressive conditional duration

---

[3] Diamond and Verrecchia (1987), Admati and Pfleiderer (1988) and Easley and O'Hara (1992) have written theoretical models that have implications on how news and transaction frequency are related.

model (ACD) proposed by Engle and Russell (1998), which focuses on time elapsed between trades, forms the framework for incorporating duration information into the analysis of irregularly-spaced high frequency data. In this paper, we model the joint density of the marked point process of durations and tick by tick returns. This paper extends the ACD/UHF-GARCH framework proposed by Engle (2000). First, we model the duration variable as an ACD process that could potentially depend on past returns. Then, we model the 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 we specify (1) the distribution of duration conditional on past information (return and duration) and (2) the distribution of return conditional on current duration and past information, we can then obtain the joint distribution of return and duration. Proceeding thus, we can forecast return volatility during any arbitrary length of time using simulation.

Our paper aims to make three contributions. First, we model the noise by allowing most of the important statistical features emphasized in the microstructure literature. The noise has general autocorrelation and is allowed to be correlated in arbitrary fashion with the innovation in the efficient price. The efficient price innovation is itself allowed to have time varying volatility. Our paper finds that microstructure noise is time dependent and is informative about unobserved efficient price change. On average, the informational component accounts for 45% of the total variation of microstructure noise. One notion that distinct our paper from the current realized volatility literature is that most of the current studies assume microstructure noise to be uncorrelated with the underlying efficient price. Under such assumption, noise is always "pure" noise, and they simply add to the volatility estimator, therefore an volatility estimator without adjusting the noise is always upward biased. In contrast, our model recognize the possibility of negative correlation between noise and private information suggested by the adverse selection model, so a realized volatility estimator of the efficient price change that does not account of microstructure noise can be either upward or downward biased, depending on the correlation structure of the microstructure noise.

Second, our paper studies extensively on how one important attribute of trades – durations between trades - reflects the underlying information. The current literature pays little attention to the duration between trades. For example, in the standard Brownian Motion framework, the volatility of price increases linearly in time. This is based on the assumption that trades take place at the process independent of information. In contrast, our model allows for volatility to depend on duration nonlinearly and treats the degree of nonlinearity as a parameter to be estimated. We find that 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 infrequently 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. We also find that the nonlinear properties of duration-volatility relationship is not changed even if another important attributed of trades – sizes of trades – is incorporated into the model.

Third, our paper can serve as a benchmark or laboratory for examining the widely used models of realized volatility such as Andersen, Bollerslev, Diebold and … (citation Zhang, Mykland and Ait-Sahalia (2003)), Hansen and Lunde (2006), Russell and Bandi() . By generating simulated tick by tick data with a well specified model and corresponding measure of realized volatility, it is straight forward to examine various methods for handling data and estimating realized volatility. The accuracy and efficiency of different methods can be compared. In following this procedure, we find some important distinctions between the methods.

In terms of econometric modeling, the paper that is most 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.

Comments on microstructure empirical literature that empirically estimate the impact of asymmetric information on security prices. (Hasbrouck (1988), Harris(1990), Glosten & Harris (1988)).

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. Section four applies the model to a sample of NYSE stocks. Section five discusses implications of the results on the realized volatility estimation. Section six discusses the volume’s effect on volatility and durations, and Section seven concludes.

II. Microstructure Noise

Our paper assumes a typical price generating mechanism. Specifically, we assume an auction-dealer mechanism in which market specialists quote bid and ask prices at which they are willing to trade, and orders are executed at either the bid or the ask. In this paper, we consider the fair value of the stock conditioning on all available information (public as well as private) as the efficient price (denoted by m_t). Strong form of the efficient market hypothesis implies that m_t is a martingale process. Transaction price can often stray from the full information fair market price. First, the market maker usually holds some opinions about the fundamental value of the stock. However, she may not be willing to transact at the price level equal to this belief because of order processing cost or inventory control needs. Instead, she will ask for a price concession for both incoming buy and sell orders, making the observed transaction price different from the market maker's estimate of the efficient price. Second, even if the market maker is willing to buy and sell at her belief of the efficient price, her estimate of the efficient price may very well be different from the estimate of another market participant possessing private information. We define microstructure noise as the difference between the transaction price and the efficient price. In particular, let p_t be the transaction price and u_t be the microstructure noise, then $p_t = m_t + u_t$. Usually the observed price cannot deviate too much from the fundamental price, so u_t is assumed to be stationary. However, various sources of friction will impose different structures on u_t. It is difficult to point out which of these sources are observable and whether there is a dominant one. The section summarizes various types of frictions suggested in the literature. Understanding them will help to shed the light on properties of the various frictions and also the microstructure noise structure in our model.

1. Roll Model (1984)

In this model, the only market friction is the order processing cost of the market maker. In particular,

$$p_t = m_t + cq_t$$

where c is the cost of processing the order, and 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. The Roll model assumes that orders randomly bounce between bid and ask quotes, i.e. q_t is an i.i.d. process.

2. Stale Prices

Although it is rare cases in nowadays trading systems, theoretically if the operational systems are relatively slow compared with the speed of order submission, trades may 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 correlated 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 Milgrom(1985) and Glosten and Harris (1988). The evolution of the full information 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) private information that is partially reflected by the order flow q_t (iii) private information that is uncorrelated with current order flow η_t. In particular, $\omega_t = \lambda(q_t - E(q_t \mid q_{t-1})) + \eta_t + \varepsilon_t$, where q_t, η_t and ε_t are uncorrelated. In the

Glosten and Milgrom (1985) adverse selection model, the market maker put up a regret free price p_t, which equals the post-trade expected value of the stock conditional upon public information and the order flow information q_t. To make the effect of asymmetric information on price process clearer, we omit any transitory microstructure effect such as inventory or order processing cost. Let I_t be the market maker's information set at time t, which includes any public information ε_t and order flow information q_t

$$p_t = E(m_t \mid I_t) = m_{t-1} + \lambda(q_t - E(q_t \mid q_{t-1})) + \varepsilon_t = m_t - \eta_t$$

It can be easily seen that the microstructure noise is time independent but is negatively 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. The full information efficient price is the same as in the asymmetric information case above. Beside the permanent impact of order flow on prices through adverse selection model, there is also a 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 $q_t = \rho q_{t-1} + \xi_t$. In particular, the observed transaction price is given by

$$p_t = E(m_t \mid I_t) + cq_t = m_t - \eta_t + c\frac{\xi_t}{1-\rho L}$$

The equation can be reorganized into

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

Where $a = \frac{c\lambda E(\xi_t^2) - E(\eta_t^2)}{\lambda^2 E(\xi_t^2) + E(\varepsilon_t^2) + E(\eta_t^2)}$, $b = \frac{\rho E(\eta_t^2)}{\lambda^2 E(\xi_t^2) + E(\varepsilon_t^2) + E(\eta_t^2)}$ and

$e_t = c\xi_t - \eta_t + \rho\eta_{t-1} - a\omega_t - b\omega_{t-1}$

It can be shown that e_t follows an MA(1) process and it is uncorrelated with ω_t. Therefore, the model implies both time dependence of the noise and price noise correlation. One thing to notice is that the correlation between microstructure noise and innovation of efficient price can be either positive or negative depending on the relative magnitude of private information $E(\eta_t^2)$ and temporary price impact of order flow c. The correlation is negative if the private information not yet revealed by current order flow is relatively bigger and positive otherwise.

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 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, $\text{var}(\Sigma_{i=1}^{n-1} r_i \mid F_0) = E(n-1 \mid F_0) Var(r_1 \mid F_0)$.

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 interdependent, 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.

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

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

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

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

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[4]:

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

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

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

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

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

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

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 $\tilde{B}_1 = (B_1 - 1)$, $\tilde{B}_2 = (B_2 - B_1) ...$

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

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.

IV. 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 .

[Insert Table 1]

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), a company owns and operate amusement and water parks in the unite states, 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. 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.

[Insert Figure 1 and Figure 2]

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.

[Insert Table 2 and Table 3]

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.

[Insert Figure 3]

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 white noise process for the informational component and an ARMA (1,1) 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.

V. 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. 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 transaction take places as if it is a random draws of the underlying price process, with sampling frequency uncorrelated with whether there is information or not. 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.

[Insert Table 4]

Appealing as it looks, however, our estimation results do not support the above assumption. Table 4 summarizes the estimated δ for each dataset. All the data give estimates of δ significantly less than 1. In fact, 6 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.

(need to be changed) 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 specific to the 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 numbers 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

The general assumption of the variance structure of microstructure noise complicates the econometric techniques that are required to estimate the model. Given the existing simpler non-parametric procedures that estimate realized volatility, we would like to know how well these simpler procedures perform in the true data generating process is more general than the assumptions that they are based on. We evaluate the performance of these methodologies in two ways: first we compare our estimator of volatility with theirs for the same set of data. However, our estimator should be a benchmark only if it is our model is correctly specified. In reality, it is unlikely that one cannot tell whether the unobserved price follows a particular process, so we run into a problem of not knowing the “true” target. One advantage of our fully specified parametric model is that we can generate the price and trade series assuming our model is the true data generating process, and evaluate the performance of various models on the simulated data. With the simulated data, we would know the true daily volatility of efficient price innovation. We then estimate the daily realized volatility using various exiting procedure. Therefore, in the second part, we evaluate the performance of various measures using the simulated data.

The most naïve way to estimate volatility using tick by tick data is simply adding up all the tick by tick squared returns. There are at least two sources of biases of such a measure. First, it ignores information of duration between trades. As pointed out by O'Hara (1995), when trading frequency is dependent on whether there is new information or not, variance computed from only transaction data is imparted an upward bias, no trading also bares information of the underlying price process, and therefore should be taken into accounted when estimate the "true" volatility. Second, as most of the realized volatility literatures are aware of, ignoring the microstructure noise will also render a bias of the volatility estimator. This problem is most serious with the tick by tick data since microstructure noise such as order processing cost does not shrink with time.

[Insert Table 5]

Table 5 summarizes the ratio average daily realized volatility computed from our model and the daily realized volatility computed from raw returns $\frac{E(\sigma^2_{\omega(t,i)})}{\mathrm{var}(r_{t,i})}$. We find a lot of variation in the ratio. In general, we find pretty high magnitude of bias, the degree of upward bias can be as high as 100% as in some of the stocks. The direction of bias can be both upward or downward, three out of ten cases render downward biased estimates, suggesting the iid assumption of the microstructure noise is not an accurate description of the data.

[Insert Table 6]

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. All of the previous measures are estimated using equally spaced data. Table 6 calculates the ratio between daily volatility of the efficient price change and the daily realized volatility. Four popular volatility estimates are compared in the table. 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.

To better understand how the nonparametric estimators are affected by the properties of the microstructure noise, we run a simulation experiment. We simulate 10 pseudo companies, using the 10 sets of parameters estimated from the 10 companies in our datasets. On the one hand, simulation generated from the parameters estimated from the real-world data will make our study more likely to reflect cases in real life; on the other hand, with the simulated data, we can easily compute the true realized volatility and use it as a benchmark to evaluate other volatility estimators.

[Insert Figure 4, 5 and 6]

For each company, we simulate at least 100 days. The total number of data generated ranges from 100,000 for the least frequently traded stock – FUN, to more than 4,000,000 for the most frequently traded stock – IBM. Several measures are used to compare the estimates with the true variance. First, we compute the average ratio of the estimate to the true daily volatility, this measure gauges overall biasedness of each estimate. Second, we calculate correlation between estimated and true realized variance. As the true daily realized variance is time varying, we would like to have an estimator that gives a higher value when the true volatility during a day is also higher. The first two measures each reflect part of the story. To summarize all the information into one measure, we use the root mean squared error (RMSE). The ratio, correlation and RMSE of each companies are plotted side by side in Figure 4, 5 and 6 respectively. As we can see, the roll measure performs the worst in terms of biasedness, the ratio between the estimate and true measure can as high as 200% for some stocks and as low as 50% for other stocks. The three other measures all give ratios fairly close to one. In contrast, roll's model perform the best in terms of its correlation with the true measure, although the two-scale measure also performs consistently well with a correlation above 70% with the true measure. The RMSE in some sense combines the information of both ratio and correlation. Again, the Roll measure exhibits large variation in performance, the 30 minutes realized volatility also performs not satisfactorily for some of the stocks. Both HL and ZMA measures beat 30 minutes measure. ZMA measure persistently performs well for all stocks.

[Insert Table 7]

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.

Figure 7

VI. Volume's effect on volatility

Previous researches show that there are multiple attributes of trades that can help in the inference about private information. Among these attributes, the effect of the sizes of trades has been studied most extensively. Both theoretical and empirical studies have shown that trading volume has an impact on market makers' inference about the stocks' value (Easley and O'Hara (1987), citation). Since duration between trades and size of trades are both affected by underlying information, they are likely to be correlated. It is interesting to see how the inference about duration is affected when trading volume information is included.

To incorporate volume information, we made a slight modification of the model for the innovation of efficient price. In particular, we add another multiplicative term to equation (3.6). The new specification is now $\sigma_{\omega(t,i)} = \sqrt{h_t s_{t,i} g_{t,i} d_{t,i}{}^{\delta} (a_1 \text{Large}+a_2 \text{Medium})}$. Large (Medium) is a dummy variable taking a value of one if the trade size is large (medium). Similar to duration, we first take out the time of day effect of trading volume. We then use trading volume information over the past month to determine whether the current trade is a large or small one. A trade is a big (medium) one if it is larger than the 2/3 (between 1/3 and 2/3) of the trades over the past month.

[Insert Table 8]

Table 8 reports the some of the key parameters of the new model. We find that including the volume information does help in terms of overall goodness of fit, both the log likelihood and AICC favors the new model. Consistent with the previous literature, we find that large trades have significantly higher impacts on efficient prices relative to median and small trades. In contrast, medium and small trades don't differentiate between each other in terms of price impact. Finally, the inference about δ is not even when volume information is incorporated. All 10 stocks render a coefficient significantly smaller than one and size out of ten are insignificantly different from zero.

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 dependence 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.

Easley, David, and Maureen O'Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance* 47, 577 – 606.

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

We lay out the following state space model.

Let the state equation be:

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

and the observation equation be:

$$r_{t,i} = \begin{bmatrix} \tilde{\theta}_0 & \tilde{\theta}_1 & \cdots & \tilde{\theta}_{\kappa_1-1} & 1 & \tilde{B}_1 & \cdots & \tilde{B}_{\kappa_2-1} \end{bmatrix} \begin{bmatrix} \chi_i \\ \zeta_i \end{bmatrix} \quad \text{(A.2)}$$

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).

Prove that equation (A.1)-(A.2) can be written in the form(3.11):

Let $\chi_{j,i}$ be the jth entry of the vector χ_i ,from **Error! Reference source not found.** we have

$$
\begin{gathered}
\chi_{1,i+1} = \phi_1 \chi_{1,i} + \phi_2 \chi_{2,i} + \ldots + \phi_{r_1} \chi_{r_1,i} + \sigma_{\omega(t,i+1)} \omega_{i+1} \\
\chi_{2,i+1} = \chi_{1,i} = L \chi_{1,i+1} \\
\chi_{3,i+1} = \chi_{2,i} = L^2 \chi_{2,i+1} \\
\vdots \\
\chi_{r_1,i+1} = L^{r_1-1} \chi_{1,i+1}
\end{gathered} \tag{A.3}
$$

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

$$\chi_{1,i+1} = \frac{\sigma_{\omega(t,i+1)} \omega_{i+1}}{1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_{r_1} L^{r_1}} \tag{A.4}$$

Similarly, let $\zeta_{j,i}$ be the jth entry of the vector ζ_i, then from (A.1) we have

$$\zeta_{1,i+1} = \frac{\eta_i}{1 - A_1 L - \ldots - A_{r_2} L^{r_2}} \tag{A.5}$$

Expanding Equation (A.2)

$$
\begin{aligned}
r_{t,i} &= (\tilde{\theta}_0 \chi_{1,i} + \tilde{\theta}_1 \chi_{2,i} + \ldots + \tilde{\theta}_{r-1} \chi_{r_1-1,i}) + (\zeta_{1,i} + \tilde{B}_1 \zeta_{2,i} + \ldots + \tilde{B}_{r_2-1} \zeta_{r_2-1,i}) \\
&= (\tilde{\theta}_0 + \tilde{\theta}_1 L + \ldots + \tilde{\theta}_{r-1} L^{r_1-1}) \chi_{1,i} + (1 + \tilde{B}_1 L + \ldots + \tilde{B}_{r_2-1} L^{r_2-1}) \zeta_{1,i}
\end{aligned} \tag{A.6}
$$

Finally, Equation (3.11) can be obtained by plugging (A.4) and (A.5) into (A.6).

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 $E_{i-1}(y_i y_{i+j})=C_j\dfrac{\sigma_i^2}{E(\sigma_i^2)}+D_j$,where C_j is a constant determined by the parameters $\left[\phi_1,\ldots,\phi_{p_1}, A_1,\ldots,A_{p_2},\theta_0,\ldots\theta_{q_1},E(\sigma_i^2)\right]$ and D_j is a constant determined by $(\phi_1,\ldots\phi_{p_1},B_1,\ldots B_{q_2},\sigma_\eta^2)$. 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 $\left[\phi_1,...\phi_{p_1},B_1,...,B_{q_2},\sigma^2_\eta\right]$. Therefore, 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 $(B_1,...,B_{q_2},\sigma^2_\eta)$ 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

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

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

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: $h_t, s_{t,i}, g_{t,i}, d^{\delta}_{t,i}$. 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	0.0172	0.0227	0.4486
ρ	5.59E-07	1.65E-07	0.0007
α	0.0672	0.0138	0.0000
β	0.9080	0.0203	0.0000
θ_0	0.1537	0.0416	0.0002
A_1	-0.9356	0.1359	0.0000
B_1	0.7898	0.1034	0.0000
Ω	0.0315	3.0351	0.9917
L	115826		
AIC	231689		

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

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

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

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

Ticker	# of obs	δ	ρ	β	α	Ω
FUN	11280	-0.0226	4.9e-7	0.8068	0.0795	3.8944
		0.0155	*7.5e-8*	*0.0215*	*0.0082*	*0.1279*
RGR	13456	0.0039	1.2e-6	0.8366	0.0753	43.0901
		0.0203	*2.0e-7*	*0.0237*	*0.0486*	*6.1572*
CDI	14478	0.0220	1.1e-6	0.8833	0.0684	1.5000
		0.0565	*5.6e-6*	*0.2947*	*0.1646*	*0.3857*
WSO	15580	**-0.0529**	3.1e-7	0.9152	0.0624	0.7158
		0.0148	*9.4e-8*	*0.0198*	*0.0059*	*0.0574*
OMM	16301	-0.0049	5.4e-7	0.9034	0.0752	39.4885
		0.0450	*2.1e-7*	*0.0222*	*0.0229*	*6.3608*
ASL	23832	0.0172	5.6e-7	0.9080	0.0672	0.0315
		0.0227	*1.7e-7*	*0.0203*	*0.0138*	*3.0351*
LUK	30038	-0.0096	7.4e-8	0.9076	0.0573	0.0378
		0.1744	*2.1e-7*	*0.0331*	*0.0107*	*2.1870*
CTX	196920	**0.1427**	6.6e-8	0.9111	0.0349	0.5276
		0.0471	*3.9e-8*	*0.0394*	*0.0098*	*0.0788*
BA	279900	**0.1040**	9.0e-9	0.9690	0.0229	1.5894
		0.0276	*7.0e-9*	*0.0110*	*0.0058*	*0.2331*
IBM	419994	**0.2480**	6.6e-8	0.7030	0.1009	0.2649
		0.0940	*2.6e-8*	*0.0399*	*0.0128*	*0.8586*

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	$AR_\omega(1)+ARMA_\eta(1,1)$	8808.375	17206.85	0.51	0.95
RGR	$MA_\omega(1)+ARMA_\eta(1,1)$	37807.98	78824.85	0.48	1.08
CDI	$AR_\omega(1)+WN_\eta$	54325.92	22083.51	2.46	-0.59
WSO	$MA_\eta(1)$	33437.34	37181.00	0.90	0.11
OMM	$ARMA_\omega(1,1)+WN_\eta$	79652.77	85982.80	0.93	0.08
ASL	$WN_\omega+ARMA_\eta(1,1)$	105050	125782.46	0.84	0.20
LUK	$AR_\omega(1)+AR_\eta(1)$	15866.95	14111.86	1.12	-0.11
CTX	$AR_\omega(1)+MA_\eta(1)$	30037.05	24244.75	1.24	-0.19
BA	$AR_\eta(1)$	49479.75	51161.27	0.97	0.03
IBM	$AR_\omega(1,1)+WN_\eta$	20044.42	24847.91	0.81	0.24

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 $\frac{Var(r_{t,i})}{E(\sigma^2_{\omega(t,i)})}-1$.

Ticker	$\frac{E(\Sigma_{i=1}^{n}\sigma_{\omega(t,i)}^{2})}{E(\Sigma_{i=1}^{n}r_{t,i}^{2})}$	**Roll**	**30 minutes**	**HL**	**ZMA**
FUN	0.51	0.8201	0.6861	0.7019	0.3863
RGR	0.48	0.7735	0.5856	0.5861	0.3846
CDI	2.46	1.1194	1.5233	1.4632	1.0377
WSO	0.90	0.8289	0.8583	0.8698	0.5789
OMM	0.93	0.8384	0.8243	0.8045	0.5493
ASL	0.84	0.8047	0.7434	0.7777	0.4411
LUK	1.12	0.9698	0.9612	1.0461	0.7682
CTX	1.24	0.9527	1.3914	1.2263	0.9749
BA	0.97	0.8188	0.4556	0.3623	0.6938
IBM	0.81	0.776	0.8506	0.8192	0.5798
Median	0.91	0.8245	0.8375	0.8118	0.5794
Correlation with our model		0.920411	0.858537	0.978308	0.758022

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.

Ticker	**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.

Ticker	Model	δ	Std. dev.	Large	Std. dev.	Med	Std.dev.	-LogL	AICc
FUN	1	-0.0226	*0.0155*					42941	85919
	2	-0.0204	*0.0166*	**0.1126**	*0.0540*	0.0718	*0.0671*	42936	85908
RGR	1	0.0039	*0.0203*					59336	118708
	2	0.0132	*0.0085*	**0.2825**	*0.0434*	**0.1473**	*0.0390*	59306	118648
CDI	1	0.0220	*0.0565*					53053	106142
	2	0.0252	*0.0139*	**0.2431**	*0.0959*	0.0744	*0.2115*	53024	106083
WSO	1	**-0.0529**	*0.0148*					60287	120600
	2	**-0.0405**	*0.0183*	**0.3486**	*0.0203*	0.0067	*0.0359*	60175	120383
OMM	1	-0.0049	*0.0450*					70642	141311
	2	-0.0070	*0.0351*	0.1249	*0.0979*	0.0040	*0.0437*	70628	141289
ASL	1	0.0172	*0.0227*					115826	231689
	2	0.0159	*0.0102*	**0.3521**	*0.0258*	0.2372	*0.0041*	115716	231459
LUK	1	-0.0096	*0.1744*					89928	179892
	2	-0.0026	*0.0167*	**0.5200**	*0.0262*	0.2128	*0.0500*	89675	179387
CTX	1	0.1427	*0.5000*					56388	112812
	2	**0.1124**	*0.0312*	**0.2648**	*0.1153*	**-0.1351**	*0.0657*	56288	112611
BA	1	**0.1040**	*0.0276*					60895	121816
	2	**0.0987**	*0.0269*	0.0902	*0.1400*	-0.0602	*0.1424*	60876	121778
IBM	1	**0.2480**	*0.0940*					48432	96900
	2	**0.1949**	*0.0744*	**0.3719**	*0.0698*	0.0033	*0.0619*	48324	96679

Table 8 : Volume's effect on volatility

Figure 1 Daily Seasonality for Duration

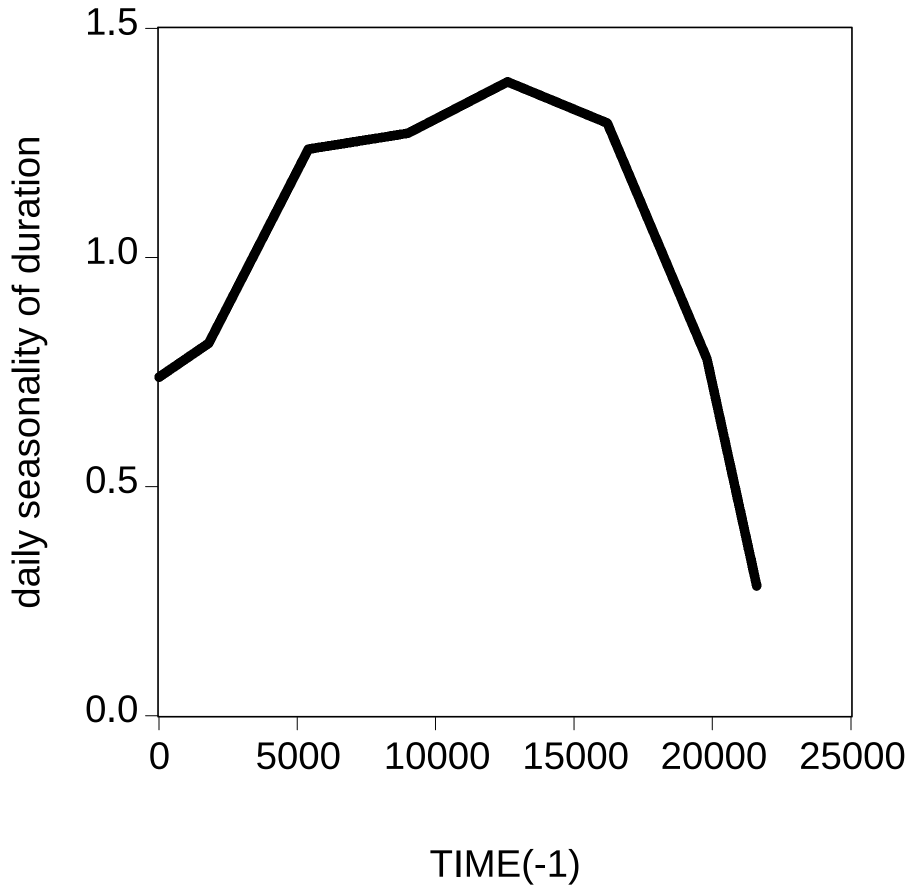

Figure 2 Daily Seasonality for Tick Volatility

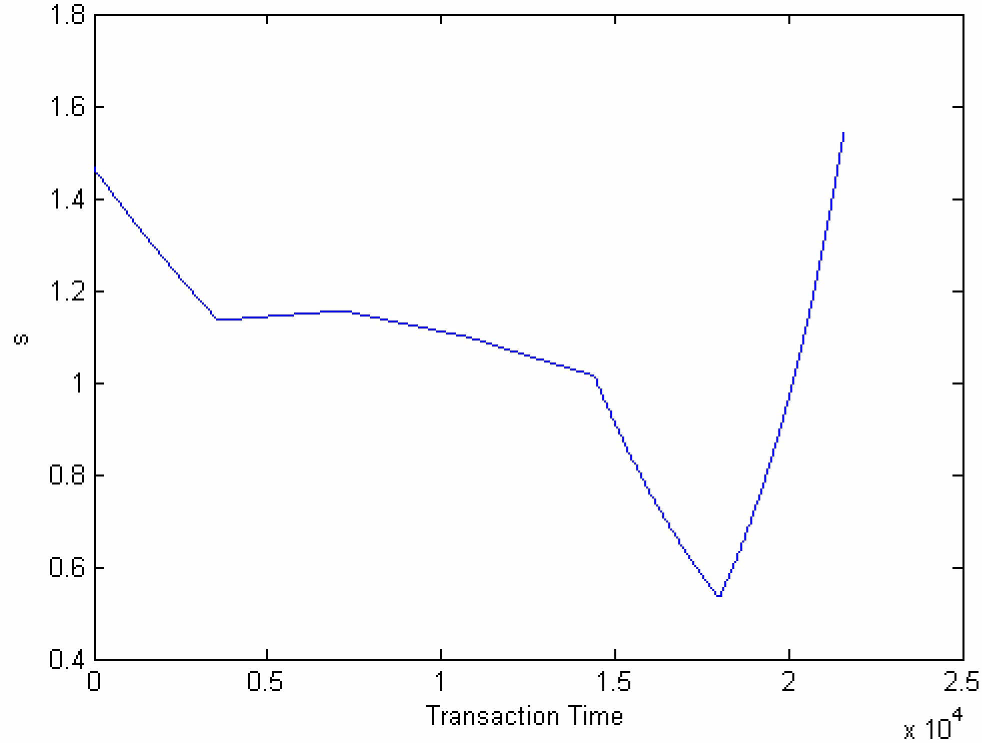

Figure 3: Generalized Gamma Probability Plot

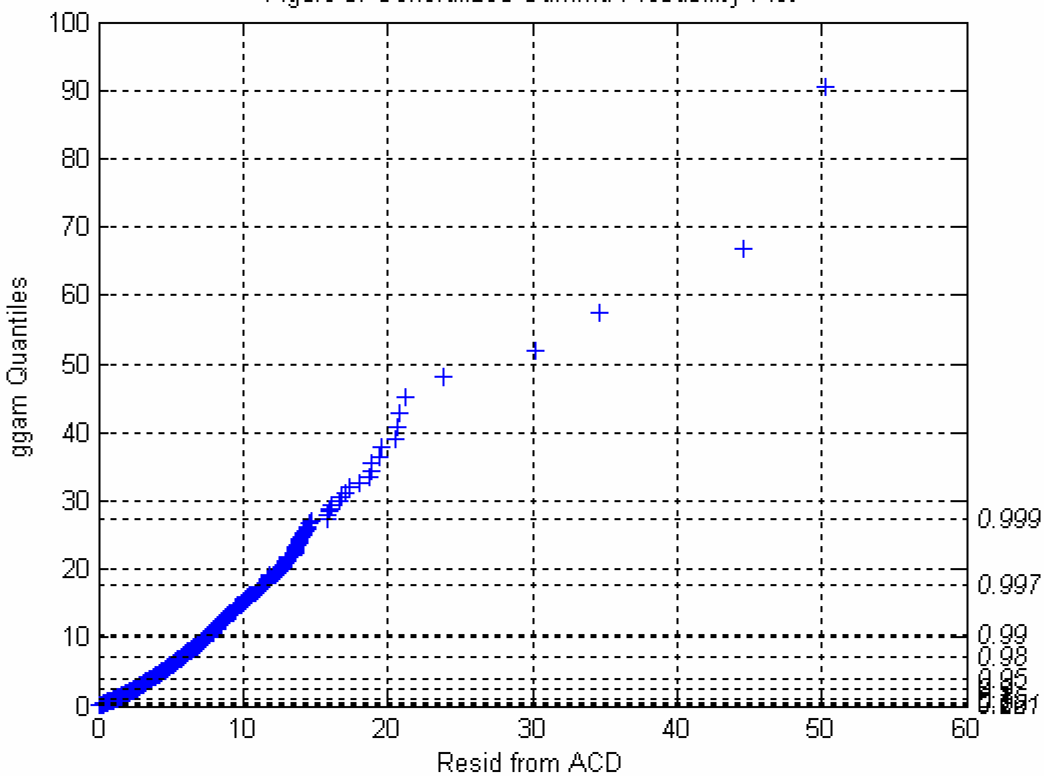

Figure 4 : Ratio of estimated volatility versus true

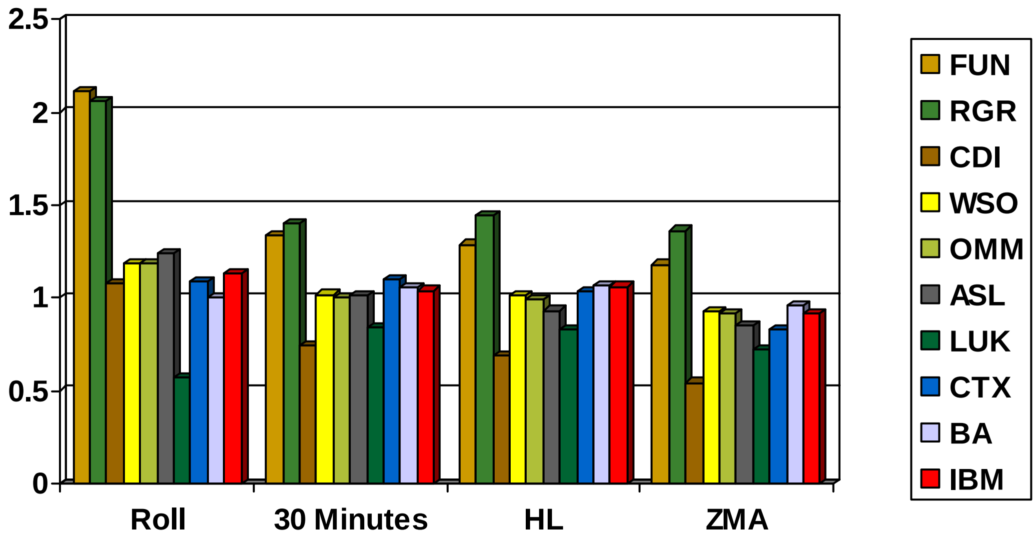

Figure 5 : Correlation of estimated volatility to true

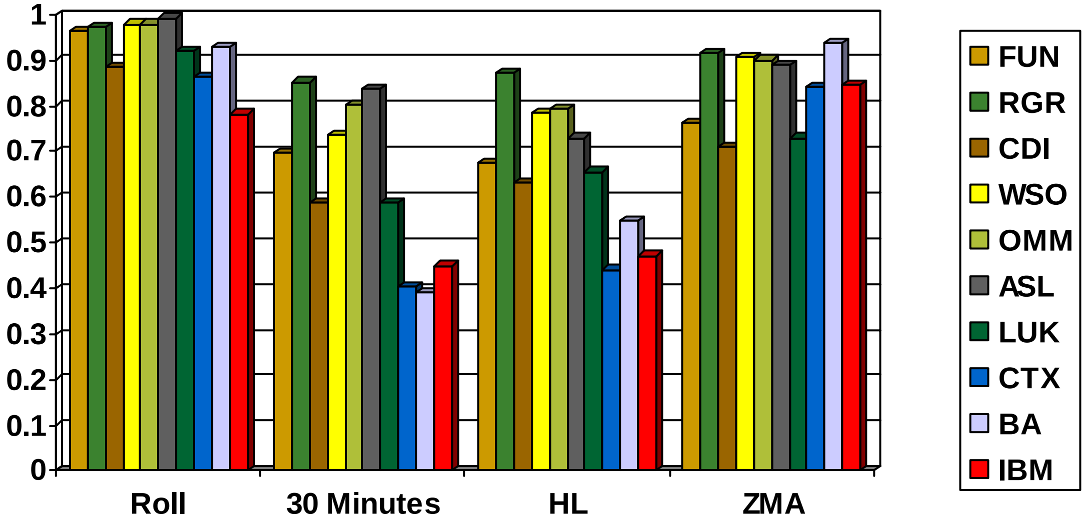

Figure 6 : RMSE standardised by daily efficient price volatility

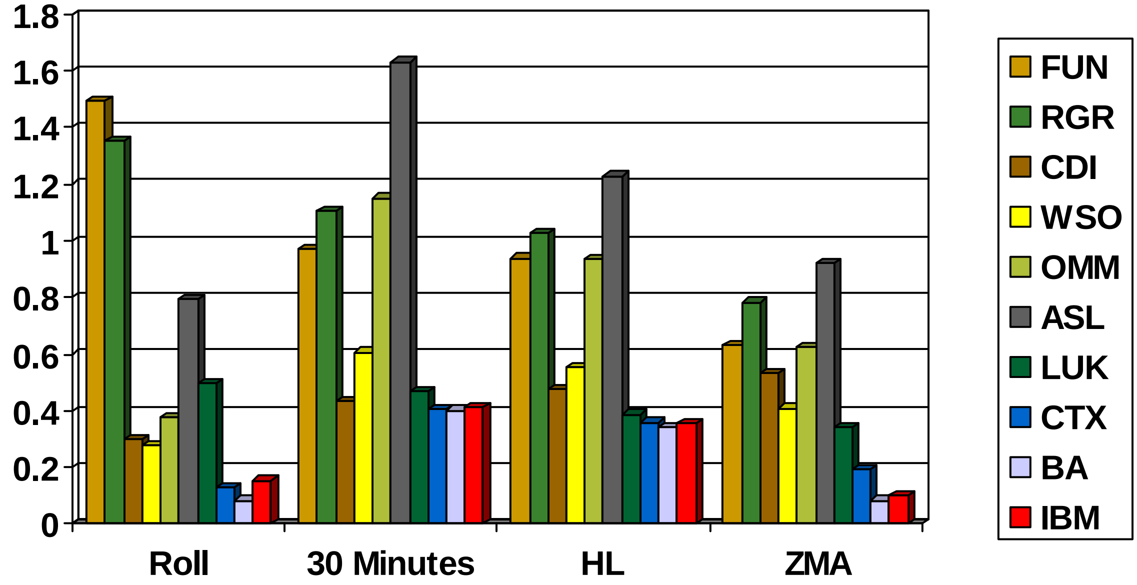

Figure 7 : Impulse response function