Joint modeling of high-frequency price and duration data

Denis Pelletier
North Carolina State University*

and

Haiqing Zheng
North Carolina State University†

Current version: May 31, 2013

Abstract

High frequency financial data is irregularly spaced in time. The information content that determines the time between subsequent trades introduces is potentially related to volatility. We introduce a new continuous time model to jointly model stock return and duration between trades. This model include a bivariate Ornstein-Uhlenbeck process for two latent processes: log-volatility of stock returns and log-intensity of the elapsed time between trades. We apply this model to tick-by-tick stock price data. We find that volatility and intensity have strong persistence and are contemporaneously positively correlated. A Monte Carlo study points out that more accurate measurement of volatility can be obtain by conditioning on observed duration between trades in addition to conditioning on the returns.

* Mailing address: Department of Economics, Campus Box 8110, North Carolina State University, Raleigh, 27695-8110, USA. e-mail: denis_pelletier@ncsu.edu. Web site: http://www4.ncsu.edu/~dpellet.
† Mailing address: Department of Economics, Campus Box 8110, North Carolina State University, Raleigh, 27695-8110, USA. e-mail: hzheng2@ncsu.edu. Web site: http://www4.ncsu.edu/~hzheng2.
1 Introduction

Being critical to risk management, portfolio allocation and asset pricing, volatility has been an active area of research in financial econometrics. It is of great interest in financial econometrics to accurately measure and forecast it. Since at least the ARCH model of Engle (1982), volatility has generally been considered as time-varying and persistent. There is a large body of literature covering extension of the ARCH methodology (see the book chapter: Bollerslev, Engle, and Nelson (1994); comprehensive review papers: Bera and Higgins (1995); Bollerslev, Chou, and Kroner (1992) and Poon and Granger (2003)). More recently, however, the focus on volatility modeling has shifted away from ARCH-type models due to their drawbacks of being expressed in discrete time, not closed under temporal aggregation. A greater focus in now on nonparametric volatility measurements such as realized volatility estimated with high-frequency data (tick-by-tick data).

In an ideal frictionless market environment, ex-post integrated volatility can be consistently measured by the realized volatility estimator, which is simply a sum of all the intraday squared returns over short fixed time intervals. This estimator is model-free and computationally convenient. Theoretically when the sampling frequency gets finer and finer, to the limit it converges to the quadratic variation over the same period (see Andersen and Bollerslev (1998); Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2002)). Therefore, in this environment, realized volatility estimates should be computed with returns over smallest time intervals possible.

However in reality, this is not the case. The equilibrium or true prices are contaminated by market microstructure noise which could possibly include bid-ask spread, price discretization, data recording error, etc. (see Stoll (2000)). There is a consensus in the literature that realized volatility derived from observed prices contains the component of market microstructure noise and is no longer a consistent estimator. In more recent papers, different authors have been interested in deriving consistent and efficient estimator of quadratic variation or integrated variance of financial returns under the presence of market frictions (see review papers: Bandi and Russell (2007) and Andersen and Benzoni (2008)). Among these studies, Andersen, Bollerslev, Diebold, and Labys (2001) propose a sparse sampling scheme. They suggest a sampling frequency lower than the highest frequency available to balance efficient sampling and bias-inducing noise. The basic idea is that if the sampling interval separating
two prices becomes larger, the volatility of underlying true price process increases while market microstructure noise component remains constant. In terms of signal-to-noise ratio, lower frequency would be a better choice. Zhang, Mykland, and Ait-Sahalia (2005) develop a benchmark “Two scales volatility estimator (TSRV)”. This estimator is computed from averaged realized volatilities of sub-samples and corrected for the remaining errors. Another approach to deal with market microstructure noise is to design kernels that are robust to it. Zhou (1996) was the first to use a kernel method to handle microstructure effects and his approach is based on autocorrelation of the return in the i.i.d. noise case. More recently, Hansen and Lunde (2006) extended this approach to non i.i.d. noise and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) develop a kernel based method robust to endogenous and/or dependent noise.

One feature of high frequency data that differs from daily, weekly or monthly data is that the prices are recorded at irregularly spaced points in time. Since trades are induced by market participants reacting in part to new information hitting financial markets, it is very likely that the time between trades is not a random exogenous process. It ought to interact with the price formation process. So if we use the information that is in the durations between trades on top of the information from the returns between trades, it could potentially improve the accuracy of our estimates of volatility. There is some evidence in the literature that durations are endogenous. Li, Mykland, Renault, Zhang, and Zheng (2009) find evidence of endogeneity of sampling time in high-frequency data and establish a central limit theorem for realized volatility in an endogenous time setting.

There are papers that use parametric methods to jointly model returns and trade durations, such as Ghysels and Jasiak (1998), Gramming and Wellner (2002), Engle (2000) and Engle and Sun (2007). However, these studies are ACD-GARCH type and some of them haven’t taken market microstructure noise into account.

In this paper, we develop a model for the joint evolution of returns and time between trades. The return over consecutive trades is assumed to follow a basic stochastic volatility model. The duration between trades is assumed to be exponentially distributed conditional on a latent stochastic trade intensity. It is assumed that the two latent variables, the volatility and the trade intensity, follow a bivariate Ornstein-Uhlenbeck (OU) process. We also allow the observed returns to be contaminated by market microstructure noise. By specifying the model in continuous time, the parameter values
of the discretized version of the model will be consistent across the different durations, unlike the ACD-GARCH type models. With our model, we are able to get estimates of volatility conditional on both the observed returns and the durations between trades.

In an empirical application where we estimate this model with tick-by-tick data for AMD stock, we find that both intensity and volatility have strong persistence and they are contemporaneously positively correlated. Market microstructure noise is quite significant but its impact can be greatly mitigated by combining a fixed number of trades. Using Monte Carlo simulations we show that we can get more precise measurements of volatility if we take into account not only the returns but also the durations.

The rest of the paper is structured as follow. Section 2 develops the model for joint distribution of return and durations, and writes it as a linear state space representation. The estimation strategy is described in section 3. In Section 4, we estimate the model with tick-by-tick data and report the empirical results. Monte Carlo simulation results are presented in section 5 and section 6 contains the concluding remarks.

2 Model for price and duration

In this section we develop a model for the joint distribution of returns and durations between trades.

2.1 Specification in continuous time

As we have mentioned earlier, in high frequency data, the observations are irregularly spaced in time. Trade events can be modeled as a point process. Let us denote the arrival time of the trade events by the sequence \( \{t_i\}_{i \in \{1, 2, \ldots, n\}} \). This is the realization of a point process defined on some probability space \( (\Omega, \mathcal{F}, P) \). These trade events could consist of all the individual trades but in the general they will consist of a number of trades (e.g., the tenth trade, twentieth trade, thirtieth trade, ...). The number of trade events until time \( t \) in \( [0, \infty) \) can be viewed as the corresponding right-continuous counting function to this realization, which is defined as \( N_t := \sum_{i \geq 1} 1_{\{t_i \leq t\}} \). Denote the filtration \( \mathcal{F}_t \) as the history of observed and some unobserved processes up to time \( t \), the \( \mathcal{F}_t \)-intensity function of
this counting process is defined by
\[
\lambda(t; \mathcal{F}_t) := \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\{[N_{t+\Delta} - N_t] > 0|\mathcal{F}_t\}. \tag{1}
\]

The intensity summarizes the arriving rate of the next occurrence in the near future conditional on the past history.

For the trade arrival process, the assumption of homogeneous Poisson process with constant arrival rate is not adequate. In the case of non-homogeneous Poisson process, the rate function is a known function of time and some observed explanatory variable. If the rate function includes additional unobserved characteristics, the resulting process is said to be a doubly stochastic Poisson process, also known as Cox process (Cox (1955) and Cox and Isham (1980)). Loosely speaking, it is a Poisson process whose rate is modulated by a second stochastic process governed by some probabilistic law. A formal definition from Bremaud (1981) is as follows.

**Definition 1** Let \( N_t \) be a point process adapted to a history \( \mathcal{F}_t \), and let \( \lambda_t \) be a nonnegative measurable process [all given on the same probability space \((\Omega, \mathcal{F}, P)\)]. Suppose that \( \lambda_t \) is \( \mathcal{F}_0 \)-measurable, \( t \geq 0 \) and that \( \int_0^t \lambda_s ds < \infty \) \( P \)-a.s., \( t \geq 0 \). If for all \( 0 \leq s \leq t \) and all \( u \in \mathbb{R} \)
\[
E[e^{iu(N_t - N_s)}|\mathcal{F}_s] = \exp\left\{ (e^{iu} - 1) \int_s^t \lambda_v dv \right\},
\]

then \( N_t \) is called a \((P, \mathcal{F}_t)\)-doubly stochastic Poisson process or a \((P, \mathcal{F}_t)\)-conditional Poisson process with the (stochastic) intensity \( \lambda_t \).

Letting \( d_i \equiv t_i - t_{i-1} \) denote the time elapsed between the event \( i - 1 \) at time \( t_{i-1} \) and the event \( i \) at time \( t_i \). If every transaction is considered an event, then \( d_i \) consist in the time between consecutive transactions. We assume that conditional on the intensity \( \lambda_{t_{i-1}} \), the distribution of \( d_i \) is
\[
f(d_i|\lambda_{t_{i-1}}) = \lambda_{t_{i-1}} \exp(-\lambda_{t_{i-1}} d_i), \tag{3}
\]
i.e. an exponential distribution with mean \( 1/\lambda_{t_{i-1}} \).

Next, we consider the asset price \( S_t \) at time \( t \). Following Engle (2000) and Renault and Werker
(2010), we ignore a possible drift term in the price equation and assume that the logarithmic price evolves according to:

\[ d\ln S_t = \sigma_t dW_{0,t}, \]

where \( \sigma_t \) is the volatility and \( W_{0,t} \) is a Wiener process.

It is well known that there is deterministic seasonal pattern during the business day for volatility (e.g., see Andersen and Bollerslev (1997)) and trade intensity (e.g., see Engle and Russell (1998)). For assets that are not traded around the clock, volatility is usually higher when the markets are opening, it lowers during the middle part of the day before increasing again after that. A similar pattern exist for durations: they are relatively shorter at the opening and closing of the market, and relatively longer during the middle of the business day. Trade intensity being inversely related to durations, have a similar pattern as volatility: relatively higher trade intensity at the beginning and end of the business day than during the middle of the day.

To capture these diurnal patterns, we introduce two deterministic functions of the time of the day, \( g_\sigma(t) \) and \( g_\lambda(t) \) for the volatility and trade intensity respectively. To get a \( U \) shape for these seasonal effects, we will assume a quadratic form\(^1\). Since the volatility and trade intensity have to be positive, the functions \( g \) cannot change sign, so will have to be positive. Also, the level of the functions have to be fixed because they could not be identified separately from the unconditional mean of the volatility and trade intensity. Accordingly, we take these functions to be equal to

\[ g_\sigma(t) = \alpha_{\sigma,1}(t + \alpha_{\sigma,2})^2 + 1, \quad g_\lambda(t) = \alpha_{\lambda,1}(t + \alpha_{\lambda,2})^2 + 1. \]

and assume that the variance and trade intensity are equal to the diurnal functions \( g \) times a stationary process:

\[ \sigma_t = g_\sigma(t)\sigma^*_t, \quad \lambda_t = g_\lambda(t)\lambda^*_t. \]

To model the two latent processes, intensity \( \lambda^*_t \) and volatility \( \sigma^*_t \), we consider a bivariate Ornstein-Uhlenbeck process. The univariate Ornstein-Uhlenbeck (OU) process is a mean-reverting process

\(^1\)Any deterministic parametric function of the time of the day would work.
with wide applications, e.g. modeling biological process, psychological behavior, animal movement, interest rates [Vasicek (1977)], commodity convenience yield [Gibson and Schwartz (1990)], volatility of asset prices [Wiggins (1987)] and in the practice of pair trading and pricing and hedging of spread options. The bivariate (multivariate) OU process is a natural extension of the univariate OU process.

For the logarithm of the intensity ($\ln \lambda^*_t$) and the logarithm of the variance ($\ln \sigma^*_t$), we assume the following dynamic:

$$dX_t = -A(X_t - \mu) \, dt + SdW_{-0,t},$$

where

$$X_t = \begin{bmatrix} \ln \lambda^*_t \\ \ln \sigma^*_t \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad S = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}, \quad W_{-0,t} = \begin{bmatrix} W_{1,t} \\ W_{2,t} \end{bmatrix}.$$  

The matrix $A$ is the transition matrix, which defines the deterministic portion of the evolution of the process. The vector $\mu$ is the unconditional expectation of $X_t$. The matrix $S$ is the scatter generator, a diagonal matrix that induces the dispersion. The two Wiener processes could be contemporaneously correlated. We can also allow $W_{0,t}$ and $W_{-0,t}$ to be correlated (to allow for example the usual leverage effect):

$$< W_{0,t}, W_{1,t} > = \rho_1, \quad < W_{0,t}, W_{2,t} > = \rho_2, \quad < W_{1,t}, W_{2,t} > = \rho_3.$$  

Therefore we form a correlation matrix,

$$\Gamma = \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{bmatrix}.$$  

7
2.2 Discretized version of the model

The system described above is in continuous time. However, observations are only available at discrete time intervals. For the purpose of estimation, we discretize the continuous time model.

We first we discuss the discretization and general properties of the bivariate OU process (see Schach (1971), Dunn and Gipson (1977), Blackwell (2003) and Meucci (2010) for detailed explanations). For a duration $d_{t+1}$, integrating the system (5) we obtain the following exact autoregressive dynamic for $X_t$:

$$X_{t+1} = [I_2 - \expm(-Ad_{t+1})] \mu + \expm(-Ad_{t+1})X_t + \omega_{t+1}$$  \hspace{1cm} (7)

The term $\expm(-Ad_t)$ is the matrix exponential of $-Ad_t$. It is defined as

$$\expm(A) \equiv \sum_{k=0}^{\infty} \frac{A^k}{k!}. \hspace{1cm} (8)$$

The innovation $\omega_{t+1}$ is equal to

$$\omega_{t+1} = \int_{t_i}^{t_{i+1}} \expm(A[u-t_{i+1}])SdW_{0,u} \sim N(0, \Sigma^\omega_{t_{i+1}}). \hspace{1cm} (9)$$

The variance of this error term, $\Sigma^\omega_{t_{i+1}}$, can be expressed using the operator vec (stacks the columns of a matrix) and the Kronecker sum as

$$\text{vec}(\Sigma^\omega_{t_{i+1}}) = (A \oplus A)^{-1}(I_4 - \expm(-(A \oplus A)d_{t+1})) \text{vec}(\Sigma^\omega) \hspace{1cm} (10)$$

where $\Sigma^\omega \equiv S \left[ \begin{array}{cc} 1 & \rho_1 \\ \rho_1 & 1 \end{array} \right] S$ and $A \oplus B = A \otimes I_b + I_a \otimes B$.

Conditional on $X_t$, $X_{t+1}$ is normally distributed. The transition matrix $A$ measures the centralizing tendency. It describes the force that keeps the process from getting away from the center of activity $\mu$. It determines the dynamics of the system. If $\expm(-Ad_t)$ is close to $0$, the conditional mean is close to $\mu$. This means the process moves towards $\mu$ at a high velocity.

For general multivariate OU processes, the stationarity condition requires that $A$ be positive definite in order to satisfy the requirement of Kolmogorov forward equation [Cox and Miller (1965)].
If the square matrix $A$ is symmetric, an equivalent requirement is that all the eigenvalue of $A$ be positive. For a general real square matrix $A$, since it can be decomposed as the sum of a symmetric $(A + A^T)/2$ and skew-symmetric matrix $(A - A^T)/2$, $A$ is positive definite if and only if the symmetric part $(A + A^T)/2$ is positive definite. If the matrix $A$ is positive definite, then $\expm(-Ad) \to 0$ as $d$ goes to infinity and the unconditional distribution is:

$$X_t \sim N(\mu, \Sigma_\infty),$$

where

$$\text{vec}(\Sigma_\infty) = (A \oplus A)^{-1} \text{vec}(\Sigma^w).$$

Next, we consider the specification of the matrix $A$. The simplest case is where $A$ is an isotropic matrix [see Blackwell (1997), Blackwell (2003) and Oravecz, Tuerlinckx, and Vandekerckhove (2008)]. An isotropic matrix is a matrix satisfying the following equality, $A = \theta \times I$, i.e. a scalar $\theta$ times a matrix of ones. The matrix exponential term $\expm(-Ad_i)$ is then equal to $\exp(-\theta d_i)I$, which is easier to handle and will largely reduce the complexity of the system. This is the main reason why it is considered in many studies.

In the case where $A$ is a diagonal matrix, all the diagonal elements have to be positive for $A$ to be a positive definite matrix. If we define $A$ as the matrix exponential in this case becomes especially if we define $A = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}$ then, $\expm(A) = \begin{bmatrix} \exp(a_{11}) & 0 \\ 0 & \exp(a_{22}) \end{bmatrix}$ and the conditional mean of the process could be expressed as $[(1 - \exp(-a_{kk}d_{i+1}+1))\mu_k + \exp(-a_{kk}d_{i+1})X_{k,t_i}]_{k=1,2}$. A diagonal matrix $A$ implies that there is no cross-influence from $X_{1,t_i}$ to $X_{2,t_{i+1}}$ and from $X_{2,t_i}$ to $X_{1,t_{i+1}}$. Therefore, the model can be decomposed as two univariate OU models with potential interdependence in the innovation terms.

The most general case is where $A$ is not restricted except for the positive definiteness restriction so as to ensure the stationarity of the system. To satisfy the positive definite condition, all the eigenvalues of $(A + A^T)/2$ have to be positive. If the transition matrix $A$ has complex eigenvalues with all the real parts being positive, the process converges oscillatingly toward the mean in the long run. The real part of the complex eigenvalues determines the exponential convergence rate and the imaginary part dictates the rotation frequency. A transition matrix with real positive eigenvalues implies that the
converging path is non-oscillatory toward equilibrium. Figure 1 shows a bivariate OU process for the first case (complex eigenvalues) and figure 2 for the second case (real eigenvalues). Both processes are stationary and demonstrate the mean-reverting property. Figure 1 displays the rotation pattern and exponentially fast convergence that is not observed in Figure 2, in which we observe a less dense depiction of the connecting lines.

Regarding the discretizing the equation (4) for the evolution of the price, for a duration $d_{i+1}$ we have the following equation for the return $y^*_{t_{i+1}}$ between the trade events $i$ and $i + 1$:

$$y^*_{t_{i+1}} = \sigma_i u_{t_{i+1}}$$  \hspace{1cm} (13)

where $y^*_{t_{i+1}} = \ln S_{t_{i+1}} - \ln S_{t_i}$. We use the $*$ notation to denote that this is the “true” return, which may or may not correspond to the observed return. The innovation $u_{t_{i+1}} = \int_{t_i}^{t_{i+1}} dW_{0,t}$ has a normal distribution,

$$u_{t_{i+1}} \sim N(0, d_{i+1}).$$  \hspace{1cm} (14)

The innovation terms in equation (7) for the latent variables and the equation (13) for the return are potentially correlated:

$$\Sigma_{t_{i+1}}^{u,\omega} = Cov(u_{t_{i+1}}, \omega_{t_{i+1}}) = Cov \left( \int_{t_i}^{t_{i+1}} \exp(A(u - t_{i+1})) S dW_{-0,u}, \int_{t_i}^{t_{i+1}} dW_{0,u} \right)$$

$$= \int_{t_i}^{t_{i+1}} \exp(A[u - t_{i+1}]) S \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} du$$

$$= A^{-1} \left[ \exp(A[u - t_{i+1}]) \right]_{t_i}^{t_{i+1}} S \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

$$= A^{-1} \left[ I_2 - \exp(-Ad_{i+1}) \right] S \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}.$$  \hspace{1cm} (15)
Putting all of these variances and covariance together, we denote

\[
\Sigma t_i = \text{Var} \begin{bmatrix} t_{ti} \\ \omega t_i \end{bmatrix} = \begin{bmatrix} d_i & \Sigma^u \omega^t \\ \Sigma^u \omega & \Sigma^t \omega \end{bmatrix}.
\] (16)

Notice that in (6), simply constraining \( \rho_i \in [-1, 1], i = 1, 2, 3 \), cannot guarantee positive semidefiniteness (PSD) of the correlation matrix. In order to impose PSD, we parametrize the correlation matrix through its Cholesky decomposition. Define \( P \) as a lower triangular matrix,

\[
P = \begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix}, \quad \Gamma = PP^t = \begin{bmatrix} p_{11}^2 & p_{11}p_{21} & p_{11}p_{31} \\ p_{11}p_{21} & p_{21}^2 + p_{22}^2 & p_{21}p_{31} + p_{22}p_{32} \\ p_{11}p_{31} & p_{21}p_{31} + p_{22}p_{32} & p_{31}^2 + p_{32}^2 + p_{33}^2 \end{bmatrix}.
\]

Imposing the additional constraint that the elements on the diagonal of \( P \) are positive, the restrictions \( \Gamma_{ii} = 1 \) for \( i = 1, 2, 3 \), becomes \( p_{ii} = \sqrt{1 - \sum_{j=1}^{i-1} p_{ij}^2} \) for \( i = 1, 2, 3 \), where the sum is zero for \( i = 1 \).

### 2.3 Market micro-structure noise

It is well known that there is market microstructure noise in the high frequency financial data, the true price being contaminated by this noise. The observed return is no longer the true return and the true return is now a latent process. If the noise is assumed to be multiplicative to the true return, then the observed return is:

\[
y_{ti} = y^*_t e_{ti} \quad \text{(17)}
\]

where \( e_{ti} \) is the market microstructure noise for \( i \)th trade event. From now on, we will refer to \( y_{ti} \) as the observed return and \( y^*_t \) as the true return. We assume a multiplicative rather than additive form for the noise purely for convenience as it will become evident when we cast this model into a state-space form in the next section.

In market microstructure theory, Roll (1984)’s bid-ask model that recognizes the order processing cost and Amihud and Mendelson (1980), Demsetz (1968), Ho and Stoll (1981), Ho and Stoll (1983), and Stoll (1978)’s inventory control model which emphasize the inventory holding costs of market
makers suggest the dependence in the noise. We assume an MA(1)-like structure for the microstructure noise:

\[ e_{t_i} = \xi \sqrt{\exp(\eta_{t_i} + \theta \eta_{t_{i-1}})}, \]  

(18)

where \( \xi = E[\sqrt{\exp(\eta_{t_i} + \theta \eta_{t_{i-1}})\]}^{-1} \) and \( \eta_{t_i} \sim N(0, \sigma^2_\eta) \). By definition, the mean value of \( e_{t_i} \) is equal to one and \( e_{t_i} > 0 \). Using the properties of the log-normal distribution, we can express \( \xi \) as

\[ \xi = \exp\left(-\frac{(1 + \theta^2)\sigma^2_\eta}{8}\right) \]  

(19)

### 2.4 State space representation

Our strategy for estimating this model is based on the approach of Harvey, Ruiz, and Shephard (1994). We want to write the model as a linear state space model and estimate the parameters by QMLE. To linearize the return equation, we take the logarithm of the squared observed return and using equations (13) and (17) we get

\[ \ln(y_{t_i+1}^2) = \ln(\sigma^2_{t_i}) + \ln(u_{t_i+1}^2) + \ln(\xi^2) + \eta_{t_i+1} + \theta \eta_{t_i}. \]  

(20)

So as to have a Gaussian linear state space, we approximate the random variable \( \ln(u_{t_i+1}^2) \) by Gaussian distribution with the same mean and variance, i.e.

\[ \ln(y_{t_i+1}^2) = \ln(\sigma^2_{t_i}) - 1.2704 + \ln(d_{t_i+1}) + \varepsilon_{t_i+1} + \ln(\xi^2) + \eta_{t_i+1} + \theta \eta_{t_i}, \]  

(21)

with \( \varepsilon_{t_i+1} \sim N(0, \sigma^2_\varepsilon) \) with \( \sigma^2_\varepsilon = 4.9348 \).

For the duration \( d_{t_i+1} \) we take a similar approach. Taking the logarithm of the duration and approximating the logarithm of an exponential distribution with a mean equal to one by a Gaussian distribution with the same mean and variance, we get

\[ \ln(d_{t_i+1}) = -\ln(\lambda_{t_i}) - 0.5772 + \psi_{t_i+1}, \]  

(22)
where \( \psi_{t+1} \sim N(0, 1.6449) \).

The equations for the log squared return (21) and log duration (22) can be combined with the equation for the evolution of the log-variance and log-intensity (7) to form a Gaussian linear state space model. The observation equation is

\[
\begin{bmatrix}
\ln(d_i) \\
\ln(g_{t_i}^2) - \ln(d_i)
\end{bmatrix} = \begin{bmatrix}
-0.5772 - \ln(g_d(t_i)) \\
-1.2704 + \ln(\xi^2) + 2 \ln(g_\sigma(t_i))
\end{bmatrix} + \begin{bmatrix}
-1 & 0 & 0 \\
0 & 1 & 1 & \theta
\end{bmatrix} \begin{bmatrix}
\ln(\lambda_{t-1}^*) \\
\ln(\sigma_{t-1}^2) \\
\eta_i \\
\eta_{i-1}
\end{bmatrix} + \begin{bmatrix}
\psi_i \\
\varepsilon_{t_i}
\end{bmatrix}
\]

and the state equation is

\[
\begin{bmatrix}
\ln(\lambda_{t}^*) \\
\ln(\sigma_{t}^2) \\
\eta_{t} \\
\eta_{t-1}
\end{bmatrix} = \begin{bmatrix}
I_2 - \expm(-Ad_t) \\
0_{2 \times 1}
\end{bmatrix} \begin{bmatrix}
\expm(-Ad_t) \\
0_{2 \times 2}
\end{bmatrix} \begin{bmatrix}
\ln(\lambda_{t-1}^*) \\
\ln(\sigma_{t-1}^2) \\
\eta_{t-1} \\
\eta_{t-2}
\end{bmatrix} + \begin{bmatrix}
\omega_t \\
0
\end{bmatrix}
\]

(23)

2.5 Using the sign of the return

As discussed above, the innovation in the price process \( (u_{t_i}) \) is allowed to be correlated with the innovations of the log-intensity and log-variance \( (\omega_{t_i}) \), see equation (15). Unfortunately, these correlations are lost when we square \( u_{t_i} \), i.e. \( \text{cov}(u_{t_i}^2, \omega_{t_i}) = 0 \) for all values of \( \rho_1 \) and \( \rho_2 \). A solution to this problem is discussed in Harvey and Shephard (1996) for the case of a univariate stochastic volatility model with leverage estimated with daily data: we can use the sign of the return to recover these correlations.

We can first point out from the assumed form of the market microstructure noise that the sign of the observed return corresponds to the sign of the latent true return. Conditioning on the sign of the
return \(s_{t_i} = 1\) if \(y_{t_i} > 0\), \(s_{t_i} = -1\) if \(y_{t_i} < 0\), the linear state-space model is now

\[
\begin{bmatrix}
\ln(d_i) \\
\ln(y_{t_i}^2) - \ln(d_i)
\end{bmatrix} = \begin{bmatrix}
-0.5772 - \ln(g_d(t_i)) \\
-1.2704 + \ln(\xi^2) + 2 \ln(g_\sigma(t_i))
\end{bmatrix} + \begin{bmatrix}
-1 & 0 & 0 \\
0 & 1 & 1
\end{bmatrix} \begin{bmatrix}
\ln(\lambda_{t_{i-1}}^*) \\
\ln(\sigma_{t_{i-1}}^2) \\
\eta_i
\end{bmatrix} + \begin{bmatrix}
\psi_{t_i} \\
\varepsilon_{t_i}
\end{bmatrix}
\]

for the observation equation and

\[
\begin{bmatrix}
\ln(\lambda_{t_i}^*) \\
\ln(\sigma_{t_i}^2) \\
\eta_t \\
\eta_{t-1}
\end{bmatrix} = \begin{bmatrix}
[I_2 - \exp(-Ad_i)] \mu + \mu_{t_i}^* \\
\expm(-Ad_i) 0_{2 \times 2} \\
0_{2 \times 1} \\
0_{2 \times 1}
\end{bmatrix} + \begin{bmatrix}
\ln(\lambda_{t_{i-1}}^*) \\
\ln(\sigma_{t_{i-1}}^2) \\
\eta_{t-1} \\
\eta_{t-2}
\end{bmatrix} + \begin{bmatrix}
\omega_{t_i}^* \\
\eta_{t_i}
\end{bmatrix}
\]

(25)

for the state equation, with

\[
\begin{bmatrix}
\varepsilon_{t_i} \\
\omega_{t_i}^*
\end{bmatrix} \sim N_2 \left( \begin{bmatrix} 0 \\
0 \end{bmatrix}, \begin{bmatrix}
\sigma_{\varepsilon}^2 & \Sigma_{t_i}^{\varepsilon, \omega^*} \\
\Sigma_{t_i}^{\varepsilon, \omega^*} & \Sigma_{t_i}^{\omega^*}
\end{bmatrix} \right)
\]

(27)

where

\[
\mu_{t_i}^* = \frac{0.7979 \Sigma_{t_i}^{u, \omega}}{\sqrt{d_i}} s_{t_i},
\]

\[
\Sigma_{t_i}^{\varepsilon, \omega^*} = 1.1061 \Sigma_{t_i}^{u, \omega} (d_i)^{-1/2} s_{t_i},
\]

\[
\Sigma_{t_i}^{\omega^*} = \Sigma_{t_i}^{\omega} - \mu_{t_i}^* \mu_{t_i}^{*\prime}.
\]

The computation is means and variances conditional on the sign of the return is relegated to an Appendix.
3 Estimation

Based on our linear state space model, we can use the Kalman filter algorithm to evaluate the likelihood and get consistent estimates with QMLE (e.g., see De Jong (1991).). The Kalman filter will also provide filtered and smoothed estimates of the latent variance and intensity, although it only provides the best linear predictor. There exist more efficient methods to estimate this model that are based on the exact likelihood such as Kim, Shephard, and Chib (1998) (Bayesian estimation) or Sandmann and Koopman (1998) (classical estimation). We leave these methods for future works as they are much more computationally demanding and we will be using this model with a large amount of observations. From the basic properties of the QML estimator (see Greene (2003)), we know that it has the following asymptotic distribution:

\[
\sqrt{n}(\hat{\theta} - \theta_0) \sim N(0, I^{-1}J I^{-1}),
\]

where \(\theta_0\) denotes the true value of the parameter vector, \(\hat{\theta}\) is the maximum likelihood estimator and

\[
J = E_0 \left[ \left( \frac{\partial \ln L(\theta_0|y)}{\partial \theta_0} \right) \left( \frac{\partial \ln L(\theta_0|y)}{\partial \theta'_0} \right) \right],
\]

\[
I = E_0 \left[ \left( \frac{\partial^2 \ln L(\theta_0|y)}{\partial \theta_0 \partial \theta'_0} \right) \right].
\]

The matrices \(J\) and \(I\) can be consistently estimated by their sample equivalent,

\[
\hat{J} = n^{-1} \sum_{i=1}^{n} \left( \frac{\partial \log L_i(\theta)}{\partial \theta} \right) \left( \frac{\partial \log L_i(\theta)}{\partial \theta'} \right),
\]

\[
\hat{I} = n^{-1} \sum_{i=1}^{n} \frac{\partial^2 \log L_i(\theta)}{\partial \theta \partial \theta'}.\]
4 Application to tick-by-tick data

4.1 Summary statistics

We estimate the model with tick-by-tick transaction data from the New York Stock Exchange (NYSE) Trade and Quote (TAQ) database for AMD stocks. The time period selected cover the period July 7, 2005 to July 20, 2005. There were ten business days during this period. The tick size during this period is one cent. TAQ reports all the trades which occurred during the consolidated tape hours of operation (4:00am to 6:30pm EST). However, we restrict our analysis only to the regular trading hours, that is from 9:30am to 4:00pm EST. Furthermore, we only consider transaction records with a correction indicator equal to zero or one. We also exclude the transactions with a sale condition equal to Z, which means that the transaction is reported to the tape at the time later than it occurred. We occasionally see multiple trades recorded with the same timestamp. They are trades from multiple buyers or sellers or split-transactions occurring when the volume of an order on one side of the market is larger than the available opposing orders. We follow the common practice of aggregating these trades as a single transaction and use the volume-weighted price as the price for this aggregated transaction. Returns are calculated from the logarithm of the current and lagged prices ratios.

We consider two different types of trade events. A first type is where every single transaction is a trade event. In this case, the durations correspond to the time between consecutive trades. A second type is where an event correspond to observing five trades. The model could be estimated for any fixed number of trades and we explain below why we think that five trades is an interesting choice.

Descriptive statistics for durations and returns over consecutive trades are shown in Table 1. This sample contains 60,281 observations, so an average of 6,000 trades per trading day. For the returns, the skewness is negative but close to zero and as expected the kurtosis is much larger than the normal value of three. The distribution of the durations turns out to be skewed to the right and the sample kurtosis is also larger than the normal value of three. It is interesting to know the largest intervals between trades in this sample is around three minutes and of course the minimal interval is one second. The average waiting time for a trade to happen is slightly less than 4 seconds. As for temporal dependence, we see the usual strong negative autocorrelation of order one in the return series, often associated to the price bouncing between between the bid and ask. We also see decaying dependence in the duration series.
but the magnitude of the autocorrelation of order one is smaller for the durations than for the returns.

Table 1 also contains descriptive statistics for returns and durations for the case where a trade event consist of five consecutive trades. For returns, we see a slightly negative skewness and again a kurtosis much greater than three. The median of the returns is zero. In fact, the data includes a high percentage of observations with zero returns. About 34% of returns over five trades are zero, while this proportion is about 46% for returns over individual trades. For duration, we again have positive skewness and excess kurtosis. The median and maximum duration are of course greater when we use five consecutive trades than when we use individual trades. It is interesting to note that when going from individual trades to five consecutive trades, the ACFs suggest that the data is less noisy. The magnitude of the autocorrelation of order one for the return series has shrunk by half and the magnitude of the ACFs for the duration series has more than doubled. By aggregating the data we lose information but on the other hand the signal to noise ratio appears to increase.

We opted to present results for the model estimated with data consisting of return and duration over five consecutive trades based the volatility signature plot. We know that market microstructure noise has many sources: bid-ask spread, price discretization, data recording errors and other properties regarding trading and market mechanism. At the highest sampling frequency, there are so many irregularities in the returns that it is hard to find the right parametric specifications to capture them. As we see from the signature plot in Figure 3, the microstructure effect is substantial at ultra high frequencies but declines and becomes stable when sampling becomes less frequent. From this Figure we can see that for the AMD stock at the point in time (July 2005), if we sample every 20 seconds or more, then the impact of market microstructure noise should not be dominant. Using five transactions as an event, we see in Table 1 that the mean duration is 19.38 seconds.

### 4.2 Estimation results for individual trades

The estimation results with individual transactions of the AMD stock are presented in Table 2 under the two “Individual trades” columns. The first thing to notice is all the estimates are statistically significant. A second thing to look at is the dynamic in the bivariate OU process. The solution of this process over any interval $d$ is a VAR of order one with the persistence equal to $\Phi = \expm(-Ad)$. We
plot the elements of $\Phi$ as a function of $d$ in Figure 4. We can see from this figure that there is limited persistence among the latent variables as the elements of $\Phi$ are all pretty much zero after 30 minutes. So when looking at individual trades, shocks to volatility or trade intensity do not propagate much over time.

There is significant correlation between the different Wiener processes. The two Wiener processes in the OU process are positively correlated ($\hat{\rho}_3 = 0.60251$), so when there is a positive shock to log-variance, then on average there is also a positive shock to the log of the trade intensity, which means shorter durations. This is an expected result. This is consistent with the result of Easley and O’Hara (1987). Factors such as characteristics of the news, market participant and market mechanism affect each individual’s trading decision. Announcement of public news normally induces an increased trading intensity and a substantial up or down movement in the price on the stock market very quickly, while private news has less impact on the market. Informed trader tends to take advantage of the information before certain time without moving market price to the adverse direction. There are type of investors who are more willing to wait and have no tolerance of the relative high price, and type of the investor who want to close the position soonest, such as liquidity trader. Block trade is normally divided into orders of smaller size. The time of waiting for the order execution depends on how much price impact the trader can take. However, we don’t observe these individual level characteristics to identify type of each trade and buyer or seller and news, what we have are variables in a market scale, e.g. stock price, quotes, trade arrival time and volume. But we know all factors that playing roles in affecting the trading intensity and price formation together. The high and positive correlation in our result also justifies the joint modelling of volatility and duration between trades. As for the correlations between the noise in the return equation and the shock to log of the intensity ($\rho_1$) and log of the variance ($\rho_2$), we get the usual negative value for $\rho_2$. This would be the usual leverage effect. As for $\rho_1$, the estimate is quite strong and positive (about 0.7). This means that negative shocks to the return are associated to negative shocks to the log-intensity, which in turns means higher expected durations.

Market microstructure noise has many sources: bid-ask spread, price discretization, data recording errors and other properties of trading and market mechanism. Bid-ask spread and the clustering of order flow are respectively related to negative and positive autocorrelations of the noises. Empir-
ical results in literature report both positive and negative autocorrelation in the noise depending on different stocks and markets. Our estimate for $\theta_{MA}$ shows a positively correlated noise with a sizable standard deviation.

The estimated seasonal patterns in volatility and trade intensity are plotted in Figure 5. We see that there are major changes in the intensity and variance through the day. The intensity $\lambda_t$ and volatility $\sigma_t$ are about 60% and 80% higher respectively when the market opens at 9:30AM than during the middle of the day around 1:00PM. They increase through the afternoon and are about 50% higher when the markets close than round 1:00PM. Durations are shorter at the opening and closing time of the market than around noon. Volatility is highest at the opening of the market, and lowest at the noon. The more frequent transaction accompanied by more fluctuations in price changes at the opening is the reaction of market participants to financial news released overnight and stock performance in other markets around the world. When the time is near the close of the market, agents wish to close their position instead of waiting to avoid potential uncertainty. In the middle of the day, transactions become less frequent because the agents get involved in other activities, such as taking lunch break, and we observe less movement in stock prices.

With the Kalman filter, we obtain the filtered estimates for the latent variables. In Figure 6 we plot the log of the trade intensity (panel a) and the log of the variance (panel b). The time series plot display a moderate level of persistence and a clustering effect.

Finally, we need to explain how we deal with the zero returns in the estimation. Similar to Harvey, Ruiz, and Shephard (1994), we add a tiny positive number (0.00008) to the squared returns.

4.3 Estimation results for combined trades

The estimation results with combined transactions of the AMD stock are presented in Table 2 under the two “Combined trades (5) - Endogenous durations” columns. As explained above, we take five consecutive transactions as a trade event. By reducing the sampling frequency, we see that the data is indeed less noisy and we can better capture the underlying dependence in the process.

A first important difference between the results for the combined trades versus the individual trades is that we capture a much higher level of dependence and co-dependence in the bivariate OU
Looking again at the persistence $\Phi = \exp(m(-Ad))$ as a function of $d$, we see the persistence decreasing much more slowly across the board. This would imply that the sequence of durations would now be much more informative about what the volatility is. A second difference to point out is the leverage effect has changed sign, $\hat{\rho}_2$ is now positive, but it is not statistically significant anymore.

Another sign that the combined trades are less noisy is the variance of the market microstructure noise. Although the MA(1) persistence is greater now ($\hat{\theta}$ is very close to one), the standard deviation $\hat{\sigma}_\eta$ is pretty much zero. As for the seasonal pattern in volatility and intensity, Figure 5 indicate stronger patterns for the combined data. The most quiet part of the day is still around 1:00PM but at the market’s opening the intensity and volatility are now 100% and 120% higher than around 1:00PM, compared to 60% and 80% we obtained with individual trades.

With a higher level of persistence in the OU process for the latent variables, the filtered log-intensity and log-variance are different than those obtained with individual trades. The series are plotted in Figure 8. We indeed see a higher level of persistence in the series compared to the series in Figure 6, corresponding to individual trades.

A first way of illustrating the gains from jointly modeling the returns and durations is to compare the results above with the results from a simplified version of our model where the durations are exogenous, i.e. no link between durations and intensity with returns and volatility: $a_{12} = a_{21} = \rho_1 = \rho_3 = 0$. Estimation results for this model are presented in Table 2 under the two “Combined trades (5) - Exogenous durations” columns. It is particularly interesting to compare the filtered estimates of the log-variance between the endogenous and exogenous duration models. They are plotted in Figure 9. Taking into account the log-scale of the series, we can see differences exceeding 20% in the filtered estimates of the variance when we don’t (incorrectly) assume that the durations are exogenous.

The filtered log-variance estimates give us estimates of the spot volatility. From these estimates, we can compute an estimate of the integrated variance, $\int_{t_i}^{t_f} \sigma_t dt$, for any length of time (one day for example). According to our model, the filtered log-variances are normally distributed with conditional means and variances given by the Kalman filter. From the properties of the log-normal distribution, we know that $E[\sigma_t^2] = \exp\left(E[\ln(\sigma_t^2)] + \frac{\text{var}[\ln(\sigma_t^2)]}{2}\right)g_a(t)^2$. We can take the sum of these predicted variances times the corresponding duration to get an estimate of the integrated variance. A non-parametric alternative to estimate the integrated variance is the realized volatility (RV) estimator.
and its refinements such as the method proposed by Hansen and Lunde (2004). The basic RV estimator consists in taking returns over fixed intervals (five minutes in our application), squaring them and taking their sum. In the absence of market microstructure noise, the RV estimator will consistently estimate the integrated variance as length of the time intervals shrink to zero. Estimates of the integrated variance with our model, RV and the Hansen-Lunde (HL) method are reported in Table 3. Our estimates are generally higher than HL, which are higher than RV.

5 Monte Carlo Simulation

In this section, we extend the comparison of the results between the models where the durations are either assumed endogenous or endogenous. We use Monte Carlo simulations to study the loss of accuracy in filtering the variance. The setup is the following. We simulate data from a model with endogenous durations, then we compared the filtered estimates of the variance from (1) a model allowing for endogenous durations estimated with the simulated data and (2) the restricted version of our model where the durations are assumed exogenous.

We simulate the joint dynamics of returns and durations according to the model described in Section 2 with the estimates corresponding to the individual trades reported in Table 2. Results for the combined trades will be added in the next draft of the paper. Each simulation is done with 20 days’ worth of observations (more than 60,000 trades). The models are re-estimated for each sample and each time the filtered estimates of the variance are kept and compared to the true variance. We report the Mean Square Error (MSE) and the Mean Absolute Error (MAE). The results for the MSE and MAE for each of the 15 replications are reported in Table 4. We can see that, not surprisingly, for 14 out of 15 simulations the MSE or MAE of the model that allows for endogenous durations is smaller than for the model with exogenous durations. Averaging the results over the 15 replications, we see that the MSE and MAE of the exogenous duration model are 29% and 11% bigger than for the endogenous duration model.

---

2This realized volatility estimator is computed by summing up squared returns over one minute intervals and then adjusting with autocorrelation up to 10 minutes.
6 Conclusion

We believe that the time between trades carries information about the volatility. To capture this information we develop a joint for the joint evolution of returns and time between trades (durations). The returns follow a basic stochastic volatility model and the durations follow an exponential distribution conditional on a latent trade intensity. The two latent variables, log-variance and log-intensity, follow a bivariate OU process. We propose a classical estimation procedure based on a linear state-space representation of the model. In an application with tick-by-tick transaction data for the AMD stock, we show that the time between trades is endogenous and by conditioning on trade durations we can get more accurate measurements of volatility than if we were to only use returns.
References


24


Table 1: Summary Statistics for the duration and return series for AMD for individual transactions (1) and for sequences of five transactions (5). The data is over the period July 7, 2005 to July 20, 2005.

<table>
<thead>
<tr>
<th></th>
<th>Return (1)</th>
<th>Duration (1)</th>
<th>Return (5)</th>
<th>Duration (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Obs</td>
<td>60281</td>
<td>60281</td>
<td>12053</td>
<td>12053</td>
</tr>
<tr>
<td>Mean</td>
<td>0.000213</td>
<td>3.876329</td>
<td>0.001068</td>
<td>19.38372</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>0.052345</td>
<td>4.905678</td>
<td>0.067154</td>
<td>13.86797</td>
</tr>
<tr>
<td>Min</td>
<td>-1.67327</td>
<td>1</td>
<td>-1.61627</td>
<td>5</td>
</tr>
<tr>
<td>Max</td>
<td>1.673267</td>
<td>184</td>
<td>1.616266</td>
<td>208</td>
</tr>
<tr>
<td>Median</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.1716</td>
<td>4.9663</td>
<td>-0.1446</td>
<td>2.5401</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>111.8810</td>
<td>64.1980</td>
<td>83.6957</td>
<td>14.8331</td>
</tr>
<tr>
<td>ACF(1)</td>
<td>-0.4253</td>
<td>0.1729</td>
<td>-0.20637</td>
<td>0.3867</td>
</tr>
<tr>
<td>ACF(2)</td>
<td>-0.0033</td>
<td>0.1401</td>
<td>-0.01336</td>
<td>0.3283</td>
</tr>
<tr>
<td>ACF(3)</td>
<td>-0.0044</td>
<td>0.1352</td>
<td>0.02292</td>
<td>0.3210</td>
</tr>
<tr>
<td>ACF(4)</td>
<td>0.0065</td>
<td>0.1222</td>
<td>-0.00664</td>
<td>0.3067</td>
</tr>
<tr>
<td>ACF(5)</td>
<td>0.0120</td>
<td>0.1145</td>
<td>0.0117</td>
<td>0.3000</td>
</tr>
</tbody>
</table>
Table 2: Estimation results for AMD stock data.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Individual trades</th>
<th>Combined trades (5)</th>
<th>Combined trades (5)</th>
<th>Combined trades (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates</td>
<td>Std. Dev.</td>
<td>Endogenous durations</td>
<td>Exogenous durations</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-2.00965</td>
<td>0.0057862</td>
<td>-3.98381</td>
<td>0.038867</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-8.63638</td>
<td>0.015605</td>
<td>-9.77690</td>
<td>0.075089</td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.0033857</td>
<td>0.00005936</td>
<td>5.435E-05</td>
<td>0.0006369</td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.0036804</td>
<td>0.00007108</td>
<td>8.418E-05</td>
<td>0.0007729</td>
</tr>
<tr>
<td>$a_{12}$</td>
<td>-0.00079256</td>
<td>0.00002873</td>
<td>-0.0002425</td>
<td>2.116E-05</td>
</tr>
<tr>
<td>$a_{21}$</td>
<td>-0.0062666</td>
<td>0.00009638</td>
<td>-0.001161</td>
<td>9.197E-05</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.010929</td>
<td>0.00013494</td>
<td>0.0040779</td>
<td>0.002026</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.046178</td>
<td>0.00074167</td>
<td>0.012500</td>
<td>0.0023874</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>1.44350</td>
<td>0.00505204</td>
<td>6.951E-07</td>
<td>1.455E-07</td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.69339</td>
<td>0.0015082</td>
<td>0.69748</td>
<td>0.014142</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>-0.15533</td>
<td>0.0036103</td>
<td>0.076837</td>
<td>0.090509</td>
</tr>
<tr>
<td>$\rho_3$</td>
<td>0.60251</td>
<td>0.0035879</td>
<td>0.76808</td>
<td>0.068331</td>
</tr>
<tr>
<td>$\alpha_{\lambda,1}$</td>
<td>0.052346</td>
<td>0.0020090</td>
<td>0.081815</td>
<td>0.010052</td>
</tr>
<tr>
<td>$\alpha_{\lambda,2}$</td>
<td>-3.47155</td>
<td>0.015535</td>
<td>-3.71089</td>
<td>0.073415</td>
</tr>
<tr>
<td>$\alpha_{\sigma,1}$</td>
<td>0.058698</td>
<td>0.0028050</td>
<td>0.071818</td>
<td>0.011923</td>
</tr>
<tr>
<td>$\alpha_{\sigma,2}$</td>
<td>-3.68816</td>
<td>0.0058811</td>
<td>-4.092713</td>
<td>0.129628</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.56622</td>
<td>0.0021075</td>
<td>0.99930</td>
<td>0.001722</td>
</tr>
</tbody>
</table>

Table 3: Daily volatility measurements for AMD Stock. First two columns are the integrated volatilities estimated from the full model (endogenous durations) and partial model (exogenous durations). The third column is a basic realized volatility measurement using 5 minutes returns. The last column is a realized volatility measurement computed with the method introduced by Hansen and Lunde (2004).

<table>
<thead>
<tr>
<th>Day</th>
<th>Full</th>
<th>Partial</th>
<th>RV(5min)</th>
<th>HL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.7248</td>
<td>5.0745</td>
<td>5.2791</td>
<td>5.1948</td>
</tr>
<tr>
<td>2</td>
<td>3.6976</td>
<td>3.6964</td>
<td>2.4531</td>
<td>2.8715</td>
</tr>
<tr>
<td>3</td>
<td>4.1648</td>
<td>4.2148</td>
<td>2.2244</td>
<td>2.3264</td>
</tr>
<tr>
<td>4</td>
<td>3.3171</td>
<td>3.4288</td>
<td>2.5984</td>
<td>2.2222</td>
</tr>
<tr>
<td>5</td>
<td>4.1407</td>
<td>4.2618</td>
<td>2.6626</td>
<td>3.0894</td>
</tr>
<tr>
<td>6</td>
<td>8.5538</td>
<td>8.4149</td>
<td>6.3278</td>
<td>9.0669</td>
</tr>
<tr>
<td>7</td>
<td>3.9300</td>
<td>4.0367</td>
<td>2.0171</td>
<td>2.1958</td>
</tr>
<tr>
<td>8</td>
<td>2.5087</td>
<td>2.5827</td>
<td>1.1797</td>
<td>1.5214</td>
</tr>
<tr>
<td>9</td>
<td>4.9339</td>
<td>5.0154</td>
<td>2.7206</td>
<td>3.6448</td>
</tr>
<tr>
<td>10</td>
<td>4.1641</td>
<td>4.2617</td>
<td>3.9215</td>
<td>3.5833</td>
</tr>
<tr>
<td>Sum</td>
<td>44.1354</td>
<td>44.9877</td>
<td>31.3843</td>
<td>35.7164</td>
</tr>
</tbody>
</table>
Table 4: MSE and MAE of the filtered variances for the models where the durations are either endogenous or exogenous.

<table>
<thead>
<tr>
<th>Replication</th>
<th>MSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endogenous</td>
<td>Exogenous</td>
</tr>
<tr>
<td>1</td>
<td>0.262008</td>
<td>0.302045</td>
</tr>
<tr>
<td>2</td>
<td>0.260519</td>
<td>0.345235</td>
</tr>
<tr>
<td>3</td>
<td>0.261878</td>
<td>0.312254</td>
</tr>
<tr>
<td>4</td>
<td>0.270338</td>
<td>0.303050</td>
</tr>
<tr>
<td>5</td>
<td>0.270338</td>
<td>0.315043</td>
</tr>
<tr>
<td>6</td>
<td>0.261898</td>
<td>0.290822</td>
</tr>
<tr>
<td>7</td>
<td>0.250870</td>
<td>0.419258</td>
</tr>
<tr>
<td>8</td>
<td>0.317047</td>
<td>0.296896</td>
</tr>
<tr>
<td>9</td>
<td>0.299483</td>
<td>0.351124</td>
</tr>
<tr>
<td>10</td>
<td>0.250499</td>
<td>0.345299</td>
</tr>
<tr>
<td>11</td>
<td>0.271676</td>
<td>0.317399</td>
</tr>
<tr>
<td>12</td>
<td>0.251818</td>
<td>0.289465</td>
</tr>
<tr>
<td>13</td>
<td>0.259300</td>
<td>0.309527</td>
</tr>
<tr>
<td>14</td>
<td>0.276145</td>
<td>0.298936</td>
</tr>
<tr>
<td>15</td>
<td>0.236187</td>
<td>0.317546</td>
</tr>
<tr>
<td>mean</td>
<td>0.248561</td>
<td>0.320927</td>
</tr>
</tbody>
</table>
Figure 1: Sample path from a bivariate OU process with a transition matrix $A$ with complex eigenvalues. The transition matrix is equal to $A = \begin{bmatrix} 0.005 & 0.6 \\ -0.6 & 0.007 \end{bmatrix}$ and the eigenvalues are $\text{eig}(A) = \begin{bmatrix} 0.006+0.6i \\ 0.006-0.6i \end{bmatrix}$. 
Figure 2: Sample path from a bivariate OU process with a transition matrix $A$ with real eigenvalues. The transition matrix is equal to $A = \begin{bmatrix} 0.001 & 0.0015 \\ 0.002 & 0.0 \end{bmatrix}$ and the eigenvalues are $\text{eig}(A) = \begin{bmatrix} 0.000678174 \\ 0.0100321825 \end{bmatrix}$. 
Figure 3: Signature plot for AMD (daily realized volatility against sampling frequency in seconds)
Figure 4: Impulse-response for the model estimated with individual trades
Figure 5: Plot of the daily trends for duration and volatility.
Figure 6: Filtered estimates of the latent log-trade intensity and log-variance.
Figure 7: Impulse-response for the model estimated with observations corresponding to five consecutive trades
Figure 8: Filtered Estimates, AMD (5 trades, 10 days’ data in July 2005). Filtered estimates are computed based on the estimators. (a) is the plot for estimated log intensity for AMD stock during this 10 day period; (b) is the plot for estimated log volatility for AMD stock during this 10 day period.
Figure 9: Filtered log-variance for the models with endogenous and exogenous durations. Panel (a) presents the series for the ten days of data used to estimate the model while panel (b) presents the results for the first 1500 observations.
Appendix

\[ \mu^* = E[\omega_i | s_t] \]
\[ = E[E[\omega_i | u_t] | s_t] \]
\[ = E[\Sigma_t^{n,\omega}(d_i)^{-1} u_t | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} E[u_t | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} \sqrt{d_i} \sqrt{2/\pi} s_t \]
\[ = \frac{0.7979}{\sqrt{d_i}} \Sigma_t^{n,\omega} s_t \]

and

\[ \Sigma_t^{n,\omega} = Cov(\omega_i, \varepsilon_t | s_t) \]
\[ = E[\omega_t \varepsilon_t | s_t] - E[\omega_t | s_t]E[\varepsilon_t | s_t] \]
\[ = E[E[\omega_t \varepsilon_t | u_t] | s_t] - E[\varepsilon_t | s_t] \ E[\omega_t | s_t] \]
\[ = E[s_t E[\ln(u_t^2) - E[\ln(u_t^2)]] | s_t] \]
\[ = E[\mu^* E[\ln(u_t^2) - E[\ln(u_t^2)]] | s_t] \]
\[ = E[\mu^* E[\ln(u_t^2)] - E[\ln(u_t^2)] | s_t] \]
\[ = E[\varepsilon_t E[\ln(u_t^2)] | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} E[\varepsilon_t u_t | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} E[u_t \ln(u_t^2) - E[\ln(u_t^2)] | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} E[u_t \ln(u_t^2) - E[\ln(u_t^2)] | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} E[u_t \ln(u_t^2)] - \Sigma_t^{n,\omega}(d_i)^{-1} E[\ln(u_t^2)] E[u_t | s_t] \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1} E[u_t \ln(u_t^2)] s_t - \Sigma_t^{n,\omega}(d_i)^{-1} (\ln(d_i) - 1.2704) \sqrt{d_i} \sqrt{2/\pi} s_t \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1/2} [0.7979 \ln(d_i) + 0.0925] s_t - \Sigma_t^{n,\omega}(d_i)^{-1/2}(\ln(d_i) - 1.2704) \sqrt{2/\pi} s_t \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1/2} [0.7979 \ln(d_i) + 0.0925] s_t - \Sigma_t^{n,\omega}(d_i)^{-1/2}(0.7979 \ln(d_i) - 1.2704) s_t \]
\[ = \Sigma_t^{n,\omega}(d_i)^{-1/2} [0.0925 + 0.7979 \times (-1.2704)] s_t \]
\[ = 1.1061 \Sigma_t^{n,\omega}(d_i)^{-1/2} s_t \]