A Jump Diffusion Model for Volatility and Duration

Wei Wei*

6th April 2013

Abstract

This paper puts forward a stochastic volatility and stochastic conditional duration with cojumps (SVSDCJ) model to analyze returns and durations. In high frequency data, transactions are irregularly spaced, and the durations between transactions carry information about volatility as suggested by the market microstructure theory. Traditional measures of volatility do not utilize durations. I adopt a jump diffusion process to model the persistence of intraday volatility and conditional duration, and their interdependence. The jump component is disentangled from the continuous part of the price, volatility and conditional duration process. I develop a MCMC algorithm for the inference of irregularly spaced multivariate process with jumps. The algorithm provides smoothed estimates of the latent variables such as spot volatility, jump times and jump sizes. I apply this model to IBM data and I find meaningful relationship between volatility and conditional duration. Also, jumps play an important role in the total variation, but the jump variation is smaller than traditional measures that use returns sampled at lower frequency.

*Department of Economics, North Carolina State University, Raleigh, NC 27606, USA, email: wwei6@ncsu.edu

1 Introduction

The recent availability of high frequency data has provided an unprecedented opportunity to look into financial markets at a microscopic level. With this type of data, every transaction is recorded. For various reasons, it is common to aggregate the individual trades over a fixed time interval such as five minutes. At this level of aggregation, high frequency returns exhibit fat tails, volatility clustering, and jumps similar to returns obtained at lower frequencies. With daily or lower frequency returns, these features have inspired GARCH and stochastic volatility models to capture the predictability of volatility. High frequency data has advanced another volatility measure: realized volatility. This non-parametric estimator uses returns sampled at shorter horizon (such as 5 minutes) to measure the volatility at a longer horizon (such as a day). The theoretical foundation was laid by Andersen, Bollerslev, Diebold, and Labys (2001), Andersen, Bollerslev, Diebold, and Labys (2003), Barndorff-Nielsen and Shephard (2001) and Barndorff-Nielsen and Shephard (2002). Since then, a huge literature has been devoted to the implementation of this estimator.

Realized volatility measures do not utilize the persistence of volatility or estimate intraday spot volatility. Also, fixed time aggregation loses potentially valuable information, such as the durations between transactions. I propose a jump diffusion process to model the movement of returns and durations jointly. My stochastic volatility and stochastic duration with cojumps (SVSDCJ) model helps fill the gap between traditional stochastic volatility models and irregularly spaced high frequency data. With this model, I can measure intraday volatility by exploiting the persistence of volatility as well as the information conveyed in durations. I also disentangle the jump component with the continuous part and estimate the jump variation.

The asymmetric information models of market microstructure suggest that the durations between trades provide information to market participants. Both the presence and the absence of trade impacts price adjustments. In the seminal work of Easley and O'Hara (1992), a fraction of traders are informed with a signal (news). Informed traders buy or sell only when they observe a good or bad signal. A long interval between trades is more likely to occur when no news has occurred. Increased trading intensity is associated with an information event and increased number of informed traders (See Dufour and Engle (2000) for an empirical study on how durations impact the price dynamics). Moreover, one would expect to observe that short durations are followed by short durations (duration clustering), and periods of high volatility tend to be grouped (volatility clustering). Their model also implies that long durations have negative impact on volatility, and vice versa.

Duration clustering has led to a large literature working on the direct modeling of durations. Following the idea behind GARCH, Engle and Russell (1998) propose the autocorrelated conditional duration (ACD) model. In their model, the conditional expected duration depends on past duration, return and other economic variables, such as volume or bid-ask spread. Bauwens and Giot (2000) suggest modeling the logarithm of durations. It is more flexible and does not impose parameter restrictions to ensure that durations are positive. Bauwens and Veredas (2004) put forth the stochastic conditional duration (SCD) model, which is analogous to the stochastic volatility model. In the SCD model, the expected duration becomes stochastic; I model durations in a similar fashion in this paper.

The framework of Easley and O'Hara also predicts interdependence between durations and volatility. Engle (2000) applies ACD models to IBM and examines the impact of durations on volatility. He imposes exogeneity on the duration process but allows volatility to be influenced by durations under the GARCH framework. His finding supports the Easley and O'Hara theory in which short duration leads to higher volatility, no trade being interpreted as no news. Grammig and Wellner (2002) extend Engle's model to analyse the interaction between volatility and duration. In particular, they consider the impact of volatility on trading intensity. They conclude that lagged volatility lengthens expected durations. Manganelli (2005) uses a vector autoregressive (VAR) model to incorporate volume. He allows return and volatility to interact with durations and volume. He finds that short durations follow large returns, which is in line with Easley and O'Hara theory, but the result only applies to frequently traded stocks. Ghysels and Jasiak (1998) note that the class of ACD-GARCH models can be interpreted as time deformed GARCH diffusion. Their empirical study finds that volatility has a causal relationship with durations.

Following Zheng and Pelletier (2012), I focus on the stochastic class of volatility and duration modelling, where the latent volatility and conditional duration reflect unobservable information flow. Returns are normally distributed with stochastic volatility and jumps; durations are exponentially distributed with stochastic conditional duration. The logarithm of conditional duration and the logarithm of volatility follow a bivariate Ornstein-Uhlenbeck (OU) process. The OU process is mean reverting and when discretized, it leads to a VAR model. This specification relies on two insights: first, volatility and durations are persistent, hence the conditional volatility/duration will be affected by their own past. Second, as predicted by microstructure theory and empirically documented by the ACD-GARCH models, volatility and conditional duration interact with each other. The bivariate OU process allows the expected volatility and duration to depend on the lagged value as well as correlated shocks. The estimated spot volatility is a natural supplement to realized volatility: realized volatility measures the integration of the spot volatility when no jumps are present.

The presence of jumps is an important feature of financial returns. Merton (1976) first describes returns using a continuous diffusion process and a Poisson-driven jump process. Jumps are interpreted as "abnormal" variation in price due to the arrival of important news. It is important to separate the jump component and the diffusion component in price because they are two fundamentally different sources of risk. Jump risk has different hedging possibilities and requires a different premium.

In the high frequency setting, the non-parametric realized volatility has led to a non-parametric estimator for jump variation. Barndorff-Nielsen (2004) introduce the idea of realized bipower variation, which is the summation of the cross product of return. Suppose the price process has both diffusion part and jump part, then the difference between realized bipower variation and realized volatility is a measure for the quadratic variation from the jump component. Using this tool, recent literature has suggested that jumps play an important role in the total variation of price. For reviews on realized volatility and jumps, see Andersen, Bollerslev, and Diebold (2007) and Barndorff-Nielsen and Shephard (2005).

I utilize a Merton type jump diffusion process to model price. Jumps in returns generate infrequent large movements and contribute to the fat tails in the return distribution. Without jumps, volatility needs to be extreme high to explain the occasional large fluctuations. I also consider jumps in volatility and expected durations. As Eraker, Johannes, and Polson (2003) noted, jumps in return don't affect future returns while jumps in volatility can produce a period of extreme price movements. Jumps in expected duration are included to explain the burst in transactions that accompanies the burst in volatility.

One of the most challenging complications in dealing with high frequency data is the existence of microstructure noise. Theoretically, the sum of squared return converges in probability to quadratic variation when the sampling frequency goes to infinity. However, the observed price is composed of efficient price and a noise component. Even if the noise is iid, the return will consist of efficient return and an autocorrelated noise, and the realized volatility will be a biased estimator of the actual volatility. As the sampling frequency increases, the noise to signal ratio will get higher as well.

In the realized volatility literature, there are several approaches to deal with the microstructure noise. The simplest way is to sample sparsely, for example every 5 minutes. One can also determine the sampling frequency by minimizing the mean squared error, following Russell and Bandi (2004). Bandi and Russell (2006) suggest using data at different frequencies to separate noise from volatility. Zhang, Mykland, and Ait-Sahalia (2005) proposed an estimator that utilizes subsampling, averaging and bias-correction, where the variance of the microstructure noise is estimated through the variance of returns sampled at the highest frequency. Zhou (1996) and Hansen and Lunde (2006) use the autocorrelation of returns to construct kernel-based volatility estimator. Ait-Sahalia, Mykland, and Zhang (2005) show that if the the noise term is accounted for explicitly, sampling as often as possible is optimal.

My approach in dealing microstructure noise combines several different methods. First, I model noise terms explicitly. Noise is treated as a latent variable and it is estimated in the model. Second, I sample from every Lth transaction. This sampling scheme is referred to as transaction time sampling (Oomen (2006)) or tick time sampling (Hansen and Lunde (2006)) in the realized volatility literature. Third, the autocovariance of tick-by-tick returns serves as a measure of the variance of noise. This combined approach is unique to the estimation procedure I adopt and it allows sampling at finer grid than current parametric models.

Estimating stochastic volatility model usually involves approximation. Jacquier, Polson, and Rossi (1994) introduced Bayesian Markov Chain Monte Carlo (MCMC) methods which allows for exact finite sample inference. Eraker, Johannes, and Polson (2003) use MCMC to analyse the impact of jumps in returns and volatility. I resort to MCMC for the estimation of the parameters and latent variables. My model can be viewed as a nonlinear state space model in which volatility, conditional durations, and jumps are the state variables. The observation equation describes how returns and durations change given state variables, and the evolution equation is the dynamics of state variables. One major benefit of using MCMC is that both parameters and state variables are estimated simultaneously instead of using filtering technique ad hoc. The estimated conditional volatility and jumps are very useful in applications such as Value at Risk. Another benefit of MCMC is that I can incorporate prior information properly. For example, the noise variance estimated from the tick-by-tick returns can be used to form an informative prior. Also, if jumps are interpreted as infrequent and large movements, I can use appropriate prior to elicit such beliefs.

The rest of this paper is organized as follows. In section 2 I describe the model specification. Section 3 discuss the Bayesian inference and simulation studies. Section 4 presents the empirical results using IBM data. Section 5 concludes.

2 Model Specification

2.1 Setup

I start by assuming that the logarithmic asset price y_t follows the jump diffusion process

$$dy_t = \mu^y dt + \sqrt{V_t} dW_t^y + \xi_t^y dN_t^y, \tag{1}$$

where V_t is the latent spot volatility which follows a separate stochastic process, and W_t^y denotes a standard Brownian Motion. For simplicity, I assume that V_t and W_t^y are independent. Jumps follow a compound Poisson process since we're interested in large and infrequent price movements. Jump arrivals are assumed to be state independent, i.e., the jump intensity γ is constant. Given a time interval Δ, the probability of observing n jumps is $e^{-\gamma\Delta}(\gamma\Delta)^n/n!$. Jump sizes ξ_t^y are also random.

The duration D_{i+1} is defined as the time interval between an event that occurred at t_i and the next event at t_{i+1}. In my application, I'm sampling every Lth transaction; the event is defined as L transactions. For example, if $L = 100$, D_{i+1} measures the time it takes to observe 100 transactions. Let λ_{t_i} denote the conditional expectation of D_{i+1} given information set available at t_i, $E(D_{i+1}|I_{t_i}) = \lambda_{t_i}$. Following most financial duration models, D_{i+1} is modeled as λ_{t_i} times a i.i.d random variable with positive support, i.e., $D_{i+1} = \lambda_{t_i} e_i$. I assume an exponential distribution for e_t in this paper. It is noted that my model can be easily extended to accommodate other distributions.

To create persistence and interdependence between volatility and duration, I follow Zheng and Pelletier (2012) and model the logarithm of λ_t and V_t using a bivariate OU process. As noted by Andersen, Bollerslev, Diebold, and Ebens (2001), the logarithmic volatility are closer to being normal than the raw volatility. Also, modelling logarithmic volatility and duration has the benefit of imposing non-negativity without putting extra constraint on parameters. To explain a prolonged effect from news, I consider adding a Poisson jump component to the Gaussian OU process. Let $X_t = (\log(V_t), \log(\lambda_t))'$, X_t solves:

$$dX_t = -\Psi(X_t - \mu^x)dt + S_x dW_t^x + \xi_t^x dN_t^x, \tag{2}$$

where Ψ is a 2×2 matrix that measures the mean reversion and dependence between conditional duration and volatility. The OU process mean reverts to μ^x, the diffusive long-run mean. S_x measures the variation of the logarithmic volatility and the logarithmic duration, and $S_x = diag(\sigma_v, \sigma_\lambda)$. W_t^x is a Brownian motion in $\mathcal{R}^2$ with $dW_t^v dW_t^\lambda = \rho dt$, where ρ is the instantaneous correlation. The instantaneous covariance matrix is given by

$$\Sigma_x = S_x \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} S_x = \begin{pmatrix} \sigma_v^2 & \rho\sigma_v\sigma_\lambda \\ \rho\sigma_v\sigma_\lambda & \sigma_\lambda^2 \end{pmatrix}.$$

N_t^x is a Poisson process in $\mathcal{R}^2$, $N_t^x = (N_t^v, N_t^y)'$. Jumps are interpreted as unexpected important news, and when a jump occurs, it occurs to price, volatility and conditional durations. In other words, I model jumps with contemporaneous arrivals, $N_t^y = N_t^v = N_t^\lambda$. This type of jumps is referred to as cojumps. Since cojumps reflect impacts from the same news, the jump sizes in returns, volatility and conditional durations are correlated. Let ξ_t denote the vector of jump sizes in returns, volatility and conditional duration, $\xi_t = (\xi_t^y, \xi_t^v, \xi_t^\lambda)'$, we specify a multivariate normal distribution for ξ_t:

$$\xi_t \sim N(\mu_J, \Sigma_J). \tag{3}$$

Jumps in volatility account for the rapid increases in volatility that have been observed in financial markets. The persistence of volatility also allows periods of large price movement before volatility reverts to its diffusive long-run mean. Jumps in conditional duration explain the rapid change in trading intensity associated with the rapid change in volatility.

The continuous time framework allows straightforward discretization of the model over unequally spaced data. Using an Euler approximation, I discretize dy_t over the durations:

$$y_{i+1} - y_i = r^e_{i+1} = \mu^y D_{i+1} + \sqrt{V_t D_{i+1}}\epsilon^y_{i+1} + \xi^y_{i+1} J_{i+1}, \tag{4}$$

where the subscript i denotes the time of ith event, t_i. Jumps are assumed to be rare, γ is close to zero, so the probability of observing no jumps in the time interval D_{i+1} can be approximated by $1 - \gamma D_{i+1}$. Furthermore, there is at most one jump in D_{i+1}, with $Pr(J_{i+1} = 1|\gamma) = \gamma D_{i+1}$. Hence, J_{i+1} is referred to as jump indicator.

The efficient log price process y_t is unobservable in high frequency financial data due to market frictions. The observed logarithmic price is the sum of the logarithmic efficient price and market microstructure noise,

$$y^o_i = y_i + m_i. \tag{5}$$

I assume that microstructure noise is i.i.d mean zero and normal, $m_i \sim N(0, \sigma^2_m)$, and that the microstructure noise is independent of the efficient price. The observed return has a MA(1) contamination:

$$r^o_{i+1} = y^o_{i+1} - y^o_i = r^e_{i+1} + m_{i+1} - m_i. \tag{6}$$

The logarithmic volatility and conditional duration X_t is discretized using the exact solution of OU process with jumps (See Section 2.2 for descriptions of the solution). Equation (4), (6), and the time discretization of X_t over durations form the SVSDCJ model:

$$\begin{aligned} r^o_{i+1} &= \mu^y D_{i+1} + \sqrt{V_i D_{i+1}}\epsilon^y_{i+1} + \xi^y_{i+1} J_{i+1} + m_{i+1} - m_i \\ X_{i+1} &= (I_2 - e^{-\Psi D_{i+1}})\mu^x + e^{-\Psi D_{i+1}} X_i + \xi^x_{i+1} J_{i+1} + U_{i+1}, \end{aligned} \tag{7}$$

where

$$U_{i+1} \sim N(0, \Sigma_{i+1})$$
$$vec(\Sigma_{i+1}) = (\Psi \oplus \Psi)^{-1}(I_2 - e^{-(\Psi\oplus\Psi)D_{i+1}})vec(\Sigma_x).$$

2.2 Properties

The exact solution to the SDE (2) is given by

$$X_t = (I_2 - e^{-\Psi t})\mu^x + e^{-\Psi t}X_0 + U_t + Z_t,$$

where $Z_t = \sum_{j=0}^{N_t} e^{-\Psi(t-\tau_j)}\xi^x_{\tau_j}$ (See Kloeden and Platen (1992); Ross (1996)). Assuming that there is at most one jump in D_{i+1} with jump probability γD_{i+1}, we would arrive at the discretize version in (7). Note that the long-run or unconditional mean of X_t is no longer μ^x in the presence of jumps. Under the assumption of contemporaneous jump arrivals and constant intensity, the long-run mean and variance of X_t is given by

$$E(X_t) = \mu^x + \lambda\Psi^{-1}\mu^x_J$$
$$vec(Var(X_t)) = (\Psi \oplus \Psi)^{-1}\left(vec\left(\Sigma_X\right) + \lambda vec\left(\left(\mu^x_J\right)\left(\mu^x_J\right)' + \Sigma^x_J\right)\right). \quad (8)$$

To ensure the existence of a stationary solution, it is necessary that Ψ has only eigenvalues with positive real parts so that $e^{-\Psi t} \to 0$ as $t \to 0$ (See Gardiner (2009)). I discuss two important subsets of the parameter space. First, if $\Psi_{11} > 0$, $\Psi_{22} > 0$, Ψ_{12} and Ψ_{21} have the same sign, and $\det(\Psi) > 0$, then all the eigenvalues of Ψ will be real and positive. In this case, the system reverts to its diffusive mean following an exponential decay. This encompass the case when Ψ is diagonal and the diagonal elements are positive. Second, if $\Psi_{11} > 0$, $\Psi_{22} > 0$, Ψ_{12} and Ψ_{21} has opposite sign, and

$(\Psi_{11} - \Psi_{22})^2 < -4\Psi_{12}\Psi_{21}$, the eigenvalues of Ψ have positive real parts with imaginary parts, and the the system oscillates to the diffusive mean.

I use a Euler discretization for X_t without jumps to gain some insight about the parameters:

$$X_{i+1} = \Psi\mu^x D_{i+1} + (I_2 - \Psi D_{i+1})X_i + \Sigma_x^{1/2}\sqrt{D_{i+1}}\epsilon_{i+1}^x,$$

rearranging,

$$\begin{pmatrix} \log V_{i+1} - \mu^v \\ \log \lambda_{i+1} - \mu^\lambda \end{pmatrix} = \begin{pmatrix} 1 - \Psi_{11}D_{i+1} & -\Psi_{12}D_{i+1} \\ -\Psi_{21}D_{i+1} & 1 - \Psi_{22}D_{i+1} \end{pmatrix} \begin{pmatrix} \log V_i - \mu^v \\ \log \lambda_i - \mu^\lambda \end{pmatrix} + \Sigma_x^{\frac{1}{2}}\sqrt{D_{i+1}}\epsilon_{i+1}^x.$$

The persistence in the logarithmic volatility and conditional duration are measured by Ψ_{11} and Ψ_{22}, respectively. If Ψ_{11} are positive and close to zero, volatility is highly persistent and the speed of mean-reversion is low. Ψ_{12} is the feedback effect from conditional duration to volatility. If Ψ_{12} is positive, longer duration will lead to lower volatility, and vice versa. Ψ_{21} is the impact of lagged volatility on duration. If Ψ_{21} is positive, high volatility will have a negative impact on expected duration. The instantaneous correlation between volatility and expected duration is measured by ρ. Easley and O'Hara theory predicts positive value for Ψ_{12} and Ψ_{21}, and a negative value for ρ.

Todorov and Tauchen (2011) find strong evidence of cojumps in volatility and price using a nonparametric framework. They also find that almost all of the common jumps in price and volatility occur in opposite directions. In particular, a negative price jump is usually associated with a positive jump in volatility. This dependence suggests jumps are an important source for leverage effect. To accommodate this correlation, I assume the following covariance matrix for jumps sizes:

$$\Sigma^J = \begin{pmatrix} \sigma_{J,y}^2 & \rho_{yv}\sigma_{J,y}\sigma_{J,v} & \rho_{yd}\sigma_{J,y}\sigma_{J,d} \\ \rho_{yv}\sigma_{J,y}\sigma_{J,v} & \sigma_{J,v}^2 & \rho_{vd}\sigma_{J,v}\sigma_{J,d} \\ \rho_{yd}\sigma_{J,y}\sigma_{J,d} & \rho_{vd}\sigma_{J,d}\sigma_{J,v} & \sigma_{J,d}^2 \end{pmatrix}$$

If a negative price jump is correlated with increased volatility, ρ_{yv} should be negative. The dependence between price jumps and the jumps in expected duration is measured by ρ_{yd}. If it's positive, it indicates that negative price jump leads to shorter durations between trades. Lastly, ρ_{vd} is the correlation of jumps sizes in volatility and expected duration. It is expected to have the same sign as the correlation between volatility and expected durations ρ.

3 Bayesian Inference

The model can be considered as a non-linear non-Gaussian state space model. Let Y, Θ and Z denote the observables, parameters and state variables respectively. The observables, parameters and state variables in my model are:

$$Y = \{y_i, D_i\}_{i=1}^N$$

$$\Theta = \{\Psi, \mu^x, \Sigma_x, \mu_J, \Sigma_J, \gamma, \sigma_m^2\}$$

$$Z = \{V_i, \lambda_i, \xi_i^y, \xi_i^x, J_i, m_i\}_{i=1}^N$$

Traditional likelihood based estimation requires evaluating the marginal likelihood $p(Y|\Theta)$. However, computation of $p(Y|\Theta)$ involves integrating out the latent time-varying variables Z, and this high dimensional integration is usually intractable. One solution is to employ a linear and Gaussian approximation and use Kalman Filter to obtain the likelihood. This method produces a Quasi Maximum Likelihood estimator. In a standard stochastic volatility model, the adequacy of the approximation depends on the variation of volatility (Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994) and Harvey and Shephard (1996)). The latent volatility and conditional durations in my model exhibit time varying variation. Also, in the presence of jumps and microstructure noise, logarithmic squared returns do not have a linear state space representation. Hence, I adopt a Bayesian MCMC algorithm consisting of Gibbs and Metropolis-Hastings sampler for the estimation.

Bayesian inference in a state space model focuses on the marginal posterior distribution $p(\Theta|Y)$ and $p(Z|Y)$. The key feature of Gibbs sampler is that if we draw G random samples $\{\Theta^{(g)}, Z^{(g)}\}_{g=1}^{G}$ from their conditional distributions $p(\Theta|Y, Z)$ and $p(Z|\Theta, Y)$ sequentially, then$\{\Theta^{(g)}\}_{g=1}^{G}$ and $\{Z^{(g)}\}_{g=1}^{G}$ converges to the marginal distributions of interest as $G \to \infty$. The conditional posterior $p(Z|Y, \Theta)$ updates the prior distribution $p(Z|\Theta)$ with information from the augmented likelihood $p(Y|\Theta, Z)$. If Θ or Z consists more than one elements and they can not be updated in one block, I divide them into blocks where conditionals are available. I also combine Metropolis-Hastings steps in the algorithm if conditional distributions cannot be sampled directly. With sufficiently large draws $\{\Theta^{(g)}\}_{g=1}^{G}$ and $\{Z^{(g)}\}_{g=1}^{G}$, a commonly used point estimate is simply the sample mean after discarding the first K draws for burning in, i.e., $\hat{\Theta} \approx \frac{1}{G-K}\sum_{g=K+1}^{G}\Theta^{(g)}$ and $\hat{Z}_i = \frac{1}{G-K}\sum_{g=K+1}^{G} Z_i^{(g)}$. For an overview of MCMC methods in finance, see Johannes and Polson (2002).

I outline the algorithm as follows:

1. Initialize $Z^{(0)}$ and $\Theta^{(0)}$.
2. Sample $\{V_i\}$ and $\{\lambda_i\}$.
3. Sample $\{\xi_i\}$ and $\{J_i\}$.

4. Sample $\{m_i\}$.

5. Sample Ψ, μ^x and Σ_x.

6. Sample μ_J and Σ_J.

7. Sample γ.

8. Sample σ_m^2.

The priors for parameters are explained at section 3.1. Section 3.2 describes how to update the state variables. Section 3.3 explains the posteriors of parameters. For a full description of the updating procedures, see the Appendix.

3.1 Priors

Bayesian analysis requires the formulation of prior distributions. First, I choose a prior for Ψ to ensure that the OU process stays in the stationary region. Other than imposing stationarity, the prior is diffuse. Specifically, I choose a truncated multivariate normal with large variance. Second, the prior for γ and Σ_J reflects our belief that jumps are large and infrequent. Jump intensity γ is restricted in the region where γD doesn't exceed one.

Third, I use returns sampled at the highest frequency to form an informative prior for σ_m^2. Efficient price has independent increments, but the observed tick-by-tick return has significant negative autocorrelation. Suppose observed price is composed of efficient price and uncorrelated noise m_i, observed return is composed of independent efficient returns and an $MA(1)$ noise. In other words, the autocorrelation of tick time return is induced by the microstructure noise and I can use the first-order autocovariance as a measure of σ_m^2. Under the assumption of uncorrelated noise, $\hat{\sigma}_m^2 = -\Sigma_{i=1}^{N-1} r_i r_{i+1}/N$. I specify an inverse Gamma prior for σ_m^2 such that the prior mean is equal to $\hat{\sigma}_m^2$. The posterior mean $E(\sigma_m^2|m)$ is a weighted average between prior mean and variance of m_i. I use returns sampled from every Lth transaction to estimate the model, and I choose the weight of prior, or how tight/informative the prior is, according to L. The larger L is, the more sparse we are sampling, the more informative prior is.

Last, I choose uninformative and conjugate prior for μ^x and μ_J since there is no prior belief about their value. There is no conjugate prior for Σ_x; I specify a diffuse inverse Wishart prior.

3.2 Posteriors of State variables

I update each state variable sequentially using their posterior distributions. First, I decompose the joint posterior of X into univariate conditionals and update each element one at a time. Specifically, I break $p(X|r, D, \Theta, Z_{-X})$ into $p(V_i|V_{-i}, \lambda, r_{i+1}, \Theta, Z_{-X})$ and $p(\lambda_i|\lambda_{-i}, V, D_{i+1}, \Theta, Z_{-X})$ for $i = 1, ..., N$, where V_{-i} denotes the vector of V expect V_i, and λ_{-i} denotes the vector of λ expect λ_i. Using Bayes rule, the posterior for V_i is given by

$$p(V_i|rest) \propto \frac{1}{V_i} p(X_{i+1}|X_i, D_{i+1}, \Theta, Z_{-X}) p(X_i|X_{i-1}, D_i, r_{i+1}, \Theta, Z_{-X}) p(r_{i+1}|V_i, \Theta, Z_{-X}).$$

The posterior is not a standard distribution, hence the adoption of independent Metropolis-Hastings algorithm. Let $\pi(V_i)$ denote $p(V_i|rest)$, the target density, I draw V_i from $q(V_i)$, the proposal density, then accept the draw with probability α where

$$\alpha = min \left\{ \frac{\pi(V_i^{(g+1)}) q(V_i^{(g)})}{\pi(V_i^{(g)}) q(V_i^{(g+1)})}, 1 \right\}.$$

To find a proposal density, notice that the target density has a lognormal kernel from the evolution equation and a inverse gamma kernel from the observation equation. Following Jacquier, Polson, and Rossi (1994), I choose an inverse gamma distribution to approximate the lognormal kernel and combine it with the other inverse gamma kernel. The updating of λ_i follows the same procedure.

The conditional posteriors of J_i is Bernoulli with $J_i = 1$ indicating a jump arrival. I update J following the algorithm in Eraker, Johannes, and Polson (2003). Jumps sizes ξ_i have a multivariate normal posterior conditional on J, X, m, Θ and Y.

Last, define $\check{r}_{i+1} = r_{i+1} - \left(\mu^y D_{i+1} + \xi^y_{i+1} J_{i+1}\right)$, the posterior of m_i can be simplified to $p(m_i|m_{-i},\, \check{r}, \sigma_m^2)$, and drawn directly.

3.3 Posteriors of Parameters

From a Bayesian perspective, models with latent variables have a hierarchical structure. In other words, the conditional distribution of parameters governing the evolution of latent variables only depend on the latent variables. For example, I draw Ψ, Σ_x and μ^x sequentially from $p(\Psi|\mu^x,\, \Sigma_x,\, Z)$, $p(\Sigma_x|\mu^x,\, \Psi,\, Z)$ and $p(\mu^x|\Psi,\, \Sigma_x,\, Z)$. The posterior of Ψ can not be sampled from directly, so I adopt an independent Metropolis-Hastings step with a proposal density derived from the Euler discretization of X. To ensure that the proposal density bounds the tails of the target density, Ψ is drawn from a multivariate t distribution rather than Normal. Σ_x has a posterior that is well approximated by an Inverse Wishart Distribution, and I use Metropolis-Hasting algorithm to update it. μ^x has conjugate multivariate normal distribution and it can be drawn directly.

The conditional distribution of μ_J and Σ_J only depend on ξ and J. Since ξ are normal, the posterior of μ_J and Σ_J can be derived from standard linear models. Conditional on J, the posterior of γ is independent from other state variables and parameters.

The posterior of γ is sampled using a Metropolis-Hastings step. The posterior of σ_m^2 does not depend on Y or state variables other than m, i.e., $p(\sigma_m^2|rest) = p(\sigma_m^2|m)$. It has an inverse gamma kernel.

4 Simulation Studies

I use simulation studies to demonstrate the reliability of my estimation procedure. The simulated sample size is 5000. The posterior mean and the standard deviation is reported in Table 1.

Table 1: Parameter Inference using Simulated data.

	True value	SVSDCJ
Ψ_{11}	0.009	0.0109 (0.0021)
Ψ_{21}	0.002	0.0003 (0.0017)
Ψ_{12}	0.003	0.0036 (0.0023)
Ψ_{22}	0.006	0.0068 (0.0014)
μ^v	-8.8	-8.7874 (0.0565)
μ^d	2.0	1.9134 (0.0521)
Σ_{11}	0.0036	0.0052 (0.0013)
Σ_{12}	-0.0006	-0.0006 (0.0007)
Σ_{22}	0.0025	0.0029 (0.0005)
$\mu_{J,r}$	-0.1	-0.0913 (0.0799)
$\mu_{J,v}$	0.7	0.7071 (0.3306)
$\mu_{J,d}$	-0.4	-0.3801 (0.2232)
Σ_{11}^J	0.04	0.0526 (0.0294)
Σ_{12}^J	-0.06	-0.0032 (0.0632)
Σ_{22}^J	0.36	0.1764 (0.1611)
Σ_{13}^J	0.006	-0.0402 (0.0417)
Σ_{23}^J	-0.09	-0.0528 (0.0874)
Σ_{33}^J	0.09	0.0895 (0.0851)
γ	0.0005	0.0004 (0.0002)
σ_m	0.01	0.0099 (0.0002)

Note: This table reports the posterior mean and the standard deviation of the posterior (in parentheses). I run 20000 iterations and discard the first 1000 draws for burning in.

Formal test of the consistency of the posterior simulator can be employed following the method proposed in Geweke (2004).

5 Empirical Results

5.1 Data

I apply my model to the milli-second time stamped IBM trade data in the US Equity Data provided by tickdata. The sample period is September 2011 (21 trading days). I follow the cleaning procedure proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) to filter out the potentially erroneous data. First, entries with correction indicator other than 0 are deleted. Second, I delete entries with abnormal sales condition. (See the TAQ manual for a complete reference on the correction indicator and sales condition). Third, observations from outside of the normal opening time are omitted. Fourth, I delete entries from the first five minutes after opening to eliminate the price changes due to information accumulated overnight. Last, I treat entries with the same time stamp as one observation and use the mean price.

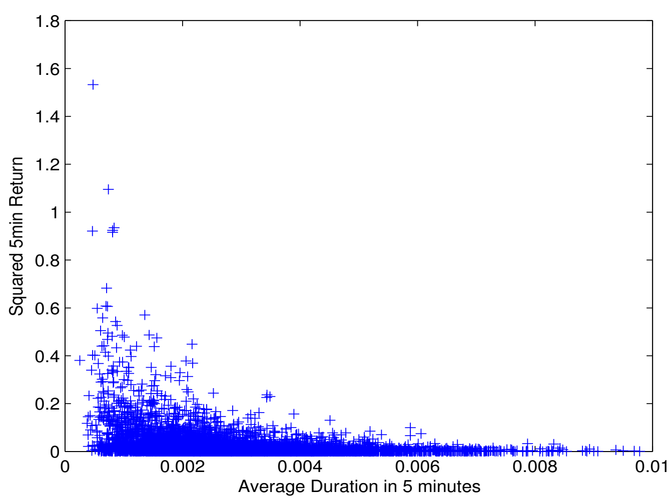

Figure 1: The dependence between duration and squared return.

Intraday returns can be constructed using different sampling schemes. First, I use a fixed five-minute sampling frequency to illustrate the motivation for modeling volatility and duration jointly. Figure 1 depicts the squared five-minute returns versus five-minute average duration, where the average durations are computed by counting the number of trades in the five-minute sampling period. The dependence between squared returns (proxy for volatility) and durations (proxy for conditional durations) is evident.

To preserve the information from durations and mitigate the effect of market microstructure noise, I sample from every Lth transaction rather than using tick-by-tick data. At ultra high frequency, unconditional returns display very high kurtosis. Under the assumption that returns are conditionally normal mixed with Poisson jumps, it is hard to produce such high kurtosis. Also, the discreteness of return is a predominant feature in tick-by-tick data since price changes has to be multiples of 1 cent (See Russell and Engle (2005)). The discreteness of durations induces measurement error as well. Moreover, this measurement error would affect shorter durations/smaller returns more than the longer durations/larger returns and hence biase the estimation. Another problem in tick time is that the microstructure noise is autocorrelated (See Hansen and Lunde (2006)). The time dependence of noise becomes negligible as sampling frequency decreases. Considering these factors, I choose L to be 100, leaving 6038 observations after. At this frequency, the mean duration is about 78 seconds. Although a large portion of data is tossed out, assumptions underlying my model are better met and this allows a more reliable estimation.

Intraday volatility and duration have well known diurnal patterns. Transactions happen more frequently near the opening time and closing time, and less frequently during the middle of a day. Before I apply the data to the stochastic model, this deterministic diurnal pattern needs to be filtered out. Durations are adjusted using

$D_i^a = D_i/gd_i$, where D_i^a is the adjusted duration, D_i is the original duration, and gd_i is the diurnal effect at time t_i. A nonparametric estimate of gd_i is obtained using a Normal kernel on the five-minute average durations. The level of the diurnal pattern has to be specified, otherwise the mean of conditional durations will be unidentified. I set gd_i at a level such that the mean of gd_i equals to one. Returns also need to be adjusted to account for the diurnal effect in duration and volatility. Diurnal volatility gv_i is obtained from using the Normal kernel on five-minute average squared returns. Adjusted return r_i^a is equal to $r_i/\sqrt{gd_i gv_{i-1}}$. The diurnal pattern gd_i and gv_i are plotted in Figure 2. Summary statistics for the adjusted returns and durations are given in Table 2.

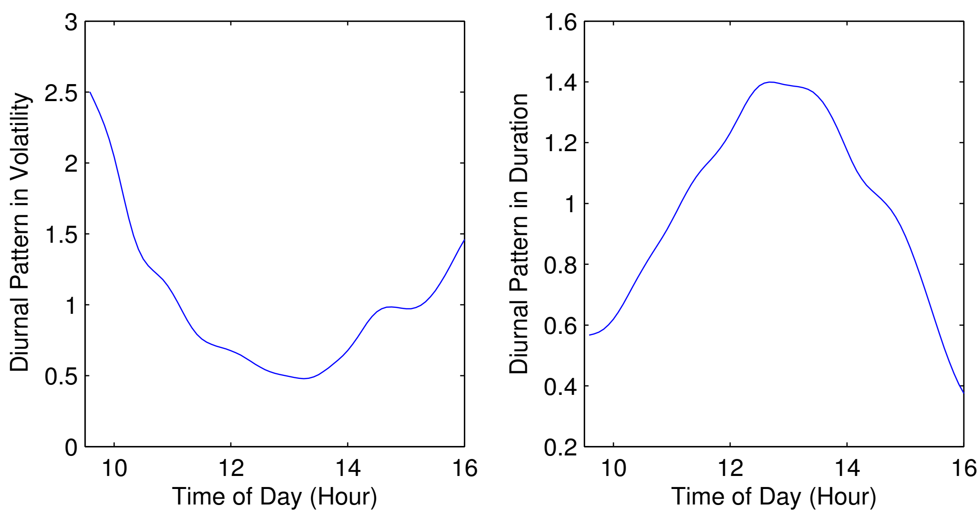

Figure 2: Nonparametric estimate of Diurnal Pattern.

Table 2: Summary statistics for adjusted IBM returns and durations.

	Mean	Std	Autocorrelation
r^a	0.00038	0.09421	0.01483
D^a	78.276	37.377	0.54025

5.2 Estimation

The posterior mean of the parameters for SVSDCJ model are given in Table 3. For comparison, I also estimate two nested models: stochastic volatility and stochastic duration with jumps in return (SVSDJ), and stochastic volatility and stochastic duration model (SVSD). Parameters for SVSDJ and SVSD are presented in the second and third column of Table 3.

Table 3: Parameter Inference

	SVSDCJ	SVSDJ	SVSD
Ψ_{11}	0.0009 (0.0002)	0.0004 (0.0001)	0.0012 (0.0002)
Ψ_{21}	0.0002 (0.0002)	0.0001 (0.0000)	0.0000 (0.0000)
Ψ_{12}	0.0014 (0.0007)	0.0005 (0.0002)	0.0011 (0.0003)
Ψ_{22}	0.0013 (0.0003)	0.0003 (0.0001)	0.0004 (0.0001)
μ^v	-9.3528 (0.0787)	-9.1931 (0.0885)	-9.1919 (0.0575)
μ^d	4.5818 (0.0579)	4.6221 (0.0625)	4.5343 (0.0830)
Σ_{11}	0.0006 (0.0001)	0.0004 (0.00005)	0.00142 (0.00025)
Σ_{12}	-2.29e-005 (7.17e-006)	1.66e-005 (1.02e-005)	-1.58e-005 (1.87e-005)
Σ_{22}	6.45e-006 (1.41e-006)	4.58e-006 (8.73e-007)	7.38e-006 (1.59e-006)
$\mu_{J,r}$	0.0406 (0.0399)	0.0086 (0.0142)	
$\mu_{J,v}$	1.0130 (0.2044)		
$\mu_{J,d}$	-0.0016 (0.0322)		
Σ_{11}	0.0273 (0.0087)	0.0154 (0.00332)	
Σ^J_{12}	-0.0105 (0.0222)		
Σ^J_{22}	0.2602 (0.1375)		
Σ^J_{13}	0.0019 (0.0064)		
Σ^J_{23}	0.0114 (0.0197)		
Σ^J_{33}	0.0283 (0.0104)		
γ	0.0001 (0.0000)	0.0005 (0.0002)	
σ_m	0.0093 (0.0003)	0.0092 (0.0003)	0.0092 (0.0003)

Note: This table reports the posterior mean and the standard deviation of the posterior (in parentheses). I run 6000 iterations and discard the first 1000 draws for burning in.

The posterior mean of Ψ_{11} in the SVSDJ model is the smallest among three models. Along with the smallest Σ_{11} estimate, it produces the most persistent spot volatility, as shown in Figure 3. The spot volatility estimates is obtained from the posterior mean of the draws of V_i. The volatility process in the SVSD model is expected to be less persistent and more volatile since all the variation in returns is attributed to volatility. This is confirmed by the larger Ψ_{11} and Σ_{11} estimates, and also the volatility estimates in Figure 3. The SVSDCJ model allows cojumps in returns and volatility, so the volatility process has less variation in the diffusive part (smaller Σ_{11}) and mean reverts more slowly (smaller Ψ_{11}) than what the SVSD model suggests. The estimated spot volatility for the SVSD model is presented in the bottom panel of Figure 3. The diffusive mean of the volatility process μ_v is smaller in the SVSDCJ model than the SVSDJ or SVSD model. This is consistent with the large positive jumps I find in the volatility process.

I use QQ-plot of the return residuals to assess the specification of the models. First, I plot the standardized unconditional returns in the top left panel in Figure 4. The distribution of unconditional returns is clearly not normal and it exhibits high kurtosis. The residuals in the SVSDCJ model are given by

$$\frac{r^a_{i+1} - \xi^y_{i+1} J_{i+1} - m_{i+1} + m_i}{\sqrt{V_i D^a_{i+1}}}$$

where the state variables are estimated from their posterior mean. If the model is correctly specified, the residuals should be approximately normally distributed. From the bottom right panel in Figure 4, SVSDCJ model shows no clear sign of misspecification. Bottom left panel in Figure 4 is the QQ-plot of residuals in the SVSDJ model. The residuals show signs of thin tails, indicating that the jumps could be overstated in this model. This also explains the smooth volatility path in Figure 3: since large movements of returns are attributed to jumps, variation of volatility is small. Residuals in the SVSD model have fat tails. Jumps or leptokurtic distributions are needed to capture the conditional nonnomalities in returns. These results are consistent with what Eraker, Johannes, and Polson (2003) find using daily level data.

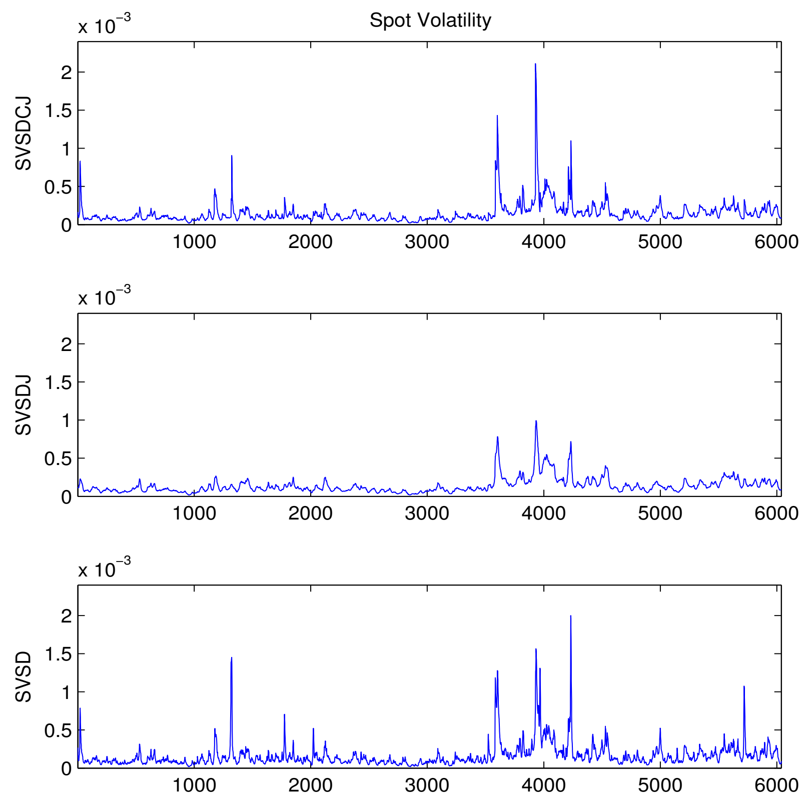

Figure 3: Estimated spot volatility. The estimate is obtained from the posterior mean (average of draws after burning in).

The estimated conditional duration λ_i is presented in Figure 5. Jumps in return does not have a direct impact on the duration process, hence the duration process in the SVSD model and SVSDJ model have similar dynamics. SVSDCJ model has a less persistent conditional duration path (larger Ψ_{22}). The duration residuals in all three models exhibit underdispersion, indicating the need for a more flexible duration

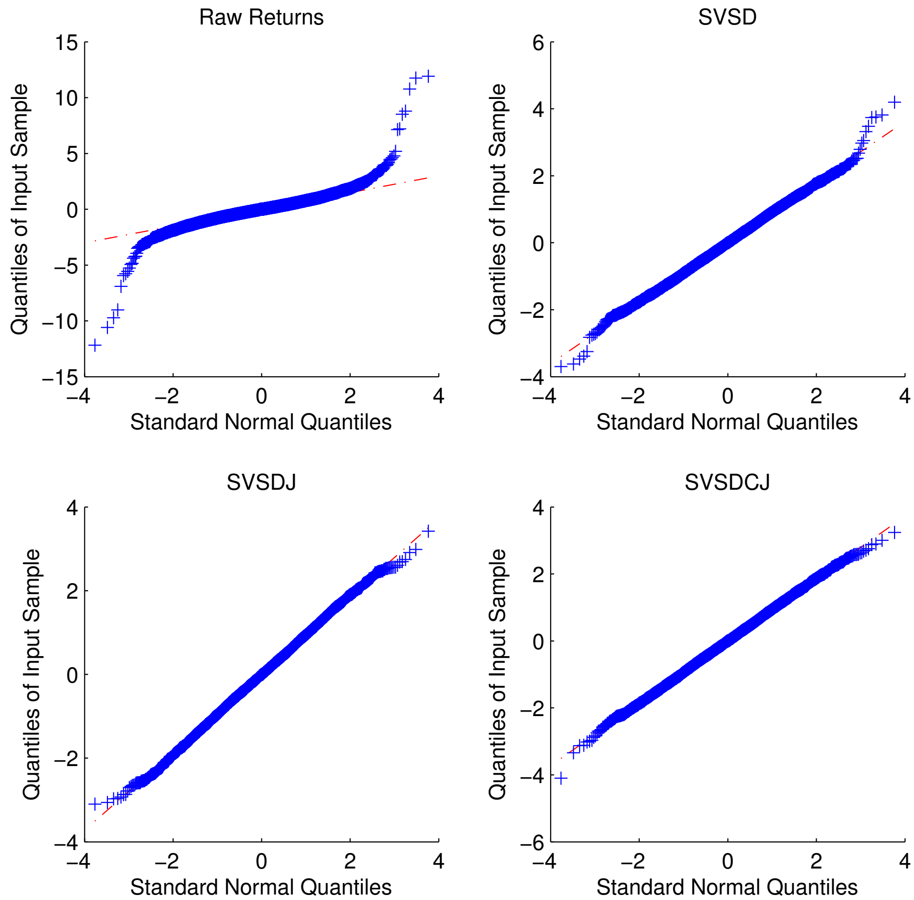

Figure 4: QQ plots of the return residuals.

distribution.

The off diagonal element of Ψ and Σ measures the interdependence between volatility and duration process. My findings are consistent with the Easley and O'Hara theory: both the presence and the lack of trade convey information about volatility. The positive posterior mean of Ψ_{12} and Ψ_{21} suggests that high volatility leads to short conditional duration, and short conditional duration leads to high volatility. Also, since they have the same sign, the system reverts to its diffusive mean following an exponential decay. Σ_{12} is negative, hence the contemporaneous correlation between the two

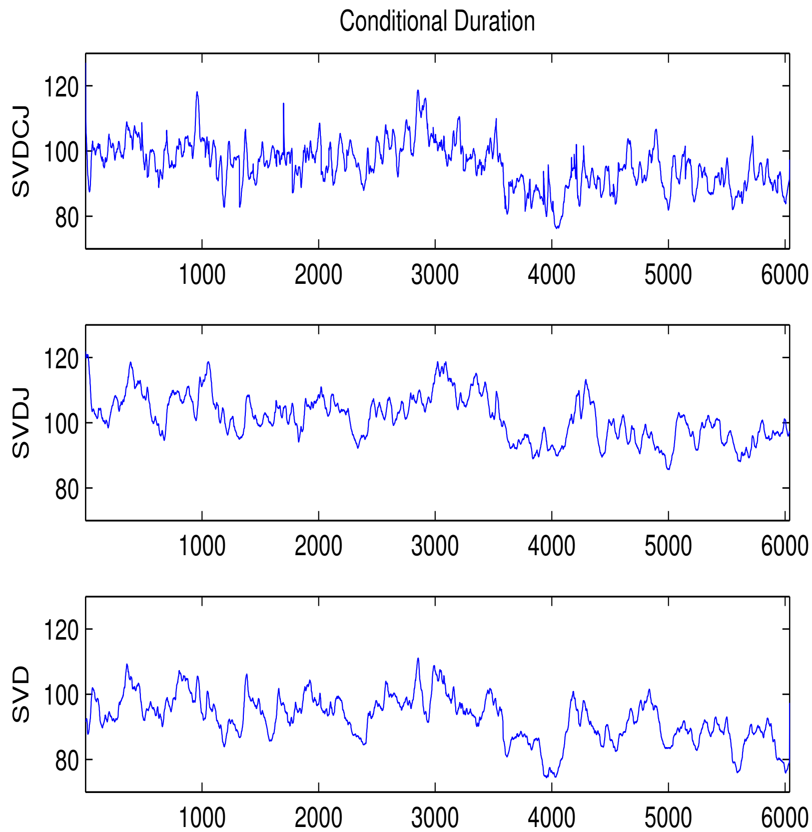

Figure 5: Estimated conditional duration. The estimate is obtained from the posterior mean (average of draws after burning in).

Brownian motion W_t^v and W_t^λ is negative. In other words, short conditional duration is accompanies by high volatility.

Figure 6 depicts the estimated jump sizes in returns from the SVSDCJ and SVSDJ model. More jumps are identified in the SVSDJ model than the SVSDCJ model. Given the evidence in the QQ-plot, some of the jumps in the SVSDJ model might be spurious. The estimated jump intensity γ is also higher in the SVSDJ model. Jumps in volatility reduce the need for jumps in returns as expected. The correlation between jump sizes in return and jump sizes in volatility provides source for leverage effect. The estimated correlation is negative as indicated by the sign of $\Sigma_J(2, 1)$, but it's not significantly different from zero. Other correlations between jump sizes are not significant either. Considering that jumps are rare and latent, I would need larger samples to estimate the correlation.

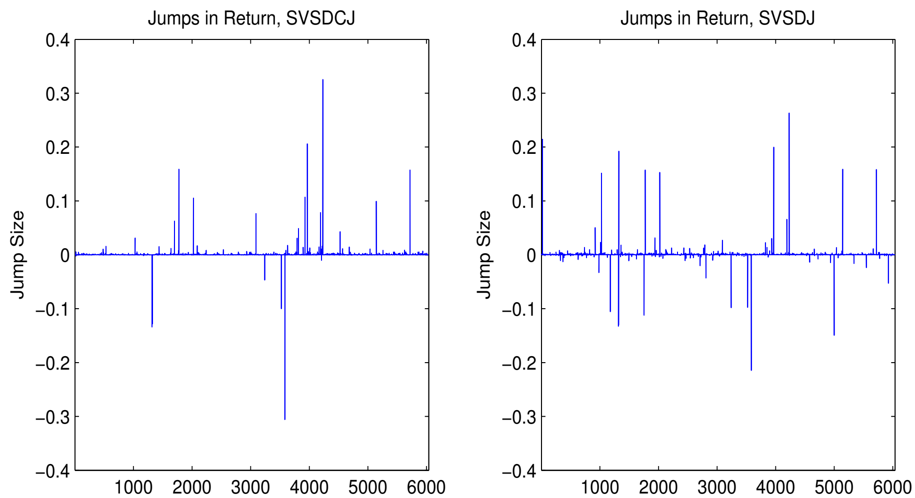

Figure 6: Estimated jumps.

I compare the estimated volatility and jumps to the popularly utilized bipower variation and jump variation. To get estimated integrated volatility in one day, I take the sum of the spot volatility multiplied by the duration, i.e., $\widehat{IV}_t = \sum_{i \in day(t)} V_i D_{i+1}$. Bipower variation is constructed using five-minute returns, $BV_t = \mu_1^{-2} \sum_{j=2}^{1/\Delta} |r_{t+j\Delta}||r_{r+(j-1)\Delta}|$, where $\Delta = 1/78$ and $\mu_1 = \sqrt{2/\pi}$. The estimated integrated volatility and bipower variation are plotted in Figure 7. They show similar pattern, with EIV lying slightly above in most days.

In the realized volatility literature, jump variation is measured by the difference between realized volatility and bipower variation. Realized volatility is computed using five-minute returns, $RV_t = \sum_{j=1}^{1/\Delta} r_{t+j\Delta}^2$. The difference can be negative with finite Δ, so the empirical measure of jump variation is truncated at zero, $JV_t = max\,(RV_t - BV_t, 0)$. When $\Delta \to 0$, JV_t converges to the quadratic variation due to

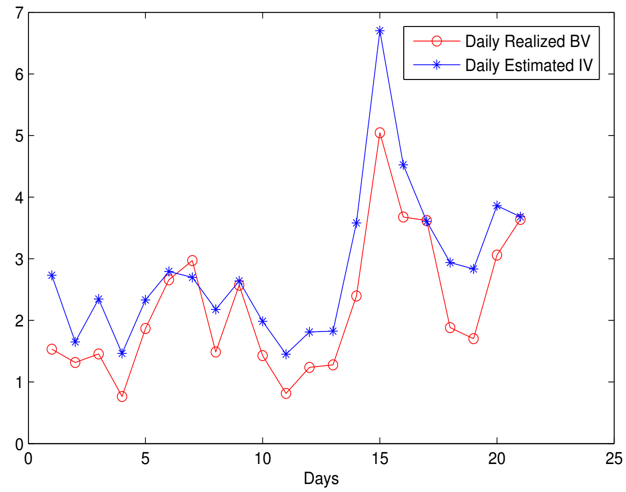

Figure 7: Estimated daily integrated volatility and realized bipower variation.

jumps, $JV_t \to \sum_{t<s<t+1} \xi^2(s)$. Define $\widehat{IJV}_t = \sum_{i \in day(t)} \xi_i^2 J_i$, if my model is correctly specified, $\widehat{IJV}_t$ and JV_t should converge to the same value. I plot the $\widehat{IJV}_t$ and JV_t in each day in Figure 8. JV_t is a lot larger than $\widehat{IJV}_t$ in most days. This is not surprising since I'm looking at returns sampled at a finer grid. As noted by Christensen, Oomen, and Podolskij (2011), jump variation based on coarser data tend to attribute a burst in volatility to jumps in return. Figure 9 depicts the logarithmic price in the day when the JV_t is the highest in the sample period. Top panel presents the logarithmic price every five minutes. There are severe price changes that are close to one percent. In a five-minute period, these are rare and might be considered as jumps. However, if we look at the bottom panel, where prices are plotted every 100th trade, there's no clear indication of large discrete price movement.

The proportion of variation due to jumps can be computed by

$$\frac{\sum_t \widehat{IJV}_t}{\sum_t \widehat{IV}_t + \sum_t \widehat{IJV}_t}$$

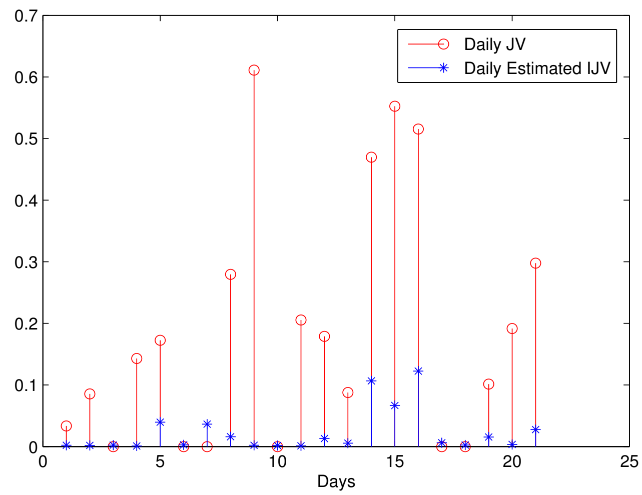

Figure 8: Estimated daily IJV and JV.

In the sample periods, 0.8% of the total variation is from jumps. The sample period is one month and it does not cover well known periods of market stress, so the estimated proportion doesn't serve as an indication of the magnitude of jump variation. However, using realized volatility, jump variation accounts for 7.8% of the total variation.

6 Conclusion

This paper puts forward a jump diffusion model SVSDCJ to jointly model the volatility and conditional duration process. Market microstructure theory suggests that durations between trades provide information to market participants, so volatility and durations are interdependent. My model analyses the interdependence and utilize this relationship to gain information about volatility. Given the nature of durations, observations are irregularly spaced. I develop MCMC algorithm for the inference of irregular spaced multivariate process. The algorithm provides smoothed estimates of the latent variables, such as spot volatility, jump times and jump sizes. Spot volatility can be easily

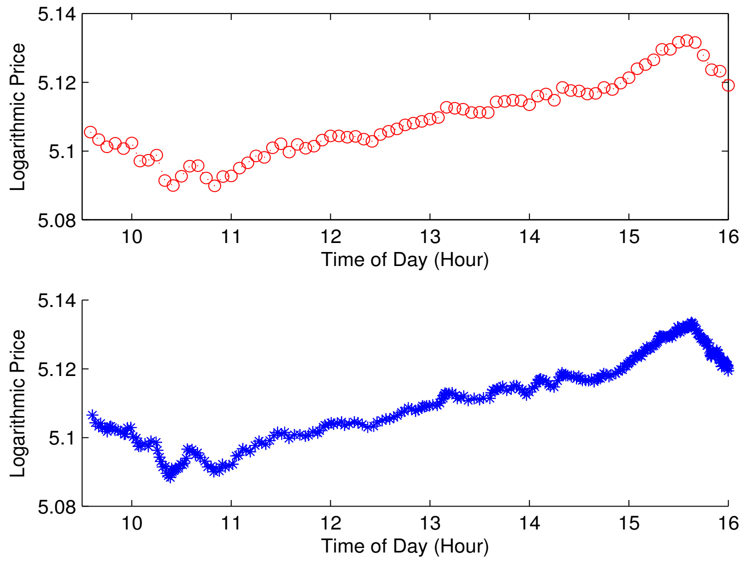

Figure 9: Logarithmic price on September 14th, the day with highest jump variation. Top panel is sampled at every 5 minutes. Bottom Panel is sampled at every 100th trade.

converted to integrated volatility in a given horizon. Knowing when jumps happen and how large the jumps are help us understand jump dynamics and price jump risk.

Applications to IBM data using my model and two nested alternatives reveal a few insights into the behavior of high frequency returns. First, jumps are important. Without jumps, stochastic volatility cannot fully capture the fat tails in the conditional distribution of returns. Second, cojump is a better specification than price jump. Jumps in volatility allow returns to change rapidly for a period of time. In addition, cojumps reduce the risk of overstating jumps. Third, total variation due to price jumps becomes smaller as I use finer returns. Last, volatility and conditional durations are interdependent, and it is consistent with what the Easley and O'Hara theory predicts.

Appendix: MCMC algorithm for SVSDCJ model

The full conditional posteriors in the MCMC algorithm are provided here:

1. Notations:

$$\Phi_{i+1} = e^{-\Psi D_{i+1}}$$

$$\mu_{i+1} = (I_2 - e^{-\Psi D_{i+1}})\mu^x + e^{-\Psi D_{i+1}} X_i$$

$$\check{r}_{i+1} = r^o_{i+1} - \xi^y_{i+1} J_{i+1}$$

$$\hat{r}_{i+1} = r^o_{i+1} - (m_{i+1} - m_i)$$

$$\hat{X}_{i+1} = X_{i+1} - (I - \Phi_{i+1})\mu^x - \Phi_{i+1} X_i,$$

$$\tilde{r}_{i+1} = r^e_{i+1} - \xi^y_{i+1} J_{i+1}$$

$$\tilde{X}_{i+1} = X_{i+1} - \xi^x_{i+1} J_{i+1}$$

2. Updating V and λ:

$$p(V_i|rest) \propto \frac{1}{V_i} exp\left(-\frac{1}{2}\frac{(logV_i - \mu_{vi})^2}{\sigma^2_{vi}}\right) \times V_i^{-\frac{1}{2}} exp\left(\frac{(\tilde{r}_{i+1} - \mu^y D_{i+1})^2}{2V_i D_{i+1}}\right)$$

$$\Sigma^*_i = \left(\Sigma_i^{-1} + \Phi'_{i+1}\Sigma^{-1}_{i+1}\Phi_{i+1}\right)^{-1}$$

$$\mu^*_i = \Sigma^*_i \left(\Sigma_i^{-1}\left(\mu_i + \xi^x_i J_i\right) + \Phi'_{i+1}\Sigma^{-1}_{i+1}\left(X_{i+1} - \left(I_2 - \Phi_{i+1}\right)\mu_x - \xi^x_{i+1} J_{i+1}\right)\right)$$

$$\mu_{vi} = \mu^*_{i,1} + \frac{\Sigma^*_i(1,2)}{\Sigma^*_i(2,2)}\left(log\lambda_i - \mu^*_{i,2}\right)$$

$$\sigma^2_{vi} = \Sigma^*_i(1,1) - \frac{\left(\Sigma^*_i(1,2)\right)^2}{\Sigma^*_i(2,2)}$$

Approximate the first half, a lognormal distribution, with an inverse gamma distribution, and combine it with the second half, we have the following Inverse

Gamma proposal distribution:

$$
\begin{aligned}
q(V_i) &\sim IG(c_1 + 0.5, c_2 + 0.5\frac{(\tilde{r}_{i+1} - \mu^y D_{i+1})^2}{D_{i+1}}) \\
c_1 &= \frac{1 - 2\exp(\sigma_{vi}^2)}{1 - \exp(\sigma_{vi}^2)} \\
c_2 &= (c_1 - 1)\exp(\mu_{vi} + 0.5\sigma_{vi}^2)
\end{aligned}
$$

Generate V_i from the proposal density, then accept it with probability

$$
\alpha = min\left\{\frac{f(V_i^{(g+1)})q(V_i^{(g)})}{f(V_i^{(g)})q(V_{i+1}^{(g+1)})}, 1\right\}.
$$

The updating of λ_i follows the same procedure:

$$
\begin{aligned}
p(\lambda_i|rest) &\propto \frac{1}{\lambda_i}exp\left(-\frac{(log\lambda_i - \mu_{\lambda i})^2}{2\sigma_{\lambda i}^2}\right)\frac{1}{\lambda_i}exp\left(-\frac{D_{i+1}}{\lambda_i}\right) \\
\mu_{\lambda i} &= \mu_{i,1}^* + \frac{\Sigma_i^*(1,2)}{\Sigma_i^*(1,1)}\left(logV_i - \mu_{i,1}^*\right) \\
\sigma_{\lambda i}^2 &= (\sigma_2^*)^2 - \frac{\left(\Sigma_i^*(1,2)\right)^2}{\Sigma_i^*(1,1)}
\end{aligned}
$$

Proposal density:

$$
\begin{aligned}
q(\lambda_i) &\sim IG(d_1 + 1, d_2 + D_{i+1}) \\
d_1 &= \frac{1 - 2exp(\sigma_{\lambda i}^2)}{1 - exp(\sigma_{\lambda i}^2)} \\
d_2 &= (d_1 - 1)exp(\mu_{\lambda i} + 0.5\sigma_{\lambda i}^2)
\end{aligned}
$$

generate λ_i from the proposal density, then accept it using Metropolis-Hastings principle.

3. Updating jump times J:

$$
\begin{aligned}
& p(J_{i+1}=1|rest) \\
\propto & exp\left(-\frac{1}{2}\left(\begin{array}{c} \hat{r}_{i+1}-\xi^y_{i+1} \\ \hat{X}_{i+1}-\xi^x_{i+1} \end{array}\right)^T \left(\begin{array}{cc} V_i D_{i+1} & 0 \\ 0 & \Sigma_{i+1} \end{array}\right)^{-1} \left(\begin{array}{c} \hat{r}_{i+1}-\xi^y_{i+1} \\ \hat{X}_{i+1}-\xi^x_{i+1} \end{array}\right)\right)\gamma D_{i+1}.
\end{aligned}
$$

Define odds ratio $or = \frac{p(J_{i+1}=1|rest)}{p(J_{i+1}=0|rest)}$, we have

$$
p(J_{i+1}=1|rest) = \frac{or}{or+1}.
$$

4. Updating jump sizes ξ:

$$
\begin{aligned}
p(\xi_{i+1}|J_{i+1}=0, rest) &\sim N(\mu_J, \Sigma_J) \\
p(\xi_{i+1}|J_{i+1}=1, rest) &\propto N(\mu^*_J, \Sigma^*_J)
\end{aligned}
$$

where

$$
\begin{aligned}
\mu^*_J &= \Sigma^*_J\left(\left(\begin{array}{c} \hat{r}_{i+1}V_i^{-1}D_{i+1}^{-1} \\ \Sigma_{i+1}^{-1}\hat{X}_{i+1} \end{array}\right) + \Sigma_J^{-1}\mu_J\right) \\
\Sigma^*_J &= \left(\left(\begin{array}{cc} V_i^{-1}D_{i+1}^{-1} & 0 \\ 0 & \Sigma_{i+1}^{-1} \end{array}\right) + \Sigma_J^{-1}\right)^{-1}.
\end{aligned}
$$

5. Updating microstructure noise m:

$$
\begin{aligned}
p(m_i|rest) &\sim N\left(\frac{S_m}{K_m}, \frac{1}{K_m}\right) \\
K_m &= \frac{1}{V_i D_{i+1}} + \frac{1}{V_{i-1} D_i} + \frac{1}{\sigma_m^2} \\
S_m &= \frac{m_{i+1} - \check{r}_{i+1}}{V_i D_{i+1}} + \frac{\check{r}_i - m_{i-1}}{V_{i-1} D_i}
\end{aligned}
$$

6. Next, the conditional posterior of Ψ with a diffusive matrix normal prior $MN(A_1, A_2, A_3)$ is given by

$$
\begin{aligned}
&p(\Psi|rest) \\
&\propto \prod_{i=1}^{N-1} \frac{1}{|\Sigma_{i+1}|^{0.5}} \exp\left(-\frac{1}{2}\left(\tilde{X}_{i+1} - \mu_{i+1}\right)' \Sigma_{i+1}^{-1} \left(\tilde{X}_{i+1} - \mu_{i+1}\right)\right) \\
&\times \exp\left(-\frac{1}{2} tr\left[A_2^{-1} \left(\Psi - A_1\right)' A_3^{-1} \left(\Psi - A_1\right)\right]\right).
\end{aligned}
$$

This is not a known distribution. We resort to the Euler discretization for proposal density:

$$
\begin{aligned}
\Psi &\sim Multivariate - t\left(\left(\left(B_2' B_2\right)^{-1} B_2' B_1\right)', \left(B_2' B_2\right)^{-1} \otimes \Sigma^x, 3\right) \\
B_1 &= \left(\frac{X_2 - X_1 - \xi_2^x J_2}{\sqrt{D_2}}, \ldots, \frac{X_N - X_{N-1} - \xi_N^x J_N}{\sqrt{D_N}}\right)' \\
B_2 &= \left(\left(\mu^x - X_1\right)\sqrt{D_2}, \ldots, \left(\mu^x - X_{N-1}\right)\sqrt{D_N}\right)'
\end{aligned}
$$

I draw Ψ from a multivariate t distribution rather than Normal so that the proposal density bounds the target density, then accept Ψ using Metropolis-Hastings principle.

7. The conditional posterior for Σ_x is given by

$$p(\Sigma_x|rest) \propto \prod_{i=1}^{N-1} \frac{1}{|\Sigma_{i+1}|^{0.5}} \exp\left(-\frac{1}{2}\left(\tilde{X}_{i+1}-\mu_{i+1}\right)' \Sigma_{i+1}^{-1}\left(\tilde{X}_{i+1}-\mu_{i+1}\right)\right).$$

Proposal density:

$$q(\Sigma_x|rest) \propto \left(\frac{1}{|\Sigma_x|}\right)^{\frac{N-1}{2}} \exp\left(-\frac{1}{2} trace\left(\Sigma_x^{-1} E\right)\right)$$

where $E = \sum_{i=1}^{N-1} E_{i+1}$, and

$$vec(E_{i+1}) = (I_2 - e^{-(\Psi \oplus \Psi)D_{i+1}})^{-1}(\Psi \oplus \Psi) vec\left(\left(\tilde{X}_{i+1}-\mu_{i+1}\right)\left(\tilde{X}_{i+1}-\mu_{i+1}\right)^T\right).$$

The proposal density has an Inverse Wishart kernel, $IW(E, N-4)$. Draw from this distribution and accept using Metropolis-Hastings principle.

8. The conditional posterior of μ_x with a multivariate normal prior $N(m_0, M_0)$ is given by

$$\begin{aligned}
&p(\mu^x|rest) \sim N(m_\mu, M_\mu)\\
&M_\mu = \left(\sum_{i=1}^{N-1}\left[(I_2-\Phi_{i+1})' \Sigma_{i+1}^{-1}(I_2-\Phi_{i+1})\right]\right)^{-1} + M_0^{-1}\\
&m_\mu = M_\mu \left\{\sum_{i=1}^{N-1}\left[(I_2-\Phi_{i+1})' \Sigma_{i+1}^{-1}\left(X_{i+1}-\Phi_{i+1}X_i-\xi_{i+1}^x J_{i+1}\right)\right] + M_0^{-1} m_0\right\}.
\end{aligned}$$

9. To update γ, we use a scaled beta prior $p(\gamma) \propto (\gamma\bar{D})^{a_\gamma-1}(1-\gamma\bar{D})^{b_\gamma}$, where $\bar{D}$ is

the mean duration. The posterior:

$$p(\gamma|J) \propto \prod_{i=2}^{N} (\gamma D_i)^{J_i} (1-\gamma D_i)^{1-J_i} (\gamma \bar{D})^{a_\gamma - 1} (1-\gamma \bar{D})^{b_\gamma - 1}.$$

This is not a known distribution. To get a proposal density, we assume that $D_i = \bar{D}$:

$$q(\gamma|J) \propto (\gamma \bar{D})^{\sum J_i + a_\gamma - 1} (1-\gamma \bar{D})^{N-1-\sum J_i + b_\gamma - 1}.$$

Generate $\hat{\gamma} \sim Beta(\sum_{i=2}^{N} J_i + a_\gamma, N-1-\sum_{i=2}^{N} J_i + b_\gamma)$, then $\gamma = \hat{\gamma}/\bar{D}$ has the target kernel. Accept γ using Metropolis-Hastings principle.

10. Posterior for μ_J with a multivariate normal prior $N(M_J, Z_J)$:

$$p(\mu_J|\xi, \Sigma_J) \sim N(M_J^*, Z_J^*)$$
$$M_J^* = Z_J^* \left(Z_J^{-1} M_J + N_j \Sigma_J^{-1} \bar{\xi} \right)$$
$$Z_J^* = (Z_J^{-1} + N_j \Sigma_J^{-1})^{-1}$$

Posterior for Σ_J with an inverse Wishart prior $IW(f_J, W_J)$:

$$p(\Sigma_J|\xi, \mu_J) \sim IW(f_J^*, W_J^*)$$
$$f_J^* = f_J + N_j$$
$$W_J^* = W_J + \sum_{i:J_i=1} (\xi_i - \mu_J)(\xi_i - \mu_J)^T$$

11. To update σ_m^2, we use a conjugate Inverse Gamma prior $IG(f_m, W_m)$. The pos-

terior is

$$p(\sigma_m^2|m) \sim IG\left(f_m + \frac{N}{2}, W_m + \sum_{i=1}^{N} m_i^2\right).$$

I use an informative prior with sparsely sampled data. The autocovariance of returns at the highest frequency provides the prior mean of σ_m^2, i.e., $E(\sigma_m^2) = \frac{W_m}{f_m-1}$. Set the prior to data ratio to a, then f_m and W_m is given by $f_m = \frac{aN}{2} + 1$ and $W_m = E(\sigma_m^2)(f_m - 1)$.

References

Ait-Sahalia, Y., P. A. Mykland, and L. Zhang (2005): "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18(2), 351–416.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility," *The Review of Economics and Statistics*, 89(4), 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61(1), 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

Bandi, F. M., and J. R. Russell (2006): "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79(3), 655–692.

Barndorff-Nielsen, O. E. (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2(1), 1–37.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2009): "Realized kernels in practice: trades and quotes," *Econometrics Journal*, 12(3), C1–C32.

Barndorff-Nielsen, O. E., and Shephard (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal Of The Royal Statistical Society Series B*, 64(2), 253–280.

Barndorff-Nielsen, O. E., and N. Shephard (2001): "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics," *Journal Of The Royal Statistical Society Series B*, 63(2), 167–241.

——— (2005): "Variation, jumps, market frictions and high frequency data in financial econometrics," Economics Papers 2005-W16, Economics Group, Nuffield College, University of Oxford.

Bauwens, L., and P. Giot (2000): "The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks," *Annales d'Economie et de Statistique*, (60), 06.

Bauwens, L., and D. Veredas (2004): "The stochastic conditional duration model: a latent variable model for the analysis of financial durations," *Journal of Econometrics*, 119(2), 381–412.

Christensen, K., R. Oomen, and M. Podolskij (2011): "Fact or friction: Jumps at ultra high frequency," Creates research papers, School of Economics and Management, University of Aarhus.

Dufour, A., and R. F. Engle (2000): "Time and the Price Impact of a Trade," *Journal of Finance*, 55(6), 2467–2498.

Easley, D., and M. O'Hara (1992): "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47(2), 576–605.

Engle, R. F. (2000): "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68(1), 1–22.

Engle, R. F., and J. R. Russell (1998): "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66(5), 1127–1162.

Eraker, B., M. Johannes, and N. Polson (2003): "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58(3), 1269–1300.

Gardiner, C. (2009): *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics. Springer.

Geweke, J. (2004): "Getting It Right: Joint Distribution Tests of Posterior Simulators," *Journal of the American Statistical Association*, 99, 799–804.

Ghysels, E., and J. Jasiak (1998): "GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model," *Studies in Nonlinear Dynamics & Econometrics*, 2(4), 4.

Grammig, J., and M. Wellner (2002): "Modeling the interdependence of volatility and inter-transaction duration processes," *Journal of Econometrics*, 106(2), 369–400.

Hansen, P. R., and A. Lunde (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business & Economic Statistics*, 24, 127–161.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61(2), 247–64.

Harvey, A. C., and N. Shephard (1996): "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business & Economic Statistics*, 14(4), 429–34.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12(4), 371–89.

Johannes, M., and N. Polson (2002): "MCMC Methods for Financial Econometrics," in *Handbook of Financial Econometrics*. North-Holland. Forthcoming.

Kloeden, P. E., and E. Platen (1992): *Numerical Solution of Stochastic Differential Equations*, Stochastic Modelling and Applied Probability. Springer.

Manganelli, S. (2005): "Duration, volume and volatility impact of trades," *Journal of Financial Markets*, 8(4), 377–399.

Merton, R. C. (1976): "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics*, 3, 125 – 144.

Oomen, R. C. (2006): "Properties of Realized Variance Under Alternative Sampling Schemes," *Journal of Business & Economic Statistics*, 24, 219–237.

Ross, S. (1996): *Stochastic processes*, Wiley series in probability and statistics: Probability and statistics. Wiley.

RUSSELL, J. R., AND F. M. BANDI (2004): "Microstructure noise, realized volatility, and optimal sampling," Econometric Society 2004 Latin American Meetings 220, Econometric Society.

RUSSELL, J. R., AND R. F. ENGLE (2005): "A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive Conditional Multinomial-Autoregressive Conditional Duration Model," *Journal of Business & Economic Statistics*, 23, 166–180.

TODOROV, V., AND G. TAUCHEN (2011): "Volatility Jumps," *Journal of Business & Economic Statistics*, 29(3), 356–371.

ZHANG, L., P. A. MYKLAND, AND Y. AIT-SAHALIA (2005): "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

ZHENG, H., AND D. PELLETIER (2012): "Joint Modeling of High-Frequency Price and Duration Data," Discussion paper, North Carolina State University.

ZHOU, B. (1996): "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14(1), 45–52.