A Jump-Diffusion Model with Stochastic Volatility and Durations *

Wei Wei†, Denis Pelletier‡

August 7, 2015

Abstract

Market microstructure theories suggest that the durations between transactions carry information about volatility. This paper puts forward a model featuring stochastic volatility, stochastic conditional duration, and jumps to analyze high frequency returns and durations. Durations affect price jumps in two ways: as exogenous sampling intervals, and through the interaction with volatility. We adopt a bivariate Ornstein-Ulenbeck process to model intraday volatility and conditional duration. We develop a MCMC algorithm for the inference on irregularly spaced multivariate processes with jumps. The algorithm provides smoothed estimates of the latent variables such as spot volatility, conditional duration, jump times, and jump sizes. We apply this model to IBM data and find that volatility and conditional duration are interdependent. We also find that jumps play an important role in return variation, but joint modeling of volatility and conditional duration reduces significantly the need for jumps.

JEL Codes: C1, C5, G1

Keywords: Durations, Stochastic Volatility, Price jumps, High-frequency data, Bayesian inference

*We thank Asger Lunde, Kim Christensen, Walter Thurman, Atsushi Inoue, Peter Bloomfield and Rasmus Varneskov for helpful comments. We also want to thank participants at Computational Methods for Jump Processes Workshop (2014), and seminars at NCSU and CREATES for valuable discussions. Wei Wei acknowledges financial support from the The Danish Council for Independent Research, Social Sciences (4003-00106B/FSE) and CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

†CREATES, Department of Economics and Business Economics, Aarhus University, 8210 Aarhus V, Denmark, email: wwei@econ.au.dk

‡Department of Economics, North Carolina State University, Raleigh, NC 27695, USA, email: denis_pelletier@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. These features have inspired GARCH and stochastic volatility models to capture the predictability of volatility with daily or lower frequency returns. However, fixed time aggregation loses potentially valuable information, such as the durations between transactions. Pelletier and Zheng (2013) propose modeling returns and durations jointly using a bivariate stochastic process for the latent volatility and trading intensity. In this paper, we extend their model to allow jumps in the price process, and develop bayesian inference for this irregularly spaced multivariate model. Our proposed stochastic volatility and stochastic duration with jumps (SVSDJ) model helps fill the gap between irregularly spaced high frequency data and traditional jump-diffusion models with stochastic volatility. With this model, we can estimate intraday volatility by exploiting the persistence of volatility as well as the information conveyed in durations. We also disentangle the jump component from the continuous part, and measure the effect of trading durations on jumps.

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.

We utilize the stochastic class of volatility and duration modeling. The logarithmic

price follows a jump-diffusion process with stochastic volatility and compound Poisson jump process. Prices are sampled at random time intervals; the durations of the sampling intervals are random. Moreover, the conditional mean of the durations, or the conditional duration, follows a stochastic process. The latent volatility and conditional duration reflect unobservable information flow. Following Pelletier and Zheng (2013), we model the logarithm of conditional duration and the logarithm of volatility by 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 conditional volatility/duration will be affected by their own past. Second, as predicted by microstructure theory, volatility and conditional duration interact with each other. The bivariate OU process allows the expected volatility and duration to depend on lagged values as well as contemporaneously correlated shocks.

The presence of jumps is another important feature of financial returns. Merton (1976) first describes returns using a continuous diffusion process and a compound Poisson jump process. Jumps are interpreted as "abnormal" variation in price due to the arrival of important news. They generate infrequent large movements and contribute to the fat tails in the return distribution. Without jumps, volatility needs to be counterfactually high to explain the occasional large fluctuations, see e.g. Bates (2000). Also, 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; see e.g. Todorov (2010) or Bollerslev and Todorov (2011).

Stochastic durations affect jump estimation through two channels. Firstly, if we treat durations as exogenous sampling intervals of the continuous-time jump-diffusion process, the length of the interval would directly impact jump identification. Secondly, in our bivariate setting for latent process, the evolution of conditional duration can affect the evolution of volatility, allowing richer dynamics in the latter, and hence reduce the need for price jumps. The importance of treating durations as endogenous has been

emphasized in a small but growing body of literature, for example Renault and Werker (2011) and Renault, van der Heijden, and Werker (2014), but we extend the intuition to price jumps.

We resort to Markov chain Monte Carlo methods (MCMC) for the estimation of our model, which can be viewed as a nonlinear and non-Gaussian state space model. The latent variables, namely volatility, conditional duration, and jumps, are treated as the states, and their dynamics are described by the evolution equation. The observation equation describes how returns and durations depend on the states. MCMC was first introduced to the stochastic volatility literature by Jacquier, Polson, and Rossi (1994) as a method of exact finite sample inference. Since then, influential work such as Kim, Shephard, and Chib (1998) and Jacquier, Polson, and Rossi (2004) has refined the method and applied it to more general settings. In particular, Chib, Nardari, and Shephard (2002) and Eraker, Johannes, and Polson (2003) develop MCMC methods to incorporate jumps. Our algorithm builds upon this literature. To improve efficiency, we also borrow from particle filters, in particular Johannes, Polson, and Stroud (2009) and Pitt and Shephard (1999).

One major benefit of using MCMC is that both parameters and state variables are estimated simultaneously instead of using an *ad hoc* filtering technique. The estimated conditional volatility and jumps are useful in applications such as computation of Value at Risk. Another benefit of MCMC is that we 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, we can use an appropriate prior to elicit such beliefs.

Our model is closely related to the large literature on the direct modeling of durations. Following the idea behind GARCH, Engle and Russell (1998) propose the autocorrelated conditional duration (ACD) model. 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, where the expected duration becomes stochastic; we 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 shares and examines the impact of durations on volatility. He imposes exogeneity on the duration process but allows volatility to be influenced by durations in 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 and analyze 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. Tay, Ting, Tse, and Warachka (2011) utilize an asymmetric ACD model and they find that durations are important determinants of price dynamics and volatility.

Non-parametric volatility estimation utilizing realized measure is another important area in high frequency econometrics. In its simplest form, realized volatility uses returns sampled at a short 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, 2003), and Barndorff-Nielsen and Shephard (2001, 2002). Since then, a huge literature has been devoted to the development and implementation of realized measure.

Realized volatility estimator has led to a non-parametric estimator for jump variation.

Barndorff-Nielsen and Shephard (2004) introduce the idea of realized bipower variation, which is the summation of the cross product of returns. Suppose that the price process has both a diffusion part and a jump part, the difference between realized bipower variation and realized volatility is a measure of the quadratic variation from the jump component. Using this tool, recent literature has suggested that jumps play an important role in the quadratic variation of price. For reviews on realized volatility and jumps, see Barndorff-Nielsen and Shephard (2005).

One of the most challenging complications in dealing with high frequency data is the existence of market microstructure noise (MMN). 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 the 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 and inconsistent estimator of the actual volatility. As the sampling frequency increases, the estimator diverges to infinity.

In the realized volatility literature, there are several approaches to deal with MMN. The simplest way is to sample sparsely, for example every 5 minutes. Bandi and Russell (2006) suggest using data sampled at different frequencies to separate noise from volatility, and determine the optimal sampling frequency. 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), Hansen and Lunde (2006) and Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) use the autocovariance of returns to construct kernel-based volatility estimators. Ait-Sahalia, Mykland, and Zhang (2005) show that if the the noise term is accounted for explicitly, sampling as often as possible is optimal.

Our approach in dealing with MMN combines several different methods. First, we

model noise terms explicitly. Noise is treated as a latent variable and it is estimated in the model. Second, we sample from every Lth transaction. This sampling scheme is referred to as transaction time sampling or tick time sampling 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 we adopt and it allows sampling at finer grid than current parametric models.

The rest of this paper is organized as follows. In Section 2 we describe the model specification. Section 3 discusses our proposed Bayesian estimation method for this model. Section 4 presents the empirical results using IBM shares data. Section 5 concludes.

2 Model Specification

2.1 Returns and Durations

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

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

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, we assume that V_t and W_t^y are independent. Jumps are modeled by a compound Poisson process since we are 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 normally distributed, $\xi_t^y \sim N(\mu_J, \sigma_J^2)$.

The price process is observed when there is a transaction. 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 the application, we sample 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. The continuous time framework allows straightforward discretization of the model over unequally spaced data. Using an Euler approximation, we discretize dy_t over the durations:

$$y_{i+1} - y_i = r^e_{i+1} = \sqrt{V_i D_{i+1}}\epsilon^y_{i+1} + \xi^y_{i+1} J_{i+1}, \tag{2}$$

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 jump 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 a jump indicator, and it follows a Bernoulli distribution.

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,

$$y^o_i = y_i + m_i. \tag{3}$$

We assume that microstructure noise is i.i.d. mean zero and normal, $m_i \sim N(0, \sigma^2_m)$, and that microstructure noise is independent of the efficient price[1]. The observed return, $r_{i+1} = y^o_{i+1} - y^o_i$, are contaminated by an MA(1) noise:

$$r_{i+1} = \sqrt{V_i D_{i+1}}\epsilon^y_{i+1} + \xi^y_{i+1} J_{i+1} + m_{i+1} - m_i. \tag{4}$$

Observed returns are sampled irregularly in time. Sampling intervals have a direct impact on the estimation of jumps. Since the variation from the continuous part of r_{i+1} is proportional to D_{i+1}, when D_{i+1} is large, jumps are harder to detect. On the other hand,

[1]These assumptions may be violated for microstructure noise in tick-by-tick price series. However, the problem lessens when we sample less frequently, say every 100 trades. See Hansen and Lunde (2006) for an analysis on the properties of microstructure noise.

as noted by Christensen, Oomen, and Podolskij (2014), coarsely sampled data tends to attribute a burst in volatility to jumps in return. Equation (4) allows us to disentangle jumps and volatility while taking durations into consideration. Furthermore, as mentioned before, the dynamics of the conditional duration could impact jump estimation through volatility. To analyze the interaction between stochastic durations and volatility, we first resort to the direct modeling of durations, which is another active research area in high frequency econometrics. Following the financial duration literature, D_{i+1} is modeled as the conditional duration, λ_i, multiplied by an i.i.d random variable with positive support, i.e.,

$$D_{i+1} = \lambda_i e_i. \tag{5}$$

We specify a Gamma distribution for e_i,

$$e_i \sim \Gamma(d_s, 1/d_s),$$

where d_s is the shape parameter. When $d_s = 1$, this distribution simplifies to the standard exponential distribution. Allowing d_s to differ from 1 gives us more flexibility to fit the conditional densities of durations, especially since the observed durations are aggregated over a number of transactions. The scale parameter is restricted to $1/d_s$ so that the mean of e_i equals to 1.[2] Under this specification, λ_i represents the conditional expectation of D_{i+1}, $E(D_{i+1}|\lambda_i) = \lambda_i$.

2.2 Volatility and Conditional Durations

To create persistence and interdependence between volatility and conditional durations, we follow Pelletier and Zheng (2013) and model the logarithm of λ_t and V_t using a

[2]Alternatively, we could restrict the scale parameter to be 1, but we find that the adopted parameterization has better mixing properties in the MCMC estimation procedure.

bivariate OU process. As noted by Andersen, Bollerslev, Diebold, and Ebens (2001), logarithmic volatility is closer to being normal than raw volatility is. Also, modeling logarithmic volatility and duration has the benefit of guaranteeing non-negativity without putting extra constraints on parameters. Let $X_t = (\log(V_t),\ \log(\lambda_t))'$, X_t solves:

$$dX_t = -\Psi(X_t - \mu^x)dt + S_x dW_t^x, \tag{6}$$

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 logarithmic volatility and logarithmic duration, and $S_x = \text{diag}(\sigma_v, \sigma_\lambda)$. W_t^x is a Brownian motion in $\mathbb{R}^2$ with $dW_t^v dW_t^\lambda = \rho dt$, where ρ is the instantaneous correlation. The instantaneous covariance matrix is given by

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

The exact solution to the SDE (6) is given by

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

where U_t has a bivariate normal distribution with mean 0 and $\text{vec}(\text{Var}(U_t)) = (\Psi \oplus \Psi)^{-1}(I_2 - \text{expm}(-(\Psi \oplus \Psi)t))\text{vec}(\Sigma_x)$. The symbol $\oplus$ denotes the kronecker sum. Note that $\text{expm}(-\Psi t)$ and $\text{expm}(-(\Psi \oplus \Psi)t)$ are matrix exponentials, which differs from element-wise exponentials if Ψ is not diagonal. The long-run variance of X_t is given by $\text{vec}(\text{Var}(X_t)) = (\Psi \oplus \Psi)^{-1} (\text{vec}(\Sigma_X))$.

To ensure the existence of a stationary solution, it is sufficient that Ψ has only eigenvalues with positive real parts so that $\text{expm}(-\Psi t) \to 0$ as $t \to 0$ (See Gardiner, 2009). We 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 encompasses 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.

The logarithmic volatility and conditional duration X_t are discretized over durations:

$$X_{i+1} = (I_2 - \text{expm}(-\Psi D_{i+1}))\mu^x + \text{expm}(-\Psi D_{i+1})X_i + U_{i+1}, \tag{7}$$

where

$$\begin{aligned} &U_{i+1} \sim N(0, \Sigma_{i+1}) \\ &\text{vec}(\Sigma_{i+1}) = (\Psi \oplus \Psi)^{-1}(I_2 - \text{expm}(-(\Psi \oplus \Psi)D_{i+1}))\text{vec}(\Sigma_x). \end{aligned}$$

Equations (4), (5), and (7) form our proposed SVSDJ model. We use an Euler discretization for X_t to gain some insight about the parameters:

$$X_{i+1} = \Psi\mu^x D_{i+1} + (I_2 - \Psi D_{i+1})X_i + \Sigma_x^{1/2}\sqrt{D_{i+1}}\epsilon_{i+1}^x, \tag{8}$$

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} is 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.

For the estimation procedure, we use the exact solution (7) rather than the Euler discretization (8) since the accuracy of the Euler approximation depends on D_{i+1}. We compare the full model SVSDJ to a restricted version, SVSD, in which there is no price jump. To analyze the effect of random sampling on the estimation of price jump, we also estimate the SV and SVJ model, in which returns are sampled using durations as in (4) (with no jump in model SV), but durations are treated as exogenous. In other words, $\log V_t$ is modeled by a univariate OU process with no interaction with conditional durations. The distributions of D_{i+1} is explored by estimating the stochastic conditional duration (SD) model versus the stochastic conditional duration with exponentially distributed durations (SDexp) model. In both models, $\log \lambda_i$ follows a univariate OU process. The difference lies in the distribution of D_{i+1} conditional on λ_i. In the SD model, $D_{i+1}|\lambda_i \sim \Gamma(d_s, \lambda_i/d_s)$, while in the SDexp model, d_s is restricted to 1, and $D_{i+1}|\lambda_i \sim \text{Exp}(\lambda_i)$.

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 our model are:

$$
\begin{aligned}
Y &= \{r_i, D_i\}_{i=1}^N \\
\Theta &= \{\Psi, \mu^x, \Sigma_x, \mu_J, \sigma_J, \gamma, \sigma_m^2, d_s\} \\
Z &= \{V_i, \lambda_i, \xi_i^y, J_i, m_i\}_{i=1}^N
\end{aligned}
$$

Likelihood-based estimation requires evaluating the marginal likelihood $p(Y|\Theta)$. However, computation of $p(Y|\Theta)$ involves integrating out the latent random variables Z, and this high dimensional integration is usually intractable. One solution is to employ a linear and Gaussian approximation and use the 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 variance of discretized volatility and conditional durations in our model, i.e., $\text{Var}(U_{i+1})$ in equation (7), depends on durations and hence it is time varying. Also, in the presence of jumps and microstructure noise, logarithmic squared returns do not have a linear state space representation. Hence, we adopt a Bayesian MCMC algorithm consisting of Gibbs and Metropolis-Hastings(MH) 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 the 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 $p(Y|\Theta,Z)$. 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 (2010).

If Θ or Z consists of more than one element and they cannot be updated in one block, we divide them into blocks where conditionals are available. If the conditional distribution cannot be sampled directly, we adopt the independent Metropolis-Hastings (IMH) sampler. Our algorithm first divides $\{Z,\Theta\}$ into four blocks: the jump block, $(\{\xi_i^y, J_i\}_{i=1}^{N}, \mu_J, \sigma_J, \gamma)$; the OU block, $(\{V_i, \lambda_i\}_{i=1}^{N}, \Psi, \mu^x, \Sigma_x)$; the MMN block, $(\{m_i\}_{i=1}^{N}, \sigma_m^2)$;

and the shape parameter, (d_s). To simplify notations, we define $\hat{r}_{i+1} = r_{i+1} - (m_{i+1} - m_i)$ for the jump block, and explain the key insights of the updating procedure in Section 3.1. Similarly, we define $\tilde{r}_{i+1} = r_{i+1} - \xi^y_{i+1} J_{i+1} - m_{i+1} + m_i$. Since jumps and market microstructure noise only affect returns, the distribution of the OU block conditional on r, ξ, J, m is the same as their distribution conditional on $\tilde{r}$. The sampler for the OU block will be detailed in Section 3.2. The MMN block and the shape parameter will be explained in Sections 3.3 and 3.4. Full description of the updating procedure can be found in the Appendix.

We outline our algorithm as follows:

1. Initialize $Z^{(0)}$ and $\Theta^{(0)}$.

2. Sample the jump block by

   (a) drawing from $\{J_{i+1}|\hat{r}_{i+1}, V_i, D_{i+1}, \Theta\}_{i=1}^N$ and $\{\xi_{i+1}|J_{i+1}, \hat{r}_{i+1}, V_i, D_{i+1}, \Theta\}_{i=1}^N$;

   (b) drawing σ_J from $\sigma_J|J_{1:N}, \xi_{1:N}$ and μ_J from $\mu_J|\sigma_J, J_{1:N}, \xi_{1:N}$;

   (c) drawing γ from $\gamma|J_{1:N}, \xi_{1:N}$.

3. Sample the OU block by

   (a) drawing from $\{V_i, \lambda_i|V_{-i}, \lambda_{-i}, \tilde{r}_{i+1}, D_{i+1}, \Theta\}_{i=1}^N$, where V_{-i} denotes the vector of V except V_i, and λ_{-i} denotes the vector of λ except λ_i.

   (b) drawing Σ_x from $\Sigma_x|V_{1:N}, \lambda_{1:N}, \Psi, \mu^x$;

   (c) drawing Ψ from $\Psi|V_{1:N}, \lambda_{1:N}, \mu^x, \Sigma_x$;

   (d) drawing μ^x from $\mu^x|V_{1:N}, \lambda_{1:N}, \Psi, \Sigma_x$.

4. Sample the MMN block by

   (a) drawing from $\{m_i|m_{-i}, V_{1:N}, \lambda_{1:N}, \xi_{1:N}, J_{1:N}, r_i, D_i, \Theta\}_{i=1}^N$.

   (b) drawing σ_m^2 from $\sigma_m^2|m_{1:N}$.

5. Sample the shape parameter d_s from $d_s|D_{1:N}, \lambda_{1:N}$.

3.1 The Jump Block

We update $\{J_{i+1}, \xi_{i+1}\}$ in one block by first drawing J_{i+1} marginalized over ξ_{i+1}, then drawing ξ_{i+1} conditional on J_{i+1} and the rest (observations, other state variables, and the parameters). This is possible because the density of $\hat{r}_{i+1}$ can be marginalized over ξ_{i+1}. Specifically, $p(\hat{r}_{i+1}|V_i, D_{i+1}, J_{i+1}) = N(J_{i+1}\mu_J, V_i D_{i+1} + J_{i+1}\sigma_J^2)$. The idea is similar to Johannes, Polson, and Stroud (2009), although they are drawing the state variables in the context of particle filters.

Conditional on $(J_{1:N}, \xi_{1:N})$, the jump size parameters (μ_J, σ_J^2) can be updated in one block as in the standard regression model. We use a conjugate prior such that the posterior for (μ_J, σ_J^2) is normal-inverse-Gamma. Also, we choose the prior for σ_J^2 to reflects our belief that jumps are large.

The conditional distribution of γ is not standard because the probability of $J_i = 1$ depends on D_i. We use a truncated scaled beta prior for γ, $p(\gamma) \propto (\gamma\bar{D})^{a_\gamma - 1}(1 - \gamma\bar{D})^{b_\gamma} 1\{\gamma < 1/\bar{D}\}$, where $\bar{D}$ is the mean duration. The truncation restricts γ to be in the region where $\gamma\bar{D}$ does not exceed one. [3] The posterior of γ,

$$p(\gamma|J) \propto \prod (\gamma D_i)^{J_i}(1 - \gamma D_i)^{1-J_i}(\gamma\bar{D})^{a_\gamma - 1}(1 - \gamma\bar{D})^{b_\gamma - 1} 1\{\gamma < 1/\bar{D}\}, \tag{9}$$

can not be sampled directly and we adopt the independent MH algorithm. The Independent MH sampler tends to have better mixing properties than the random-walk MH sampler if the proposal density is well-designed. We obtain the proposal density for γ by approximating at $D_i = \bar{D}$:

$$q(\gamma|J) \propto (\gamma\bar{D})^{\sum J_i + a_\gamma - 1}(1 - \gamma\bar{D})^{N - 1 - \sum J_i + b_\gamma - 1} 1\{\gamma < 1/\bar{D}\}. \tag{10}$$

[3] We use the mean rather than the maximum so the prior would have the same support as the proposal density for γ.

If we generate $\hat{\gamma} \sim Beta(\sum J_i + a_\gamma, N - 1 - \sum J_i + b_\gamma)$, $\gamma^{(g+1)} = \hat{\gamma}/\bar{D}$ has the above kernel. Then $\gamma^{(g+1)}$ is accepted with probability α:

$$\alpha = \min\left\{\frac{p(\gamma^{(g+1)}|J)q(\gamma^{(g)}|J)}{p(\gamma^{(g)}|J)q(\gamma^{(g+1)}|J)}, 1\right\}.$$

We find this proposal density to be a good approximation to the posterior as the acceptance probability is close to 1 in our application.

3.2 The OU Block

Updating the OU block is the most challenging part of the algorithm. The conditional densities of most elements are not standard distributions and need carefully designed proposal densities. We start with updating X. Let $\mu_{i+1} = (I_2 - \text{expm}(-\Psi D_{i+1}))\mu^x + \text{expm}(-\Psi D_{i+1})X_i$, we can write $X_{i+1}|X_i, D_{i+1}, \Theta \sim N(\mu_{i+1}, \Sigma_{i+1})$, and the conditional distribution of X_i is

$$p(X_i|\text{rest}) \propto p(X_{i+1}|X_i, D_{i+1}, \Theta)p(X_i|X_{i-1}, D_{i+1}, \tilde{r}_{i+1}, \Theta)p(\tilde{r}_{i+1}|X_i, D_{i+1}, \Theta)p(D_{i+1}|X_i). \tag{11}$$

When the transition density is normal, Pitt and Shephard (1999) suggest the auxiliary particle filter which uses Taylor expansion of the logarithmic measurement density as the basis for proposal density. We extend the idea to a bivariate setting to design proposal density for X_i. Note that in particle filters the target density is the filtered distribution of X_i, while in Gibbs sampler the target is the full conditional distribution. This does not impose any difficulty since the first two densities in Equation (11) can be combined to obtain a multivariate normal kernel π_1 as in Jacquier, Polson, and Rossi (1994)

$$\pi_1 = p(X_{i+1}|X_i, D_{i+1})p(X_i|X_{i-1}, D_i) \propto \exp\left(-\frac{1}{2}\left(X_i - \mu_i^*\right)'\left(\Sigma_i^*\right)^{-1}\left(X_i - \mu_i^*\right)\right), \tag{12}$$

where

$$\Phi_{i+1} = e^{-\Psi D_{i+1}}$$
$$\Sigma_i^* = \left(\Sigma_i^{-1} + \Phi_{i+1}^{'}\Sigma_{i+1}^{-1}\Phi_{i+1}\right)^{-1}$$
$$\mu_i^* = \Sigma_i^* \left(\Sigma_i^{-1}\left(\left(I_2 - \Phi_i\right)\mu_x + \Phi_i X_{i-1}\right) + \Phi_{i+1}^{'}\Sigma_{i+1}^{-1}\left(X_{i+1} - \left(I_2 - \Phi_{i+1}\right)\mu_x\right)\right).$$

Let π_2 and π_3 denote the measurement densities $p(\tilde{r}_{i+1}|X_i, D_{i+1}, \Theta)$ and $p(D_{i+1}|X_i)$. We approximate $\log \pi_2$ by

$$\log q_2 = \frac{1}{2}\left(\frac{r_{i+1}^2}{D_{i+1}} e^{-\mu_{i,1}^*} - 1\right) X_{i,1}, \tag{13}$$

which is obtained by the first order Taylor expansion at $\mu_{i,1}^*$. Likewise, $\log \pi_3$ is approximated by

$$\log q_3 = \left(D_{i+1} d_s e^{-\mu_{i,2}^*} - d_s\right) X_{i,2}. \tag{14}$$

Since $\pi_1 q_2 q_3$ still has a multivariate normal kernel, we use it as the proposal density in the independent Metropolis-Hastings sampler.

Alternatively, we could update V_i and λ_i sequentially and construct two inverse Gamma proposal densities. However, since V_i and λ_i could be correlated, it is more efficient to sample both in one block.

Given a flat prior, the posterior for Σ_x is proportional to $p(X|\Sigma_x, \Psi, \mu_x)$, i.e.,

$$p(\Sigma_x|rest) \propto \prod \frac{1}{|\Sigma_{i+1}|^{0.5}} \exp\left(-\frac{1}{2}\left(X_{i+1} - \mu_{i+1}\right)' \Sigma_{i+1}^{-1}\left(X_{i+1} - \mu_{i+1}\right)\right). \tag{15}$$

We propose the following inverse Wishart density as an approximation:

$$q(\Sigma_x|rest) \propto \left(\frac{1}{|\Sigma_x|}\right)^{\frac{N-1}{2}} \exp\left(-\frac{1}{2} trace\left(\Sigma_x^{-1} E\right)\right), \tag{16}$$

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

$$vec(E_{i+1}) = (I_2 - \text{expm}(-(\Psi \oplus \Psi)D_{i+1}))^{-1}(\Psi \oplus \Psi)vec\left((X_{i+1} - \mu_{i+1})(X_{i+1} - \mu_{i+1})^T\right).$$

We show in the Appendix that if Σ_{i+1} can be decomposed into $A_1\Sigma_x A_2$, while A_1 and A_2 are both invertible matrices, the proposal density is proportional to the target density, and the IMH sampler becomes the Gibbs sampler, i.e., the acceptance rate is equal to 1. One sufficient condition for this decomposition is that Ψ is a diagonal matrix. In our application, we find the acceptance rate to be close to 1.

Next, the posterior of Ψ is proportional to $p(X|\Sigma_x, \Psi, \mu_x)p(\Psi)$, and it is not a known distribution. We resort to the Euler discretization for proposal density. Specifically, we use Equation (8) as a pseudo model, and rewrite it to

$$\frac{X_{i+1} - X_i}{\sqrt{D_{i+1}}} = \Psi\mu^x\sqrt{D_{i+1}} - \Psi X_i\sqrt{D_{i+1}} + w_{i+1}, \tag{17}$$

then the posterior of Ψ follows a matrix normal distribution as in a vector regression model. This posterior from the pseudo model is served as the proposal density for the true model. To ensure tail-boundness, we use a multivariate-t distribution rather than Normal. The prior for Ψ is chosen to ensure that the OU process stays in the stationary region. The stationarity is guaranteed by imposing $4\Psi_{11}\Psi_{22} > (\Psi_{12} + \Psi_{21})^2$, and we discard draws that do not satisfy this criteria. Other than imposing stationarity, the prior is diffuse. Specifically, we choose Beta$(1.05, 0.1)$ for both Ψ_{11} and Ψ_{22}, and a flat prior for Ψ_{12} and Ψ_{21}.

The posterior for μ^x given a flat prior is multivariate normal and it can be updated using Gibbs sampler.

3.3 The MMN Block

Define $\check{r}_{i+1} = r_{i+1} - \xi^y_{i+1} J_{i+1}$, , we have $\check{r}_{i+1}|\text{rest} \sim N(m_{i+1} - m_i, V_i D_{i+1})$. If the prior for m_i is normal, $m_i \sim N(0, \sigma^2_m)$, the posterior of m_i has a normal kernel,

$$p(m_i|\text{rest}) \propto p(\check{r}_{i+1}|m_i, m_{i+1}, V_i, D_{i+1}, \sigma^2_m) p(\check{r}_i|m_i, m_{i-1}, V_{i-1}, D_i, \sigma^2_m) p(m_i), \tag{18}$$

The posterior of σ^2_m follows standard regression analysis. We use an informative prior for σ^2_m formed using returns sampled at the highest frequency. Efficient price has independent increments, but the observed tick-by-tick return displays significant negative autocovariance. The autocovariance can be used to obtain an estimate for σ^2_m, as in Hansen and Lunde (2006) and Zhou (1996). Suppose observed price is composed of efficient price and uncorrelated noise m_i. Then 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 we can use the first-order autocovariance as a measure of σ^2_m. Under the assumption of uncorrelated noise, $\hat{\sigma}^2_m = -\Sigma^{N-1}_{i=1} r_i r_{i+1}/N$. We specify an inverse Gamma prior for σ^2_m such that the prior mean is equal to $\hat{\sigma}^2_m$. The posterior mean $E(\sigma^2_m|m)$ is a weighted average between prior mean and variance of m_i. We use returns sampled from every Lth transaction to estimate the model, and we 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 the prior is.

3.4 The Shape Parameter

Let $D^s_{i+1} = D_{i+1}/\lambda_i$, the posterior of d_s is given by

$$p(d_s|D, \lambda) \propto \left(\frac{d_s^{d_s}}{\Gamma(d_s)}\right)^{N-1} \left(\prod_{i=1}^{N-1} D^s_{i+1}\right)^{d_s - 1} exp(-\sum_{i=1}^{N-1} D^s_{i+1} d_s) p(d_s),$$

where $p(d_s) \propto e^{-d_s/k_{ds}}$ is an exponential prior for d_s. Following Son and Oh (2006), we use a normal proposal density $q(d_s) = N(d_s^*, z_{ds})$, where d_s^* is the mode of the logarithmic pdf. Let $\overline{\log D^s}$ denote $\sum_{i=1}^{N-1} \log D_{i+1}^s/(N-1)$, d_s^* is the numerical solution to the following FOC with respect to d_s:

$$(\log d_s + 1) - \psi(d_s) + \overline{\log D^s} - \overline{D^s} - \frac{1}{k_{ds}(N-1)} = 0,$$

where ψ is the digamma function. The variance z_{ds} is chosen to be the twice the value of the negative inverse of the second derivative of the logarithmic pdf, i.e.

$$z_{ds} = 2\left((N-1)\left(-\frac{1}{d_s} + \psi^{(1)}(d_s)\right)|_{d_s=d_s^*}\right)^{-1}.$$

4 Empirical Results

4.1 Data

We apply our model to the milli-second time stamped IBM trade data. The sample period is September 2011 (21 trading days). We follow the cleaning procedure proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) to filter out potentially erroneous data. First, entries with correction indicators other than 0 are deleted. Second, we delete entries with an 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, we delete entries from the first five minutes after opening to eliminate the price changes due to information accumulated overnight. Last, we treat entries with the same time stamp as one observation and use the mean price.

Intraday returns can be constructed using different sampling schemes. First, we use a fixed five-minute sampling frequency to illustrate the motivation for modeling volatility

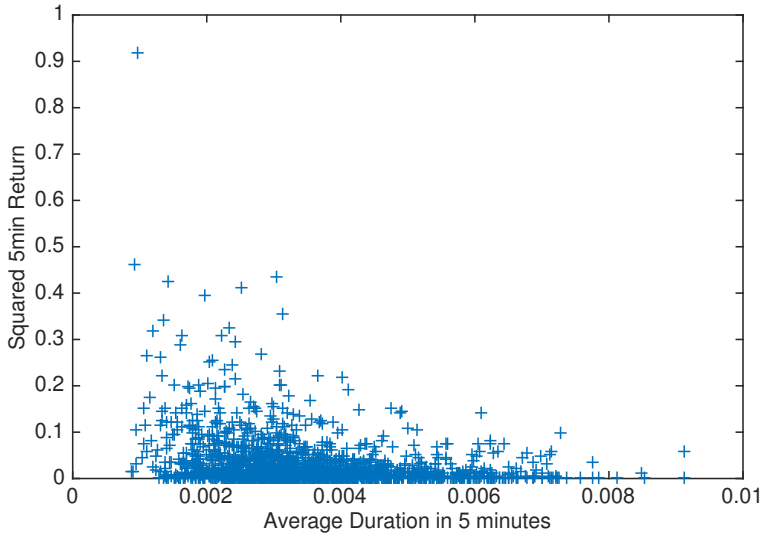

Figure 1: The dependence between durations and squared returns.

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, we 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 dominant feature in tick-by-tick data since price changes have 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 biases 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, we choose L to be 100, leaving 6038 observations. At this frequency, the mean duration is about 80 seconds. Although a large portion of data is tossed

out, assumptions underlying our model are better met and this allows for 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 we apply the data to the stochastic model, this deterministic diurnal pattern needs to be filtered out. Durations are adjusted using $D_i = D_i^u / g_d(t_i)$, where D_i^u is the unadjusted original duration, and $g_d(t_i)$ is the diurnal effect at time t_i. A nonparametric estimate of g_d is obtained by using a Normal kernel on the five-minute durations averaged over the entire year of 2011. The level of the diurnal pattern has to be specified, otherwise the mean of conditional durations will be unidentified. We set $g_d(t_i)$ at a level such that the average of $g_d(t_i)$ over the day equals to one. Diurnal volatility $g_v(t_i)$ is obtained by using the Normal kernel on five-minute average squared returns, and the level of $g_v(t_i)$ is set such that its mean equals to one. Returns need to be adjusted to account for the diurnal effect in both durations and volatility. As the unadjusted volatility V_i^u is equal to $V_i g_v(t_i)$, return is adjusted by $r_{i+1} = r_{i+1}^u / \sqrt{g_d(t_i) g_v(t_{i+1})}$. The diurnal pattern g_d and g_v are plotted in Figure 2. Summary statistics for the returns and durations before and after the adjustments are given in Table 1.

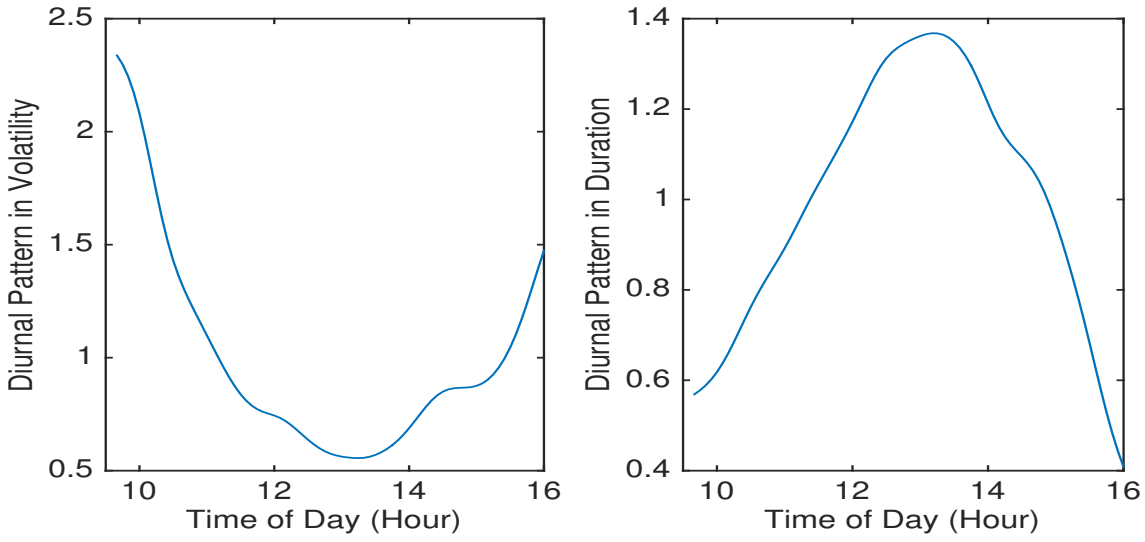

Figure 2: Nonparametric estimate of the Diurnal patterns.

Table 1: Summary Statistics for IBM returns and durations.

	r^u	D^u	r	D
Mean	0.0003	80.20	0.0003	77.99
Std	0.0896	50.76	0.0947	38.39
Autocorrelation	0.0162	0.70	0.0139	0.56

4.2 Estimation Results

We apply the Bayesian estimation procedure for the SVSDJ model and its nested alternatives to the deseasonalized returns and durations. Table 2 reports the posterior mean and standard deviation of the parameters. Models SD and SDexp estimate durations with gamma (SD) or exponential (SDexp) conditional densities, respectively. The logarithmic conditional duration is stochastic and follows a univariate OU process. SV and SVJ estimates returns with (SVJ) or without (SV) jumps, assuming that the logarithmic volatility follows an OU process. SVSD and SVSDJ model assumes that the logarithmic volatility and conditional duration are interdependent and they are modeled by a bivariate OU process.

Our MCMC algorithm also produces smoothed state variable estimates for all the models. We analyze the results in more detail in the following subsections.

4.3 Estimated Conditional Duration

The estimated conditional duration λ_i is presented in Figure 3. We omit the duration estimates in the SVSD model as it is very similar to the SVSDJ model. From Table 2, the SD model estimates for the shape parameter d_s is about 7.1 with standard deviation being around 0.2, which is significantly different from 1 in the SDexp model. Hence, we

Table 2: Parameter Estimates

	SDexp	SD	SV	SVJ	SVSD	SVSDJ
Ψ_{11}			0.0010	0.0002	0.0014	0.0010
			(0.0001)	(5.9e-05)	(0.0002)	(0.0002)
Ψ_{21}					−0.0005	−0.0004
					(7.5e-05)	(7.2e-05)
Ψ_{12}					−0.0016	−0.0016
					(0.0003)	(0.0002)
Ψ_{22}	0.0001	0.0013			0.0010	0.0011
	(3.8e-05)	(0.0001)			(0.0001)	(0.0001)
Σ_{11}			0.0015	0.0002	0.0028	0.0019
			(0.0002)	(6.2e-05)	(0.0003)	(0.0003)
Σ_{12}					−0.0011	−0.0009
					(0.0001)	(1.0e-04)
Σ_{22}	$1.6e{-}05$	0.0004			0.0005	0.0005
	(3.9e-06)	(3.3e-05)			(4.2e-05)	(4.3e-05)
μ_v			−9.1739	−9.2847	−9.3001	−9.3065
			(0.0592)	(0.1144)	(0.0543)	(0.0647)
μ_d	4.3693	4.4152			4.4144	4.4148
	(0.0495)	(0.0223)			(0.0263)	(0.0255)
d_s	1.0000	7.1370			7.5883	7.5116
		(0.1715)			(0.1869)	(0.1873)
σ_m			0.0062	0.0061	0.0061	0.0061
			(0.0002)	(0.0002)	(0.0002)	(0.0002)
γ_r				0.0017		0.0002
				(0.0005)		(0.0002)
μ_J				0.0031		−0.0139
				(0.0075)		(0.0451)
σ_J				0.0976		0.1414
				(0.0085)		(0.0290)

Note: This table reports the posterior mean and the standard deviation of the posterior (in parentheses). We run 60000 iterations and discard the first 10000 draws for burning in.

only adopt the SD specification when modelling jointly with volatility.

The effect of relaxing the shape parameter in the conditional densities of durations is evident in Figure 3. The SDexp model has a more persistent conditional duration path, which is also consistent with the smaller Ψ_{22} and smaller Σ_{22} estimates. Modeling volatility and conditional duration jointly had less impact on duration parameters or the conditional duration path.

We compare the standardized durations in Figure 4. The top row is for durations standardized by their unconditional mean. The middle and bottom rows are for dura-

tions standardized by the λ_i estimated with the model SDexp and SVSDJ, respectively. In the left column we have histograms and in the right column we have the sample autocorrelation function. For the histograms, we superimpose the pdf of an exponential distribution for the first two rows (durations standardized by the mean or SDexp's λ_i). We can see that in these two cases, the exponential distribution is not a good approximation. Looking at the autocorrelation function, we can see that although the conditional expected duration λ_i obtained from the SDexp model is able to capture the more persistent variations in durations, there are still some short-run dependence left. For the SVSDJ model, we plot the pdf of $\Gamma(d_s, 1/d_s)$ in addition to the empirical density. Both the empirical density and the autocorrelation structure shows clear improvement over the SDexp model.

4.4 Estimated Volatility and Jumps

In this subsection we analyze the dynamics of volatility and jumps. Figure 5 plots the spot volatility estimates for model SV, SVJ, SVSD and SVSDJ. These are obtained from the posterior mean of V_i. Comparing SV and SVJ, we see that the volatility path in model SV is more volatile, which is also confirmed by the larger estimates of Ψ_{11} and Σ_{11}. This is as expected since all the variation in returns is attributed to volatility in the SV model. The SVJ model produces the most persistent spot volatility among the four models, along with the smallest Σ_{11} estimate. Comparing SVSD and SVSDJ, we observe the same pattern: allowing jumps in return results in more persistent and less volatile volatility path, although the difference is less prominent.

Figure 5 shows that modeling conditional duration and volatility jointly leads to more variations in the volatility path. The off-diagonal elements of Ψ and Σ measure the interdependence between volatility and duration process. From Table 2, we find that both the presence and the lack of trade convey information about volatility. The negative posterior means of Ψ_{12} and Ψ_{21} suggests that low volatility leads to short conditional

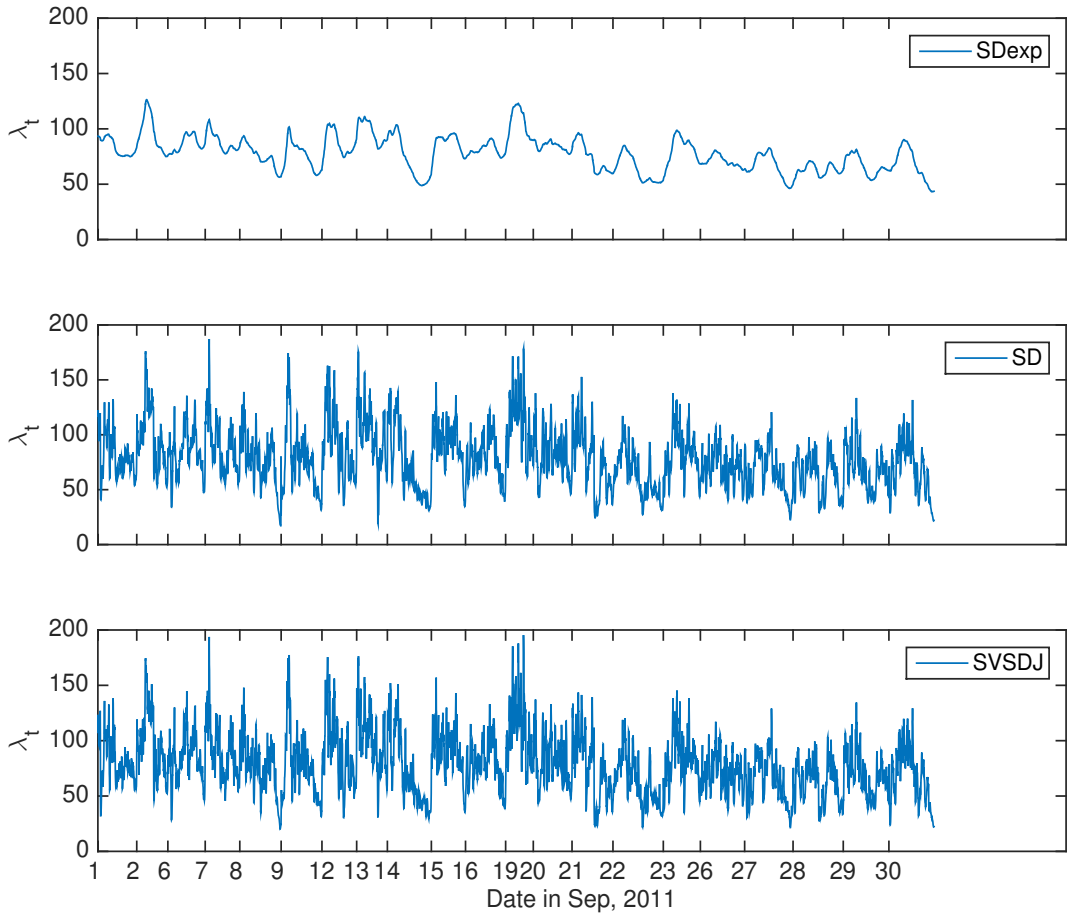

Figure 3: Estimated conditional duration. The estimate is obtained from the posterior mean (average of draws after burning in).

duration, and high conditional duration leads to high volatility. Also, since they have the same sign, the system reverts to its diffusive mean with an exponential decay. Σ_{12} is negative, hence the contemporaneous correlation between the two Brownian motion W_t^v and W_t^λ is negative. In other words, short conditional duration is accompanied by high volatility. The relationship between V_i and λ_i is opposite in the feedback effect and the contemporaneous correlation. Although this is worth further investigation, we leave it to future research.

Since conditional duration adds more variability to volatility, it also has a visible effect on jump estimation. Figure 6 depicts the estimated jump sizes in returns from the SVJ and SVSDJ model. More jumps are identified in the SVJ model than the SVSDJ

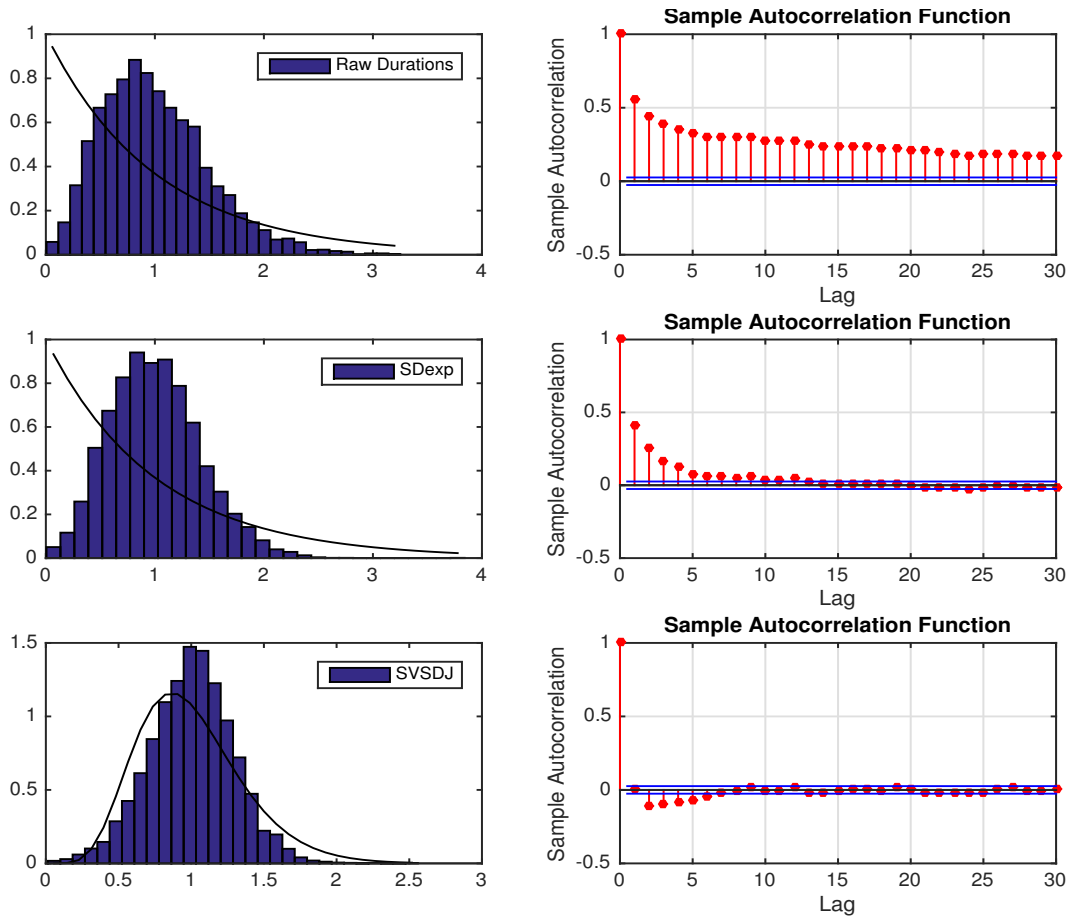

Figure 4: Standardized Durations. The left panels are the histogram of durations standardized by the conditional durations. The black line plots the theoretical conditional distribution. The right panels are the autocorrelations of the standardized durations.

model. The estimated jump intensity γ is also higher in the SVJ model. When we allow conditional durations to affect volatility, there is a significantly lesser need for price jumps.

We compare the estimated volatility and jumps to the popularly utilized bipower variation and jump variation. To estimate integrated volatility in one day, we take the sum of the spot volatility multiplied by the duration, and adjust it back using the diurnal functions, i.e., $\widehat{IV}_t = \sum_{i \in day(t)} V_i g_v(t_i) D_{i+1} g_d(t_{i+1})$. 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, $\widehat{IV}_t$ from the four parametric models and bipower variation are plotted in Figure 7. They show similar patterns, with $\widehat{IV}_t$ lying

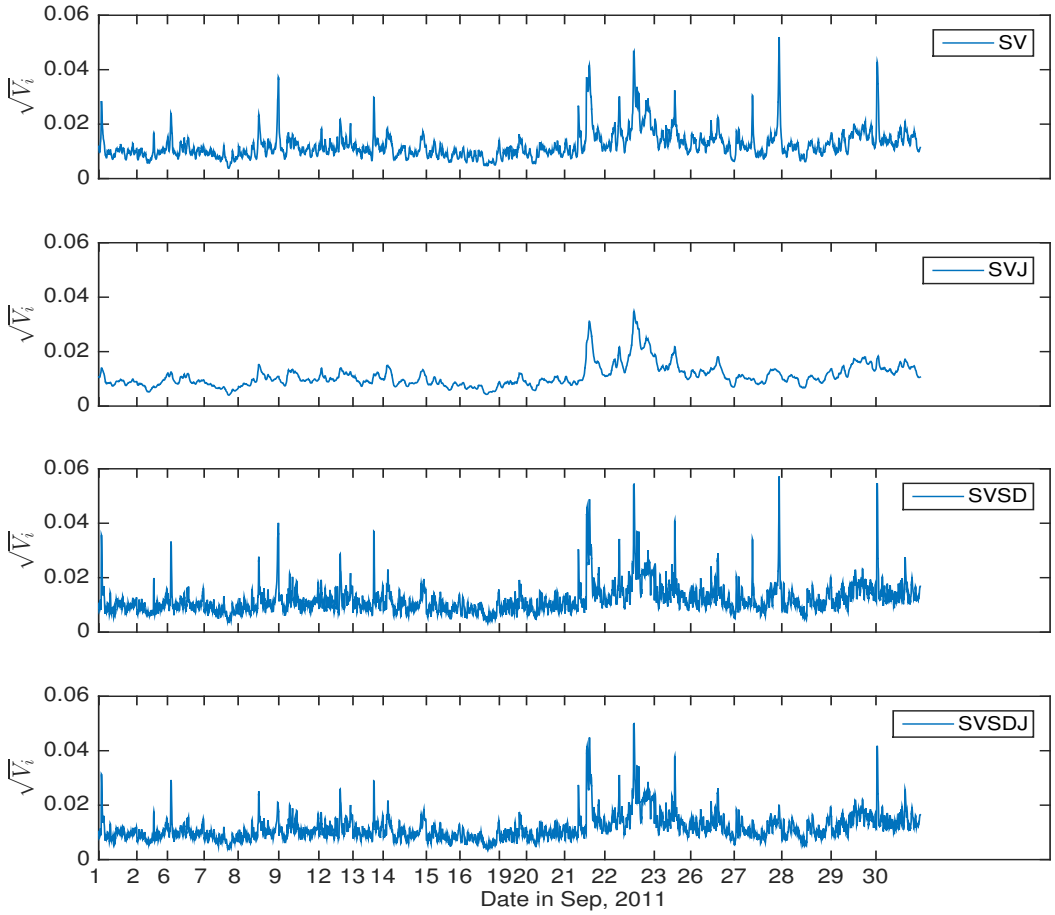

Figure 5: Estimated spot volatility. The estimate is obtained from the posterior mean (average of draws after burning in).

slightly above in most days.

In the realized volatility literature, jump variation is measured by the difference between realized volatility and bipower variation, see Barndorff-Nielsen and Shephard (2006). 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 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 our model is correctly specified, $\widehat{IJV}_t$ and JV_t should converge to the same value. We plot $\widehat{IJV}_t$ and JV_t in each day in Figure 8. JV_t is a lot larger than $\widehat{IJV}_t$ in most days. One explanation for this is that our models use returns sampled at a finer grid. As noted

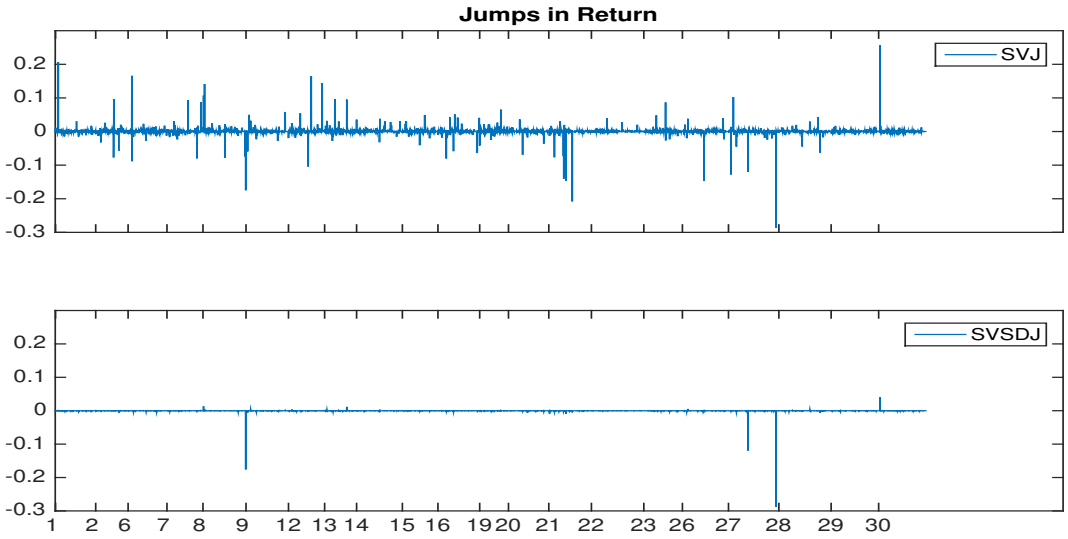

Figure 6: Estimated jumps.

by Christensen, Oomen, and Podolskij (2014), jump variation based on coarser data tend to attribute a burst in volatility to jumps in return. Figure 9 depicts the first hour of logarithmic price in the day when JV_t is the highest in the sample period, which is Sep 14th. 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 is no clear indication of large discrete price movement. Another reason for the difference is that the SVJ and SVSDJ model would take the level of volatility into consideration: large returns happened during high volatility period is less likely to be identified as jumps compare to large returns that occurred during low volatility period.

On the other hand, there are days when JV does not identify any jump but IJV does, for example, on Sep 27th. In Figure 10, we plot the returns in the first hour since opening, and compare that with Sep 14th. Although the magnitude of return variation is similar in the two plots, return variation happened over shorter durations on Sep 27th, and the SVJ model is able to identify some jumps.

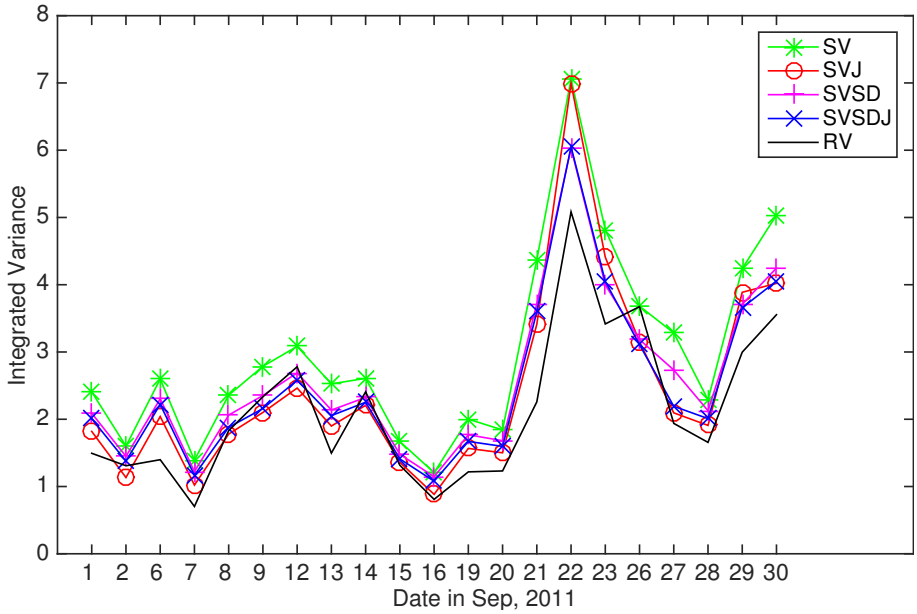

Figure 7: Estimated daily integrated volatility and realized bipower variation.

The proportion of variation due to jumps can be computed by

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

In the sample periods, 1.8% of the quadratic variation is from jumps in the SVJ model, while 0.3% of the total variation is from jumps in the SVSDJ model. The sample period is one month and it does not cover well known periods of market stress, so the estimated proportion does not serve as an indication of the magnitude of jump variation. However, using realized volatility, jump variation accounts for 8.8% of the total variation.

5 Conclusion

This paper puts forward a jump-diffusion model SVSDJ to jointly model volatility and conditional duration. Market microstructure theory suggests that durations between trades provide information to market participants, so volatility and durations are inter-dependent. Our model analyses the interdependence and utilizes this relationship to gain

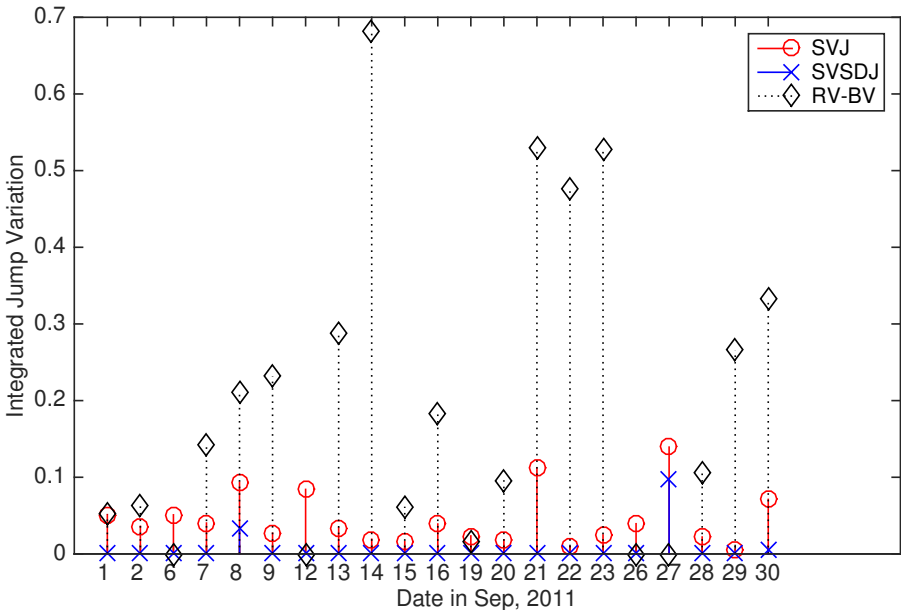

Figure 8: Estimated daily IJV and JV.

information about volatility. Given the nature of durations, observations are irregularly spaced. We develop an MCMC algorithm for inference about irregularly 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 converted to integrated volatility in a given horizon. Knowing when jumps happen and how large the jumps are helps us understand jump dynamics and price jump risk.

Applications to IBM data using our model and two nested alternatives reveal 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, total variation due to price jumps becomes smaller as we use returns sampled at finer grids. Third, volatility and conditional durations are interdependent, and modeling them jointly significantly reduces the need for price jumps.

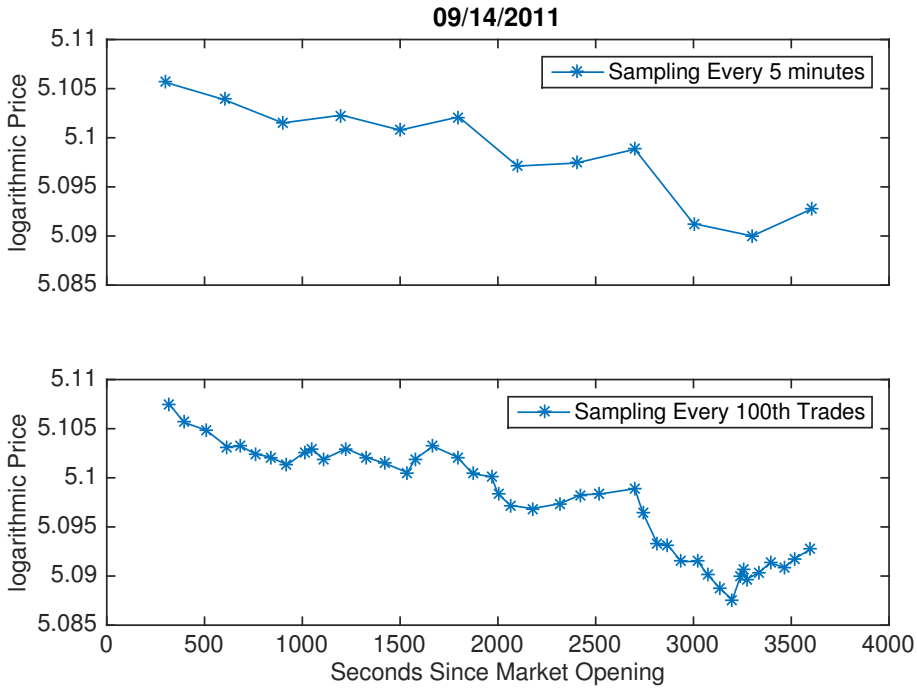

Figure 9: Logarithmic price on September 14th, the day with the highest jump variation. Top panel is sampled at every 5 minutes. Bottom Panel is sampled at every 100th trade.

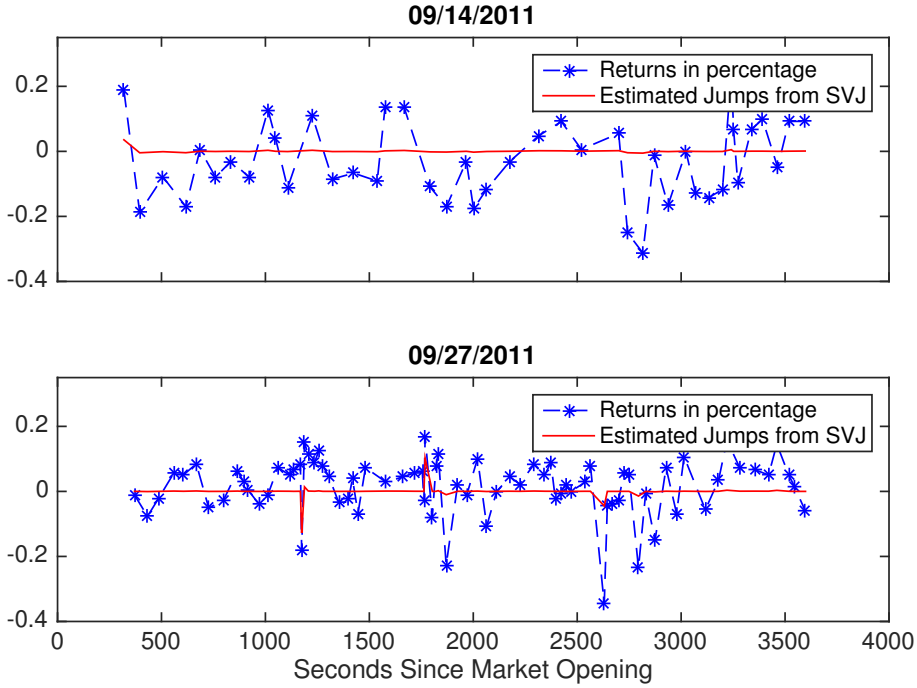

Figure 10: Returns and Jumps from model SVJ. The top panel is on September 14th, the day with the highest jump variation. The bottom panel is on September 27th, the day with the highest integrated jump variation.

Appendix: MCMC algorithm for SVSDJ model

The full detail of the MCMC algorithm for SVSDJ model are provided here. Other models are special cases of this model and the algorithm can be easily derived.

1. Initialize Z and Θ. For the SD, SDexp, SV and SVJ model, the initial value for states and parameters are obtained from QMLE estimations using Kalman Filter. For the VD and VDJ model, the initial values are taken from the estimation results from SD, SV and SVJ.

2. Sample the jump block. For notational simplicity, we use $\hat{r}_{i+1} = r_{i+1} - (m_{i+1} - m_i)$.

   (a) The conditionals of the jump indicator J_{i+1} marginalized over ξ_{i+1} is

   $$p(J_{i+1}|\text{rest}) \propto \exp\left(-\frac{1}{2}\frac{\left(\hat{r}_{i+1} - J_{i+1}\mu_J\right)^2}{V_i D_{i+1} + J_{i+1}\sigma_J^2}\right)\left(\gamma D_{i+1}\right)^{J_{i+1}}\left(1-\gamma D_{i+1}\right)^{1-J_{i+1}}.$$

   Define odds ratio or $= p(J_{i+1}=1|\text{rest})/p(J_{i+1}=0|\text{rest})$, then

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

   and it can sampled from a Bernoulli distribution. Conditional on J_{i+1}, the jump sizes ξ_{i+1}^y can be updated from

   $$p(\xi_{i+1}^y|J_{i+1}=0,\text{rest}) \sim N(\mu_J, \sigma_J^2)$$
   $$p(\xi_{i+1}^y|J_{i+1}=1,\text{rest}) \sim N(\mu_J^*, \sigma_J^{*2}),$$

   where

   $$\mu_J^* = \Sigma_J^*\left(\hat{r}_{i+1}V_i^{-1}D_{i+1}^{-1} + \mu_J/\sigma_J^2\right)$$
   $$\sigma_J^{2*} = \left(V_i^{-1}D_{i+1}^{-1} + 1/\sigma_J^2\right)^{-1}.$$

(b) To update γ, we draw from the proposal density in equation (10) and accept using the Independent MH algorithm.

(c) We use a normal-inverse-Gamma prior for μ_J and Σ_J:

$$p(\sigma_J^2) \sim \text{IG}(w_J, f_J)$$
$$p(\mu_J|\sigma_J^2) \sim N(c_J, \sigma_J^2/k_J).$$

Let $N_J = \sum J_i$ and $\bar{\xi} = \sum_{i:J_i=1} \xi_i$, the posteriors are

$$p(\sigma_J^2|\text{rest}) \sim IG(w_J^*, f_J^*)$$
$$p(\mu_J|\sigma_J^2, \text{rest}) \sim N(c_J^*, \sigma_J^2/k_J^*),$$

where

$$f_J^* = f_J + N_J/2$$
$$w_J^* = w_J + \frac{1}{2}\left(\sum_{i:J_i=1} \left(\xi_i - \bar{\xi}\right)^2 + \frac{k_J N_J}{k_J + N_J}\left(\bar{\xi} - c_J\right)^2\right)$$
$$k_J^* = k_J + N_J$$
$$c_J^* = \frac{k_J}{k_J + N_J} c_J + \frac{N_J}{k_J + N_J}\bar{\xi}.$$

3. Sample the OU block. The measurement density is given by:

$$p(D_{i+1}|\lambda_i, \Theta) \propto \lambda_i^{-d_s} D_{i+1}^{d_s - 1} exp(-\frac{D_{i+1} d_s}{\lambda_i}),$$

and $p(\tilde{r}_{i+1}|V_i) \sim N(0, V_i D_{i+1})$, where $\tilde{r}_{i+1} = r_{i+1} - \xi_{i+1}^y J_{i+1} - m_{i+1} + m_i$.

(a) The conditional distribution of X_i is

$$p(X_i|\text{rest}) \propto \underbrace{p(X_{i+1}|X_i, D_{i+1}, \Theta)p(X_i|X_{i-1}, D_{i+1}, \tilde{r}_{i+1}, \Theta)}_{\pi_1}$$
$$\times \underbrace{p(\tilde{r}_{i+1}|X_i, D_{i+1}, \Theta)}_{\pi_2} \underbrace{p(D_{i+1}|X_i)}_{\pi_3}.$$

As explained in section 3.2, π_1 has a multivariate normal kernel as in equation (12), and we can rewrite it to

$$\pi_1 \propto \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{X_{i,1}-\mu^*_{i,1}}{\sigma^*_{i,1}}\right)^2 + \left(\frac{X_{i,2}-\mu^*_{i,2}}{\sigma^*_{i,2}}\right)^2 - \frac{2\rho\left(X_{i,1}-\mu^*_{i,1}\right)\left(X_{i,2}-\mu^*_{i,2}\right)}{\sigma^*_{i,1}\sigma^*_{i,2}}\right)\right),$$

The measurement density π_2 and π_3 are approximated by q_2 and q_3 as explained in section 3.2. Let μ_a and μ_b denote the right-hand-side of equation (13) and (14), we have

$$q_1 \propto \exp\left(\mu_a X_{i,1}\right)$$
$$q_2 \propto \exp\left(\mu_b X_{i,2}\right).$$

The proposal density for X_i is multivariate normal:

$$q(X_i) \propto \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{X_{i,1}-\mu^{**}_{i,1}}{\sigma^*_{i,1}}\right)^2 + \left(\frac{X_{i,2}-\mu^{**}_{i,2}}{\sigma^*_{i,2}}\right)^2 - \frac{2\rho\left(X_{i,1}-\mu^{**}_{i,1}\right)\left(X_{i,2}-\mu^{**}_{i,2}\right)}{\sigma^*_{i,1}\sigma^*_{i,2}}\right)\right),$$

where

$$\mu_{i,1}^{**} = \mu_{i,1}^{*} + \mu_a \left(\sigma_{i,1}^{*}\right)^2 + \rho\sigma_{i,1}^{*}\sigma_{i,2}^{*}\mu_b$$
$$\mu_{i,2}^{**} = \mu_{i,2}^{*} + \mu_b \left(\sigma_{i,2}^{*}\right)^2 + \rho\sigma_{i,1}^{*}\sigma_{i,2}^{*}\mu_a.$$

Generate X_i from this multivariate normal density, then accept it with probability

$$\alpha = \min\left\{\frac{\pi_2\left(X_i^{(g+1)}\right)\pi_3\left(X_i^{(g+1)}\right)q_2\left(X_i^{(g)}\right)q_3\left(X_i^{(g)}\right)}{\pi_2\left(X_i^{(g)}\right)\pi_3\left(X_i^{(g)}\right)q_2\left(X_i^{(g+1)}\right)q_3\left(X_i^{(g+1)}\right)},\ 1\right\}.$$

(b) The posterior for Σ_x is given by equation (15). If Σ_{i+1} can be decomposed into $A_1\Sigma_x A_2$,

$$\mathrm{vec}(\Sigma_{i+1}) = (\Psi\oplus\Psi)^{-1}(I_2 - e^{-(\Psi\oplus\Psi)D_{i+1}})\mathrm{vec}(\Sigma_x) = \left(A_2^T\otimes A_1\right)\mathrm{vec}(\Sigma_x).$$

Then,

$$\begin{aligned}&\left(X_{i+1}-\mu_{i+1}\right)'\Sigma_{i+1}^{-1}\left(X_{i+1}-\mu_{i+1}\right)\\ =&\left(X_{i+1}-\mu_{i+1}\right)'A_2^{-1}\Sigma_x\left(A_1^{-1}\left(X_{i+1}-\mu_{i+1}\right)\right)\\ =&\mathrm{trace}\left(\Sigma_x^{-1}E_{i+1}\right),\end{aligned}$$

where $E_{i+1} = A_1^{-1}\left(X_{i+1}-\mu_{i+1}\right)\left(X_{i+1}-\mu_{i+1}\right)' A_2^{-1}$. It is easy to show that

$$\begin{aligned}\mathrm{vec}(E_{i+1}) =& (A_2^T\otimes A_1)^{-1}\mathrm{vec}\left(\left(X_{i+1}-\mu_{i+1}\right)\left(X_{i+1}-\mu_{i+1}\right)^T\right)\\ =&(I_2 - e^{-(\Psi\oplus\Psi)D_{i+1}})^{-1}(\Psi\oplus\Psi)\mathrm{vec}\left(\left(X_{i+1}-\mu_{i+1}\right)\left(X_{i+1}-\mu_{i+1}\right)^T\right).\end{aligned}$$

Since we also have $|\Sigma_{i+1}| = |A_1||\Sigma_x||A_2|$, the inverse Wishart proposal density in equation (16) is proportional to the target density. In other words, if

Σ_{i+1} can be decomposed into $A_1 \Sigma^x A_2$ (note that we do not need to find the decomposition, we just need it to exist), the posterior has an inverse Wishart kernel. In general, this is not the case, but the inverse Wishart proposal density still serves as a good approximation to the target density, and we us IMH to sample the posterior.

(c) Next, the posterior of Ψ is given by

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

From the Euler discretization in equation (17), the posterior for Ψ with a flat prior is given by a Matrix Normal distribution

$$\begin{aligned}
\Psi &\sim MN\left(\left(\left(B_2'B_2\right)^{-1} B_2'B_1\right)', \left(B_2'B_2\right)^{-1} \otimes \Sigma^x\right) \\
B_1 &= \left(\frac{X_2-X_1}{\sqrt{D_2}}, \ldots, \frac{X_N-X_{N-1}}{\sqrt{D_N}}\right)' \\
B_2 &= \left(\left(\mu^x - X_1\right)\sqrt{D_2}, \ldots, \left(\mu^x - X_{N-1}\right)\sqrt{D_N}\right)'.
\end{aligned}$$

To ensure tail-boundness, we use a multivariate t distribution rather than Normal as the proposal density. The difference between this proposal density and the posterior depends on the difference between Euler discretization and the exact solution. In the application, we find the acceptance rate to be close to 50%.

(d) The posterior for μ^x given a multivariate normal prior $N(C_x, Z_x)$ is

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

4. For the MMN block, we assume $E(m_i) = 0$ and the prior distribution of m_i is $N(0, \sigma_m^2)$. We use the notation $\check{r}_{i+1} = r_{i+1} - \xi_{i+1}^y J_{i+1}$.

   (a) The posterior of m_i is given by:

$$p(m_i|\text{rest}) \sim N(c_m,\ z_m)$$

   where

$$
\begin{aligned}
z_m^{-1} &= \frac{1}{V_i D_{i+1}} + \frac{1}{V_{i-1} D_i} + \frac{1}{\sigma_m^2}\\
c_m &= z_m\left(\frac{m_{i+1} - \check{r}_{i+1}}{V_i D_{i+1}} + \frac{\check{r}_i + m_{i-1}}{V_{i-1} D_i}\right).
\end{aligned}
$$

   (b) To update σ_m^2, we use a conjugate Inverse Gamma prior $IG(f_m, w_m)$. The posterior is

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

5. The shape parameter is sampled using the procedure in Section 3.4.

Acknowledgments

We thank Asger Lunde, Kim Christensen, Walter Thurman, Atsushi Inoue, Peter Bloomfield and Rasmus Varneskov for helpful comments. We also want to thank participants at Computational Methods for Jump Processes Workshop (2014), and seminars at NCSU and CREATES for valuable discussions. Wei Wei acknowledges financial support from the The Danish Council for Independent Research, Social Sciences (4003-00106B/FSE) and CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

References

Ait-Sahalia, Y., Mykland, P. A., and Zhang, L. (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., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). 'The distribution of realized stock return volatility', *Journal of Financial Economics*, 61(1): 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). 'The Distribution of Realized Exchange Rate Volatility', *Journal of the American Statistical Association*, 96: 42–55.

——— (2003). 'Modeling and Forecasting Realized Volatility', *Econometrica*, 71(2): 579–625.

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). 'Designing

Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise', *Econometrica*, 76(6): 1481–1536.

——— (2009). 'Realized kernels in practice: trades and quotes', *Econometrics Journal*, 12(3): C1–C32.

Barndorff-Nielsen, O. E., and Shephard, N. (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.

——— (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.

——— (2004). 'Power and Bipower Variation with Stochastic Volatility and Jumps', *Journal of Financial Econometrics*, 2(1): 1–37.

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

——— (2006). 'Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation', *Journal of Financial Econometrics*, 4(1): 1–30.

Bates, D. S. (2000). 'Post-'87 crash fears in the S&P 500 futures option market', *Journal of Econometrics*, 94(1-2): 181–238.

Bauwens, L., and Giot, P. (2000). 'The Logarithmic ACD Model: An Application to the Bid-Ask Quote Process of Three NYSE Stocks', *Annales d'Economie et de Statistique*, 60(60): 117–149.

Bauwens, L., and Veredas, D. (2004). 'The stochastic conditional duration model: a

latent variable model for the analysis of financial durations', *Journal of Econometrics*, 119(2): 381–412.

Bollerslev, T., and Todorov, V. (2011). 'Tails, Fears, and Risk Premia', *The Journal of Finance*, 66(6): 2165–2211.

Chib, S., Nardari, F., and Shephard, N. (2002). 'Markov chain Monte Carlo methods for stochastic volatility models', *Journal of Econometrics*, 108(2): 281–316.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). 'Fact or friction: Jumps at ultra high frequency', *Journal of Financial Economics*, 114(3): 576 – 599.

Easley, D., and O'Hara, M. (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 Russell, J. R. (1998). 'Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data', *Econometrica*, 66(5): 1127–1162.

Eraker, B., Johannes, M., and Polson, N. (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.

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

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

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

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

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

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

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). 'Bayesian analysis of stochastic volatility models with fat-tails and correlated errors', *Journal of Econometrics*, 122(1): 185–212.

Johannes, M., and Polson, N. (2010). 'MCMC Methods for Continuous-Time Financial Econometrics', in Y. A.-S. P. HANSEN (ed.), *Handbook of Financial Econometrics: Applications*, vol. 2 of *Handbooks in Finance*, pp. 1 – 72. Elsevier, San Diego.

Johannes, M. S., Polson, N. G., and Stroud, J. R. (2009). 'Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices', *Review of Financial Studies*, 22(7): 2559–2599.

Kim, S., Shephard, N., and Chib, S. (1998). 'Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models', *Review of Economic Studies*, 65(3): 361–93.

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.

Pelletier, D., and Zheng, H. (2013). 'Joint Modeling of High-Frequency Price and Duration Data', Discussion paper, North Carolina State University.

Pitt, M. K., and Shephard, N. (1999). 'Filtering via Simulation: Auxiliary Particle Filters', *Journal of the American Statistical Association*, 94(446): 590–599.

Renault, E., van der Heijden, T., and Werker, B. J. (2014). 'The dynamic mixed hitting-time model for multiple transaction prices and times', *Journal of Econometrics*, 180(2): 233 – 250.

Renault, E., and Werker, B. J. (2011). 'Causality effects in return volatility measures with random times', *Journal of Econometrics*, 160(1): 272 – 279, Realized Volatility.

Russell, J. R., and Engle, R. F. (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.

Son, Y. S., and Oh, M. (2006). 'Bayesian Estimation of the Two-Parameter Gamma Distribution', *Communications in Statistics - Simulation and Computation*, 35(2): 285–293.

Tay, A. S., Ting, C., Tse, Y. K., and Warachka, M. (2011). 'The impact of transaction duration, volume and direction on price dynamics and volatility', *Quantitative Finance*, 11(3): 447–457.

Todorov, V. (2010). 'Variance Risk-Premium Dynamics: The Role of Jumps', *Review of Financial Studies*, 23(1): 345–383.

Zhang, L., Mykland, P. A., and Ait-Sahalia, Y. (2005). 'A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data', *Journal of the American Statistical Association*, 100: 1394–1411.

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