ABSTRACT

WEI, WEI. Three Essays on Financial Durations. (Under the direction of Denis Pelletier.)

This thesis centers on volatility modeling and risk management using different types of financial durations. Chapter 1 models the durations between trades in high-frequency data jointly with financial returns. We put forward a stochastic volatility and stochastic conditional duration with cojumps (SVSDCJ) model to examine the interdependent relationship between volatility and trade durations, and the role of jumps in the price, volatility and conditional duration process. We develop a MCMC algorithm for the inference of irregularly spaced multivariate process with jumps. The algorithm also provides smoothed estimates of the spot volatility, jump times and jump sizes. Chapter 2 focuses on price duration, which is the time it takes for the asset price to change a given amount. We propose a class of stochastic price duration models to estimate high-frequency volatility based on the passage time theory for Brownian motions. In particular, we extend the baseline model to incorporate information from trade durations. Chapter 3 develops a duration-based test to evaluate an important risk management tool, Value at Risk (VaR). Estimating VaR is a challenging problem, and the popular VaR forecast relies on unrealistic assumptions. Hence, assessing the performance of VaR is of great importance. We propose the geometric-VaR test, which utilizes the durations between the violations of VaR as well as the value of VaR. We conduct a Monte Carlo study based on desk-level data and show that our test has high power detecting various types of VaR misspecifications.
Three Essays on Financial Durations

by

Wei Wei

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Economics

Raleigh, North Carolina

2013

APPROVED BY:

Peter Bloomfield

Walter Thurman

Atsushi Inoue

Denis Pelletier

Chair of Advisory Committee
DEDICATION

To the memory of my mother.
BIOGRAPHY

The author was born on November 11, 1985 in a bustling little city Mianyang in southwest China. She attended Beijing Normal University for her undergraduate studies, majoring in Astronomy and co-majoring in Business Administration. In 2007, she received her Master’s degree in Finance and Economics at London School of Economics. Upon graduation, she spent a year enjoying her passion for food, books and traveling. She resumed to academic life in 2008, when she was enrolled in North Carolina State University to pursue her PhD degree in Economics. Her research interests are in the area of financial econometrics, financial economics and Bayesian econometrics.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Denis Pelletier. Without him, this dissertation would not have been possible. I learned a lot from him, not only in the subjects of finance and econometrics, but also on how to be a researcher. His insight and humor made my many visits to his office extremely useful and enjoyable.

I would also like to thank my committee members: Dr. Peter Bloomfield, Dr. Atsushi Inoue and Dr. Walter Thurman. Their valuable comments and suggestions greatly enriched this dissertation. I am also deeply grateful for their support and encouragement.

I would like to express my gratitude to Dr. Tom Vukina, who has helped me tremendously during my time at NC State. He also introduced me to teaching and I truly enjoyed working with him. I would also like to extend my thanks to the other faculty and my classmates at NC State.

Last but not least, I would like to thank my fiancé Robert Kane and my father Wei Guangwen for their endless support and love. Without them, none of this could have happened.
TABLE OF CONTENTS

LIST OF TABLES .......................................................... vii
LIST OF FIGURES .......................................................... viii

Chapter 1 A Jump Diffusion Model for Volatility and Duration ........ 1
  1.1 Introduction ......................................................... 1
  1.2 Model Specification ............................................... 7
    1.2.1 Setup ........................................................ 7
    1.2.2 Properties ................................................. 10
  1.3 Bayesian Inference ................................................ 13
    1.3.1 Priors ....................................................... 15
    1.3.2 Posteriors of State variables .............................. 16
    1.3.3 Posteriors of Parameters ................................. 17
  1.4 Simulation Studies ............................................... 18
  1.5 Empirical Results ............................................... 19
    1.5.1 Data ......................................................... 19
    1.5.2 Estimation ................................................. 22
  1.6 Conclusion ....................................................... 29

Chapter 2 A Stochastic Price Duration Model for Estimating High-
 Frequency Volatility ................................................... 32
  2.1 Introduction ....................................................... 32
  2.2 Model Specification ............................................... 36
    2.2.1 Stochastic Price Duration ................................. 36
    2.2.2 Stochastic Trade Duration ................................. 39
    2.2.3 Modeling Price Durations and Trade Durations Jointly 40
  2.3 Estimation Procedure and Simulation Studies ..................... 41
    2.3.1 Linear State Space Representation ......................... 41
    2.3.2 Simulations without Microstructure Noise ................ 44
    2.3.3 Simulation with Microstructure Noise ..................... 46
  2.4 Empirical Results ............................................... 49
    2.4.1 Data ......................................................... 49
    2.4.2 Diurnal Pattern ............................................. 51
    2.4.3 Estimation ................................................. 52
  2.5 Conclusion ....................................................... 53

Chapter 3 Backtesting Value at Risk using a Parametric Duration Model 56
  3.1 Introduction ....................................................... 56
  3.2 Duration-based Backtesting .................................... 60
3.2.1 Geometric Test ........................................... 63
3.2.2 Geometric-VaR Test ..................................... 64
3.3 Test Implementation ......................................... 67
3.4 Simulation Studies .......................................... 70
  3.4.1 Data Generating Process ................................. 70
  3.4.2 Backtesting VaR ........................................ 72
  3.4.3 Finite Sample Power of the tests ..................... 73
3.5 Empirical Results .......................................... 78
3.6 Conclusions ................................................ 81

References ...................................................... 83

Appendix ......................................................... 89
  Appendix A  MCMC algorithm for SVSDCJ model ............ 90
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Parameter Inference using Simulated data.</td>
<td>18</td>
</tr>
<tr>
<td>1.2</td>
<td>Summary statistics for adjusted IBM returns and durations.</td>
<td>22</td>
</tr>
<tr>
<td>1.3</td>
<td>Parameter estimates in the SVSDCJ, SVSDJ and SVSD model.</td>
<td>23</td>
</tr>
<tr>
<td>2.1</td>
<td>Monte Carlo results for the estimated IV</td>
<td>48</td>
</tr>
<tr>
<td>2.2</td>
<td>Summary statistics for IBM in 2011/08/01-2011/09/31</td>
<td>50</td>
</tr>
<tr>
<td>2.3</td>
<td>Parameter estimates for the SPD1 and SPD0 model</td>
<td>53</td>
</tr>
<tr>
<td>3.1</td>
<td>Parameter Estimates of NGARCH-t(d) Model for Four Business Lines</td>
<td>72</td>
</tr>
<tr>
<td>3.2</td>
<td>Power of 10% Duration-based Tests on 1% VaR in Four Business Lines</td>
<td>75</td>
</tr>
<tr>
<td>3.3</td>
<td>Power of 10% Duration-based Tests on 5% VaR in Four Business Lines</td>
<td>76</td>
</tr>
<tr>
<td>3.4</td>
<td>Fraction of Samples Where Test Is Feasible (1% and 5% VaR)</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>P/Ls and VaRs for Four Business Lines: Descriptive Statistics</td>
<td>79</td>
</tr>
<tr>
<td>3.6</td>
<td>Backtesting Actual VaRs from Four Business Lines</td>
<td>81</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

| Figure 1.1 | The dependence between durations and squared returns. | 19 |
| Figure 1.2 | Nonparametric estimate of the Diurnal patterns. | 21 |
| Figure 1.3 | Estimated spot volatility. | 25 |
| Figure 1.4 | QQ plots of the return residuals. | 26 |
| Figure 1.5 | Estimated conditional duration. | 27 |
| Figure 1.6 | Estimated jumps. | 28 |
| Figure 1.7 | Estimated daily integrated volatility and realized bipower variation. | 29 |
| Figure 1.8 | Estimated daily IJV and JV. | 30 |
| Figure 1.9 | Logarithmic price on September 14th, the day with the highest jump variation. | 31 |
| Figure 2.1 | PDF of the true distributions versus their normal approximations | 43 |
| Figure 2.2 | True volatility versus the estimated volatility from SPD0 and SPD1. | 45 |
| Figure 2.3 | Logarithmic conditional duration and its estimates from SPD1. | 46 |
| Figure 2.4 | Sampling points in one trading day. | 50 |
| Figure 2.5 | Diurnal patterns in volatility and conditional duration. | 54 |
| Figure 2.6 | Daily volatility of IBM in 2011/08/01-2011/09/31, 44 days. | 54 |
| Figure 3.1 | Time Series of P/Ls, VaRs and Violations from Four Business Lines. | 80 |
Chapter 1

A Jump Diffusion Model for Volatility and Duration

1.1 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 a 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, 2003), and Barndorff-Nielsen
and Shephard (2001, 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. We propose a jump diffusion process to model the movement of returns and durations jointly. The 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, we can measure 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 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; 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 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 analyze 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 Pelletier and Zheng (2012), we 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 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, 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 a diffusion part and a 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).

We 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 extremely high to explain the occasional large fluctuations. We also consider jumps in volatility and expected durations. As Eraker, Johannes, and Polson (2003) noted, jumps in return do not 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 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 a kernel-based volatility estimator. Ait-Sahalia,
Mykland, and Zhang (2005) show that if the noise term is accounted for explicitly, sampling as often as possible is optimal.

Our approach in dealing with microstructure noise 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 $L$th 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 we adopt and it allows sampling at finer grid than current parametric models.

Estimating a stochastic volatility model usually involves approximation. Jacquier, Polson, and Rossi (1994) introduced Bayesian Markov Chain Monte Carlo (MCMC) methods, which allow for exact finite sample inference. Eraker, Johannes, and Polson (2003) use MCMC to analyze the impact of jumps in returns and volatility. We resort to MCMC for the estimation of the parameters and latent variables. Our 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 an ad hoc filtering technique. The estimated conditional volatility and jumps are useful in applications such as 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.
The rest of this paper is organized as follows. In section 2 we describe the model specification. Section 3 discusses the Bayesian inference and simulation studies. Section 4 presents the empirical results using IBM shares data. Section 5 concludes.

1.2 Model Specification

1.2.1 Setup

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

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

(1.1)

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 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 $\gamma$ is constant. Given a time interval $\Delta$, the probability of observing $n$ jumps is $e^{-\gamma \Delta} (\gamma \Delta)^n / n!$. Jump sizes $\xi_t^y$ 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 the application, we sample every $L$th 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 $\lambda_{t_i}$ denote the conditional expectation of $D_{i+1}$ given the 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 $\lambda_{t_i}$ times a i.i.d random variable with positive support, i.e., $D_{i+1} = \lambda_{t_i} e_i$. We assume an exponential distribution for $e_i$ in this paper. Our model can be easily extended to accommodate other distributions.
To create persistence and interdependence between volatility and duration, we follow Pelletier and Zheng (2012) and model the logarithm of $\lambda_t$ and $V_t$ using a 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 imposing non-negativity without putting extra constraints on parameters. To explain a prolonged effect from news, we add a Poisson jump component to the Gaussian OU process. Let $X_t = (\log(V_t), \log(\lambda_t))'$, where $X_t$ solves:

$$dX_t = -\Psi(X_t - \mu^x)dt + S_x dW^x_t + \xi^x_t dN^x_t,$$

(1.2)

where $\Psi$ is a $2 \times 2$ matrix that measures the mean reversion and dependence between conditional duration and volatility. The OU process mean reverts to $\mu^x$, the diffusive long-run mean. $S_x$ measures the variation of logarithmic volatility and logarithmic duration, and $S_x = \text{diag}(\sigma^2_v, \sigma^2_\lambda)$. $W^x_t$ is a Brownian motion in $\mathbb{R}^2$ with $dW^v_t dW^\lambda_t = \rho dt$, where $\rho$ is the instantaneous correlation. The instantaneous covariance matrix is given by

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

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

\[
\xi_t \sim N(\mu_J, \Sigma_J).
\]

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, we discretize \( dy_t \) over the durations:

\[
y_{i+1} - y_i = r_{i+1}^c = \mu^y D_{i+1} + \sqrt{V_t D_{i+1}} \xi_{i+1}^y + \xi_{i+1}^\lambda J_{i+1},
\]

where the subscript \( i \) denotes the time of \( i \)th event, \( t_i \). Jumps are assumed to be rare, \( \gamma \) 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 a 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,

\[
y_i^o = y_i + m_i.
\]

We assume that microstructure noise is i.i.d. mean zero and normal, \( m_i \sim N(0, \sigma_m^2) \), and that microstructure noise is independent of the efficient price\(^1\). The observed return has

\(^1\)These assumptions may be violated for microstructure noise in tick-by-tick price series. However,
an MA(1) contamination:

\[ r_{i+1}^o = y_{i+1}^o - y_i^o = r_{i+1}^e + m_{i+1} - m_i. \quad (1.6) \]

The logarithmic volatility and conditional duration \( X_t \) are discretized using the exact solution of the OU process with jumps (See Section 2.2 for descriptions of the solution). Equations (1.4), (1.6), and the time discretization of \( X_t \) over durations form the SVSDCJ model:

\[
\begin{align*}
  r_{i+1}^o &= \mu^y D_{i+1} + \sqrt{V_i D_{i+1}} \epsilon_{i+1}^g + \xi_{i+1}^y J_{i+1} + m_{i+1} - m_i \\
  X_{i+1} &= (I_2 - e^{-\Psi D_{i+1}})\mu^x + e^{-\Psi D_{i+1}} X_i + \xi_{i+1}^x J_{i+1} + U_{i+1},
\end{align*}
\]

where

\[
U_{i+1} \sim N(0, \Sigma_{i+1})
\]

\[\text{vec}(\Sigma_{i+1}) = (\Psi \oplus \Psi)^{-1} (I_2 - e^{-\Psi D_{i+1}}) \text{vec}(\Sigma_x).\]

### 1.2.2 Properties

The exact solution to the SDE (2.7) is given by

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

where \( Z_t = \sum_{j=0}^{N_t} e^{-\Psi (t-\tau_j)} \xi_{\tau_j}^x \) (See Kloeden and Platen, 1992; Ross, 1996). Assuming that there is at most one jump in \( D_{i+1} \) with jump probability \( \gamma D_{i+1} \), we would arrive at 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.
the discretized version in (1.7). Note that the long-run or unconditional mean of $X_t$ is no longer $\mu^x$ 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

$$E(X_t) = \mu^x + \lambda \Psi^{-1} \mu_j^x$$
$$\text{vec}(\text{Var}(X_t)) = (\Psi \oplus \Psi)^{-1} \left( \text{vec}(\Sigma_X) + \lambda \text{vec} \left( (\mu_j^x)(\mu_j^x)' + \Sigma_j^x \right) \right).$$

(1.8)

To ensure the existence of a stationary solution, it is sufficient that $\Psi$ has only eigenvalues with positive real parts so that $e^{-\Psi t} \to 0$ as $t \to 0$ (See Gardiner, 2009). We discuss two important subsets of the parameter space. First, if $\Psi_{11} > 0, \Psi_{22} > 0, \Psi_{12}$ and $\Psi_{21}$ have the same sign, and $\det(\Psi) > 0$, then all the eigenvalues of $\Psi$ will be real and positive. In this case, the system reverts to its diffusive mean following an exponential decay. This encompasses the case when $\Psi$ is diagonal and the diagonal elements are positive. Second, if $\Psi_{11} > 0, \Psi_{22} > 0, \Psi_{12}$ and $\Psi_{21}$ has opposite sign, and $(\Psi_{11} - \Psi_{22})^2 < -4\Psi_{12}\Psi_{21}$, the eigenvalues of $\Psi$ have positive real parts with imaginary parts, and the the system oscillates to the diffusive mean.

We use an Euler discretization for $X_t$ without jumps to gain some insight about the parameters:

$$X_{i+1} = \Psi \mu^x D_{i+1} + (I_2 - \Psi D_{i+1})X_i + \Sigma^{1/2}_{i+1} \sqrt{D_{i+1}} \epsilon^x_{i+1},$$

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

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

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 that jumps are an important source of leverage effects. To accommodate this correlation, we assume the following covariance matrix for jump sizes:

\[
\Sigma^J = \begin{pmatrix}
\sigma^2_{J,y} & \rho_{yv} \sigma_{J,y} \sigma_{J,v} & \rho_{yd} \sigma_{J,y} \sigma_{J,d} \\
\rho_{yv} \sigma_{J,y} \sigma_{J,v} & \sigma^2_{J,v} & \rho_{vd} \sigma_{J,v} \sigma_{J,d} \\
\rho_{yd} \sigma_{J,y} \sigma_{J,d} & \rho_{vd} \sigma_{J,d} \sigma_{J,v} & \sigma^2_{J,d}
\end{pmatrix}.
\]

If a negative price jump is correlated with increased volatility, \(\rho_{yv}\) should be negative. The dependence between price jumps and the jumps in expected duration is measured.
by $\rho_{yd}$. If it is positive, it indicates that negative price jumps lead to shorter durations between trades. Lastly, $\rho_{vd}$ is the correlation of jump sizes in volatility and expected duration. It is expected to have the same sign as the correlation between volatility and expected durations $\rho$.

### 1.3 Bayesian Inference

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

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

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

$$
Z = \{V_i, \lambda_i, \xi^y_i, \xi^x_i, J_i, m_i\}_{i=1}^N
$$

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 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 latent volatility and conditional durations in our 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, we
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 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 \( p(Y|\Theta, Z) \). If \( \Theta \) 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. We 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 \( \hat{Z}_i = \frac{1}{G-K} \sum_{g=K+1}^{G} Z^{(g)}_i \). For an overview of MCMC methods in finance, see Johannes and Polson (2002).

We 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 \( \Psi, \mu^x \) and \( \Sigma_x \).
6. Sample \( \mu_J \) and \( \Sigma_J \).
1.3.1 Priors

Bayesian analysis requires the formulation of prior distributions. First, we choose a prior for $\Psi$ to ensure that the OU process stays in the stationary region. Other than imposing stationarity, the prior is diffuse. Specifically, we choose a truncated multivariate normal with large variance. Second, the prior for $\gamma$ and $\Sigma_J$ reflects our belief that jumps are large and infrequent. Jump intensity $\gamma$ is restricted in the region where $\gamma D$ does not exceed one.

Third, we use returns sampled at the highest frequency to form an informative prior for $\sigma_m^2$. 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 $m_i$. 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 $\sigma_m^2$. Under the assumption of uncorrelated noise, $\hat{\sigma}_m^2 = -\Sigma_i r_i r_{i+1}/N$. We specify an inverse Gamma prior for $\sigma_m^2$ 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 $m_i$. We use returns sampled from every $L$th 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 prior is.

Last, we choose an uninformative and conjugate prior for $\mu^x$ and $\mu_J$ since there is no prior belief about their values. There is no conjugate prior for $\Sigma_x$; we specify a diffuse inverse Wishart prior.

### 1.3.2 Posterior of State variables

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

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

The posterior is not a standard distribution, hence the adoption of the independent Metropolis-Hastings algorithm. Let $\pi(V_i)$ denote $p(V_i|\text{rest})$, the target density. We draw $V_i$ from $q(V_i)$, the proposal density, then accept the draw with probability $\alpha$ 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 an inverse gamma kernel from the observation equation. Following Jacquier, Polson, and Rossi (1994), we choose an inverse gamma distribution to approximate the lognormal kernel and combine it with the other inverse gamma kernel.
The updating of $\lambda_i$ follows the same procedure.

The conditional posterior of $J_i$ is Bernoulli with $J_i = 1$ indicating a jump arrival. We update $J$ following the algorithm in Eraker, Johannes, and Polson (2003). Jump sizes $\xi_i$ have a multivariate normal posterior conditional on $J, X, m, \Theta$ and $Y$.

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

### 1.3.3 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 depends on the latent variables. For example, we draw $\Psi, \Sigma_x$ and $\mu^x$ 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 $\Psi$ can not be sampled from directly, so we 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, $\Psi$ is drawn from a multivariate t distribution rather than Normal. $\Sigma_x$ has a posterior that is well approximated by an Inverse Wishart Distribution, and we use the Metropolis-Hasting algorithm to update it. $\mu^x$ has a conjugate multivariate normal distribution and it can be drawn directly.

The conditional distribution of $\mu_J$ and $\Sigma_J$ only depend on $\xi$ and $J$. Since $\xi$ are normal, the posterior of $\mu_J$ and $\Sigma_J$ can be derived from standard linear models. Conditional on $J$, the posterior of $\gamma$ is independent from other state variables and parameters. The posterior of $\gamma$ is sampled using a Metropolis-Hastings step. The posterior of $\sigma_m^2$ 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.
1.4 Simulation Studies

We use simulation studies to demonstrate the reliability of the estimation procedure. The simulated sample size is 5000. The posterior mean and the standard deviation is reported in Table 1.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>SVSDCJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_{11}$</td>
<td>0.009</td>
<td>0.0109 (0.0021)</td>
</tr>
<tr>
<td>$\psi_{21}$</td>
<td>0.002</td>
<td>0.0003 (0.0017)</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.003</td>
<td>0.0036 (0.0023)</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.006</td>
<td>0.0068 (0.0014)</td>
</tr>
<tr>
<td>$\mu_{\gamma}$</td>
<td>-8.8</td>
<td>-8.7874 (0.0565)</td>
</tr>
<tr>
<td>$\mu_{\delta}$</td>
<td>2.0</td>
<td>1.9134 (0.0521)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.0036</td>
<td>0.0052 (0.0013)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>-0.0006</td>
<td>-0.0006 (0.0007)</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.0025</td>
<td>0.0029 (0.0005)</td>
</tr>
<tr>
<td>$\mu_{J,r}$</td>
<td>-0.1</td>
<td>-0.0913 (0.0799)</td>
</tr>
<tr>
<td>$\mu_{J,v}$</td>
<td>0.7</td>
<td>0.7071 (0.3306)</td>
</tr>
<tr>
<td>$\mu_{J,d}$</td>
<td>-0.4</td>
<td>-0.3801 (0.2232)</td>
</tr>
<tr>
<td>$\sigma_{J,11}$</td>
<td>0.04</td>
<td>0.0526 (0.0294)</td>
</tr>
<tr>
<td>$\sigma_{J,12}$</td>
<td>-0.06</td>
<td>-0.0032 (0.0632)</td>
</tr>
<tr>
<td>$\sigma_{J,22}$</td>
<td>0.36</td>
<td>0.1764 (0.1611)</td>
</tr>
<tr>
<td>$\sigma_{J,13}$</td>
<td>0.006</td>
<td>-0.0402 (0.0417)</td>
</tr>
<tr>
<td>$\sigma_{J,23}$</td>
<td>-0.09</td>
<td>-0.0528 (0.0874)</td>
</tr>
<tr>
<td>$\sigma_{J,33}$</td>
<td>0.09</td>
<td>0.0895 (0.0851)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0005</td>
<td>0.0004 (0.0002)</td>
</tr>
<tr>
<td>$\sigma_m$</td>
<td>0.01</td>
<td>0.0099 (0.0002)</td>
</tr>
</tbody>
</table>

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

Formal tests of the consistency of the posterior simulator can be employed following the method proposed in Geweke (2004).
1.5 Empirical Results

1.5.1 Data

We apply our 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). 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.

![Figure 1.1: The dependence between durations and squared returns.](image-url)
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 and duration jointly. Figure 1.1 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 $L$th 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 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, we choose $L$ to be 100, leaving 6038 observations. At this frequency, the mean duration is about 78 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 $D_i^a = D_i / gd_i$,
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 by 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. We 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 by 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 1.2. Summary statistics for the adjusted returns and durations are given in Table 2.2.

Figure 1.2: Nonparametric estimate of the Diurnal patterns.
Table 1.2: Summary statistics for adjusted IBM returns and durations.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r^a )</td>
<td>0.00038</td>
<td>0.09421</td>
<td>0.01483</td>
</tr>
<tr>
<td>( D^a )</td>
<td>78.276</td>
<td>37.377</td>
<td>0.54025</td>
</tr>
</tbody>
</table>

1.5.2 Estimation

The posterior mean of the parameters for the SVSDCJ model are given in Table 1.3. For comparison, we 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 1.3.

The posterior mean of \( \Psi_{11} \) in the SVSDJ model is the smallest among the three models. Along with the smallest \( \Sigma_{11} \) estimate, it produces the most persistent spot volatility, as shown in Figure 1.3. The spot volatility estimate 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 \( \Psi_{11} \) and \( \Sigma_{11} \) estimates, and also the volatility estimates in Figure 1.3. The SVSDCJ model allows cojumps in returns and volatility, so the volatility process has less variation in the diffusive part (smaller \( \Sigma_{11} \)) and mean reverts more slowly (smaller \( \Psi_{11} \)) than what the SVSD model suggests. The estimated spot volatility for the SVSD model is presented in the bottom panel of Figure 1.3. The diffusive mean of the volatility process \( \mu_v \) is smaller in the SVSDCJ model than the SVSDJ or SVSD model. This is consistent with the large positive jumps we find in the volatility process.
Table 1.3: Parameter estimates in the SVSDCJ, SVSDJ and SVSD model

<table>
<thead>
<tr>
<th></th>
<th>SVSDCJ</th>
<th>SVSDJ</th>
<th>SVSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi_{11}$</td>
<td>0.0009 (0.0002)</td>
<td>0.0004 (0.0001)</td>
<td>0.0012 (0.0002)</td>
</tr>
<tr>
<td>$\Psi_{21}$</td>
<td>0.0002 (0.0002)</td>
<td>0.0001 (0.0000)</td>
<td>0.0000 (0.0000)</td>
</tr>
<tr>
<td>$\Psi_{12}$</td>
<td>0.0014 (0.0007)</td>
<td>0.0005 (0.0002)</td>
<td>0.0011 (0.0003)</td>
</tr>
<tr>
<td>$\Psi_{22}$</td>
<td>0.0013 (0.0003)</td>
<td>0.0003 (0.0001)</td>
<td>0.0004 (0.0001)</td>
</tr>
<tr>
<td>$\mu^v$</td>
<td>-9.3528 (0.0787)</td>
<td>-9.1931 (0.0885)</td>
<td>-9.1919 (0.0575)</td>
</tr>
<tr>
<td>$\mu^d$</td>
<td>4.5818 (0.0579)</td>
<td>4.6221 (0.0625)</td>
<td>4.5343 (0.0830)</td>
</tr>
<tr>
<td>$\Sigma_{11}$</td>
<td>0.0006 (0.0001)</td>
<td>0.0004 (0.00005)</td>
<td>0.00142 (0.00025)</td>
</tr>
<tr>
<td>$\Sigma_{12}$</td>
<td>-2.29e-005 (7.17e-006)</td>
<td>1.66e-005 (1.02e-005)</td>
<td>-1.58e-005 (1.87e-005)</td>
</tr>
<tr>
<td>$\Sigma_{22}$</td>
<td>6.45e-006 (1.41e-006)</td>
<td>4.58e-006 (8.73e-007)</td>
<td>7.38e-006 (1.59e-006)</td>
</tr>
<tr>
<td>$\mu_{J,r}$</td>
<td>0.0406 (0.0399)</td>
<td>0.0086 (0.0142)</td>
<td></td>
</tr>
<tr>
<td>$\mu_{J,v}$</td>
<td>1.0130 (0.2044)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_{J,d}$</td>
<td>-0.0016 (0.0322)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{J1}$</td>
<td>0.0273 (0.0087)</td>
<td></td>
<td>0.0154 (0.00332)</td>
</tr>
<tr>
<td>$\Sigma_{J2}$</td>
<td>-0.0105 (0.0222)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{J3}$</td>
<td>0.2602 (0.1375)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{J4}$</td>
<td>0.0019 (0.0064)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{J5}$</td>
<td>0.0114 (0.0197)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_{J6}$</td>
<td>0.0283 (0.0104)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0001 (0.0000)</td>
<td>0.0005 (0.0002)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_m$</td>
<td>0.0093 (0.0003)</td>
<td>0.0092 (0.0003)</td>
<td>0.0092 (0.0003)</td>
</tr>
</tbody>
</table>

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

We use a QQ-plot of the return residuals to assess the specification of the models\(^2\). First, we plot the standardized unconditional returns in the top left panel in Figure 1.4. The distribution of unconditional returns is clearly not normal and it exhibits high kurtosis. The residuals in the SVSDCJ model are given by

$$r_{i+1}^a - \xi_{i+1}^y J_{i+1} - m_{i+1} + m_i$$

\(^2\)If Brownian motion is adequate to describe the price process, the standardized returns should be normally distributed with mean zero and variance one.
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 1.4, the SVSDCJ model shows no clear sign of misspecification. The bottom left panel in Figure 1.4 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 1.3: 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 nonnormalities in returns. These results are consistent with what Eraker, Johannes, and Polson (2003) find using daily level data.

The estimated conditional duration $\lambda_i$ is presented in Figure 1.5. 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. The SVSDCJ model has a less persistent conditional duration path (larger $\Psi_{22}$). The duration residuals in all three models exhibit underdispersion, possibly due to the aggregation of durations. The dispersion of duration residuals $\left(\frac{\text{std}(D_{t+1}/\lambda_i)}{\text{mean}(D_{t+1}/\lambda_i)}\right)$ equals 0.46, 0.46 and 0.45 for the SVSDCJ, SVSDJ and SVSD model respectively, indicating the need for a more flexible duration distribution.

The off diagonal elements of $\Psi$ and $\Sigma$ measure the interdependence between volatility and duration process. Our 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 means of $\Psi_{12}$ and $\Psi_{21}$ 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. $\Sigma_{12}$ is negative, hence the contemporaneous correlation between the two Brownian motion $W_t^\nu$ and $W_t^\lambda$ is negative. In other words, short conditional duration is accompanied by high
Figure 1.3: Estimated spot volatility. The estimate is obtained from the posterior mean (average of draws after burning in).

volatility.

Figure 1.6 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 $\gamma$ 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 a source for leverage effects. The estimated correlation is negative as indicated by the sign of $\Sigma_J(2,1)$, but it is not significantly
different from zero. Other correlations between jump sizes are not significant either. Considering that jumps are rare and latent, we would need larger samples to estimate the correlation.

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, i.e., \( \hat{IV}_t = \sum_{t \in day(t)} V_i D_{t+1} \). Bipower variation is constructed using five-minute returns, \( BV_t = \mu_1^{-2} \sum_{j=2}^{\lceil \Delta \rceil} |r_{t+j}\Delta||r_{t+(j-1)\Delta}| \).
where $\Delta = 1/78$ and $\mu_1 = \sqrt{2/\pi}$. The estimated integrated volatility and bipower variation are plotted in Figure 1.7. They show similar patterns, 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, 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 $\Delta$, 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 \( \hat{IJV}_t = \sum_{i \in \text{day}(t)} \xi_i^2 J_i \), if our model is correctly specified, \( \hat{IJV}_t \) and \( JV_t \) should converge to the same value. We plot \( \hat{IJV}_t \) and \( JV_t \) in each day in Figure 1.8. \( JV_t \) is a lot larger than \( \hat{IJV}_t \) in most days. This is not surprising since we look 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 1.9 depicts the logarithmic price in the day when \( JV_t \) 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 is no clear indication of large discrete price movement.

The proportion of variation due to jumps can be computed by

\[
\frac{\sum_t \hat{IJV}_t}{\sum_t \hat{IV}_t + \sum_t \hat{IJV}_t}.
\]
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.

1.6 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. Our model analyses the interdependence and utilizes this relationship to gain 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, 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 we use finer returns. Last, volatility and conditional durations are interdependent, consistent with what the Easley and O’Hara theory predicts.
Figure 1.9: 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.
Chapter 2

A Stochastic Price Duration Model for Estimating High-Frequency Volatility

2.1 Introduction

Volatility plays a pivotal role in modern day financial economics. Since asset prices are generally considered to be driven by Brownian motions, the natural way to estimate volatility is to look at how much price changes in a given time interval. In particular, if volatility is constant, the variance of the return is a linear function of volatility. The passage time theory for Brownian Motion provides another approach to estimate volatility: one can look at how long it takes for the price to change by a given amount. Let price duration refer to the waiting time for the logarithmic price to travel the distance \( \delta \). The expectation of the price duration is related to the inverse of volatility. Intuitively, if volatility is high, price will be changing quickly and the expected price duration will
be relatively short. While most volatility estimation methods are based on returns, this paper utilizes price durations to model high-frequency volatility.

The time-varying nature of volatility poses challenges to its estimation. Roughly speaking, there are three return-based methods to estimate volatility, namely GARCH, stochastic volatility (SV) and realized volatility (RV). GARCH-type models\(^1\) assume that volatility is some function of past returns. In the SV-type models, volatility is assumed to be random and to follow a stochastic process. The availability of high-frequency financial data has popularized the RV estimator, which uses returns sampled at shorter horizons (such as 5 minutes) to measure volatility at a longer horizon (such as a day). The RV approach assumes volatility to be stochastic without specifying any parametric form. In the frictionless arbitrage-free world, the sum of squared returns converges in probability to integrated volatility when the sampling frequency goes to infinity. However, since observed prices are contaminated by market microstructure noise, realized volatility is a biased estimator of the actual volatility, and the problem becomes more severe when sampling frequency increases. One solution is to sample sparsely\(^2\). The optimal sampling frequency can be determined by considering the trade-off between the bias induced by microstructure noise and the variance induced by decreasing the sampling frequency. In practice, 5-minute RV is commonly used.

The literature on duration-based volatility estimation is considerably smaller and most of the work employs the autoregressive conditional duration (ACD) model. Engle and Russell (1998) propose ACD to model the durations between trades. They also apply the model to price durations by treating the price arrival times as a point process, and

\(^1\)See Hansen and Lunde (2005) for a list of 330 specifications in the GARCH universe and their evaluation.

\(^2\)There are more sophisticated ways to deal with market microstructure noise, such as subsampling (see Zhang, Mykland, and Ait-Sahalia, 2005), pre-averaging (see Jacod, Li, Mykland, Podolskij, and Vetter, 2009) or realized kernels (see Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008).
link the price arriving intensity to volatility. ACD is similar to GARCH: the volatility traced out from price intensity is assumed to be deterministic. Tse and Yang (2012) adopt the augmented ACD specification to model price durations and estimate high-frequency volatility, which they call the ACD-ICV method. They find that ACD-ICV outperforms many version of RV methods in Monte Carlo exercises. Bauwens and Veredas (2004) allow conditional duration to be random and apply the model to trade durations, price durations and volume durations. Their approach is close to a SV model although they do not directly specify or measure volatility.

Cho and Frees (1988) are the first to use passage times of Brownian Motion to estimate volatility, assuming it is constant. Andersen, Dobrev, and Schaumburg (2009) introduce a family of nonparametric volatility estimation using different types of passages times. Their method is a natural dual approach to realized volatility: both are nonparametric, assume volatility is stochastic and focus on estimating integrated volatility over longer periods, usually a day. Their duration-based estimator is robust to jumps and compares favorably to many robust RV type estimators.

This paper proposes a class of stochastic price duration models to estimate high-frequency volatility parametrically. In the baseline model which we call SPD0, logarithmic volatility follows an Ornstein-Uhlenbeck(OU) process. The OU process is mean-reverting and it leads to an AR(1) process when discretized. The SPD0 model employs SV models directly in the domain of duration-based estimator.

Interesting extensions to the baseline model can be obtained by incorporating additional information. In particular, we consider trade durations. The asymmetric information models by Easley and O’Hara (1987) suggest that trades durations have an interdependent relationship with volatility. Specifically, since a short trade duration suggests information events and an increased number of informed traders, it tends to be followed
by high volatility. On the other hand, lack of trades, or long trade durations are associated with lack of information events and hence lower volatility. Empirical studies also support the impact of duration on volatility.

We model volatility and trade duration using the stochastic volatility and stochastic duration (SVSD) model in Pelletier and Zheng (2012) and Wei and Pelletier (2013). The logarithmic volatility and conditional duration are assumed to follow a bivariate OU process to accommodate their interdependence. We call this model SPD1.

A duration-based volatility estimator faces the same challenges from market microstructure noise. The solution is the same as the return-based methods: sample sparsely so the variance of microstructure noise is small compared to volatility. The difference is the sampling scheme: the return-based approach samples at calendar time (e.g. every 5 minutes) or tick time (e.g. every 100 trades), while the duration-based estimator samples at points when the logarithmic price crosses the given threshold; decreasing sampling frequency is achieved by increasing the threshold. In this paper, we choose threshold such that the number of sampling points is comparable to a 5-minute RV estimator.

The sampling scheme renders the first benefit of using price durations over returns: since high volatility results in short price duration, we are sampling more often when the spot volatility is high, and less often when spot volatility is low. Hence, the ratio of noise variance over the volatility integrated over the sampling period is kept relatively flat. Also, if one is interested in the integrated volatility over a day, more points in the realm of high volatility would provide a better approximation to the integration.

The second benefit of using price duration is that it is robust to the discreteness of price. In an ideal world, prices are observed continuously. In the real world, the minimal price change is determined by the tick size, which has been $0.01 since 2000. Cho and Frees (1988) compare the duration-based approach with the return-based approach in
the presence of price discreteness, and they show that low-priced stocks suffer the most from price discreteness and have the most to gain from using duration-based estimators. Intuitively, if the price of stock is $1, the smallest change of return one can observe is 1% while for a $100 stock, the smallest increment for return is 0.01%. Price discreteness results in zero returns and complicates estimation for high frequency volatility. Price duration is naturally robust to price discreteness and it is particular advantageous for low-priced stock.

The benefits of using a parametric approach are threefold. First, we can utilize the persistence of volatility. Second, we can estimate intraday spot volatility while the non-parametric approach usually focuses on integrated volatility in a day. Third, we can extend the model to incorporate additional information, such as trade durations.

The rest of this chapter is organized as follows: Section 2 describes the model specification. Section 3 discusses the estimation procedure and conducts simulation studies. Section 4 presents empirical results. Section 5 concludes.

2.2 Model Specification

2.2.1 Stochastic Price Duration

We start by assuming that the logarithmic asset price $y_t$ solves the following stochastic differential equations:

$$
dy_t = \sqrt{V_t} dW_t^y$$
$$
d\log V_t = -\kappa^v (\log V_t - \mu^v) dt + \sigma^v dW_t^v, \tag{2.1}$$

36
where \( V_t \) is the latent instantaneous variance. \( W_t^y \) and \( W_t^v \) denote standard Brownian motions. For simplicity, we assume that \( W_t^y \) and \( W_t^v \) are independent, i.e., there is no leverage effect. From equation (2.1), we know that the logarithmic volatility follows the OU process:

\[
\log V_t = (1 - e^{-\kappa^v(t-s)})\mu^v + e^{-\kappa^v(t-s)}\log V_s + \sigma^v \int_s^t e^{-\kappa^v(t-s)} dW^v_s, \tag{2.2}
\]

where \( t > s \). The long-run mean of this process equals \( \mu^v \). The parameter \( \kappa^v \) and \( \sigma^v \) describe the persistence and variability of the process, respectively. The long-run variance of logarithmic volatility is given by \( \sigma^v_2/2\kappa^v \).

We use price durations to discretize the above process. Price duration is the time it takes for \( y_t \) to change by a given amount \( \delta \), also called the price threshold. Specifically, if \( \tau_{i+1} \) is the \( i+1 \)th price duration, \( \tau_{i+1} = \inf\{t > 0| |y_{t_i+t} - y_{t_i}| \geq \delta\} \), where \( t_i \) denotes the time when \( y_t \) crosses the threshold for the \( i \)th time. The sequence \( \{t_i\}_{i=0}^N \) partitions the time line \([t_0, t_N]\) into \( N \) intervals, while each interval corresponds to a price duration, i.e., \( \tau_{i+1} = t_{i+1} - t_i \).

To obtain the distribution of \( \tau_{i+1} \), we assume that volatility is constant within each price duration. In other words, we approximate volatility by a piecewise constant process, while the instantaneous volatility in the interval \([t_i, t_{i+1}]\) equals to \( V_i \), the volatility at the left end point of the interval. From the passage time theory for Brownian motion, \( \tau_{i+1} \) can be written as a function of the price threshold \( \delta \) and \( V_i \) multiplied by a random variable \( \eta_{i+1} \), see for example Andersen, Dobrev, and Schaumburg (2009). Specifically:

\[
\tau_{i+1} = \frac{\delta^2}{V_i} \eta_{i+1}. \tag{2.3}
\]
The random variable $\eta$ is the price duration when volatility and price threshold are both equal to 1. In passage time theory, $\eta$ is also referred to as the first exit time, since it measures the time it takes for a standard Brownian motion to exit the band $[-1, 1]$. The distribution of $\eta$ is given by

$$p(\eta) = \sum_{k=-\infty}^{\infty} \frac{2(1 + 4k)}{\sqrt{2\pi \eta^{3/2}}} e^{-\frac{(1+4k)^2}{2\eta}}. \quad (2.4)$$

From equation (2.2), we get the discretized logarithmic volatility using price durations:

$$\log V_{i+1} = \left(1 - e^{-\kappa \tau_{i+1}}\right) \mu^v + e^{(-\kappa \tau_{i+1})} \log V_i + u_{i+1}^v \quad (2.5)$$

where

$$u_{i+1}^v \sim N \left(0, \frac{\sigma^2_v}{2\kappa^v} (1 - e^{(-2\kappa \tau_{i+1})}) \right).$$

Equation (2.3) and (2.5) form the discretized baseline model SPD0. It is a non-linear non-Gaussian state space model where (2.3) is the observation equation and (2.5) is the evolution equation. In the baseline model, we do not consider information from other observables such as number of trades and volume in each price duration. The number of trades is particularly interesting since it reveals the trade durations, which is interdependent with volatility as suggested by the market microstructure theory. We introduce trade durations in the next subsection.
2.2.2 Stochastic Trade Duration

The trade duration $D_{j+1}$ is defined as the time interval between a trade that occurred at $t_j$ and the next trade at $t_{j+1}$. Let $\lambda_{t_j}$ denote the conditional expectation of $D_{j+1}$ given the information set available at $t_j$, $E(D_{j+1}|I_{t_j}) = \lambda_{t_j}$. We assume that trade durations are exponentially distributed given the conditional duration $\lambda_{t_j}$. Hence, $D_{j+1}$ is equal to $\lambda_{t_j}$ multiplied by an i.i.d random variable with exponential distribution, i.e., $D_{j+1} = \lambda_{t_j} e_{j+1}$. The conditional duration $\lambda_t$ can vary over time and gives rise to interesting dynamics in trade durations.

Suppose that $N$ trades happened in a time interval with length $\tau$, and we are interested in the distribution of $\tau$ given $N$ and $\lambda_t$\(^3\). For simplicity, we assume that the conditional duration $\lambda_t$ is constant within the time interval. In this case, each trade duration follows an exponential distribution with scale parameter $\lambda_t$, and $\tau$ is the sum of $N$ exponentially distributed variables. The distribution of $\tau$ is given by a gamma distribution with shape parameter $N$ and scale parameter $\lambda_t$, $\tau \sim \text{Gamma}(N, \lambda_t \tau)$. We can also look at the distribution of the average duration, $d_a = \frac{\tau}{N}$. Applying the change of variable formula we have $d_a \sim \text{Gamma}(N, \lambda_t/N)$. We can use the scaling property of the gamma distribution to write $d_a$ as $\lambda_t$ multiplied by a random variable with a $\text{Gamma}(N, 1/N)$ distribution.

In general, if we observe $N_{i+1}$ trades in the time interval $[t_i, t_{i+1}]$ with $\tau_{i+1} = t_{i+1} - t_i$, the average trade duration $d_{i+1}^a = \frac{\tau_{i+1}}{N_{i+1}}$ can be written as

$$d_{i+1}^a = \lambda_i e_{i+1}, \quad (2.6)$$

where $e_{i+1} \sim \text{Gamma}(N_{i+1}, 1/N_{i+1})$, and $\lambda_i$ is the conditional duration at the beginning of the interval. It is easily seen that $E(d_{i+1}^a|I_{t_i}) = \lambda_i$.

\(^3\)We can also use the distribution of $N$ given $\tau$. 

39
2.2.3 Modeling Price Durations and Trade Durations Jointly

To create persistence and interdependence between volatility and trade duration, we model the logarithm of $\lambda_t$ and $V_t$ using a bivariate OU process (see Wei and Pelletier, 2013, for properties of this process). Let $X_t = (\log(V_t), \log(\lambda_t))^\prime$, $X_t$ solves:

$$dX_t = -\Psi(X_t - \mu^x)dt + S_x dW^x_t,$$

(2.7)

where $\Psi$ is a $2 \times 2$ matrix that measures the mean reversion and dependence between conditional duration and volatility. The process mean reverts to $\mu^x$, the diffusive long-run mean. $S_x$ measures the variation of the logarithmic volatility and the logarithmic duration, and $S_x = \text{diag}(\sigma_v, \sigma_\lambda)$. $W^x_t$ is a Brownian motion in $\mathcal{R}^2$ with $dW^v_t dW^\lambda_t = \rho dt$, where $\rho$ 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 observables for this model are price duration $\{\tau_i\}_{i=1}^N$ and average duration $\{d^a_i\}_{i=1}^N$. We discretized the bivariate OU process using price durations. As before, we assume that volatility and conditional durations are constant within each price duration. The discretized model SPD1 is also a non-linear and non-Gaussian state space model and it is given by the observation equations

$$\tau_{i+1} = \frac{\delta^2}{V_i} \eta_{i+1},$$

$$d^a_{i+1} = \lambda_i e_{i+1},$$

(2.8)
and the evolution equation

\[ x_{i+1} = (I_2 - e^{-\Psi_{\tau_i+1}})\mu x + e^{-\Psi_{\tau_i+1}} x_i + u_{i+1}, \]

\[ u_{i+1} \sim N(0, \Sigma_{i+1}) \]

\[ \text{vec}(\Sigma_{i+1}) = (\Psi \oplus \Psi)^{-1}(I_2 - e^{-(\Psi \oplus \Psi)\tau_{i+1}})\text{vec}(\Sigma_x). \] (2.9)

### 2.3 Estimation Procedure and Simulation Studies

#### 2.3.1 Linear State Space Representation

The inference for models with stochastic volatility or stochastic conditional duration is nontrivial since the evaluation of the likelihood involves integrating out the latent variables. To avoid high dimensional integration, we adopt the quasi-maximum likelihood estimation (QMLE) method that is popular in return-based SV models (see e.g. Harvey, Ruiz, and Shephard, 1994 and Ruiz, 1994). The idea of QMLE is to approximate the non-linear non-gaussian state space model by a linear and gaussian one, and use the Kalman filter to obtain the likelihood. There also exists inference methods that evaluate the exact likelihood, such as simulated maximum likelihood (Danielsson, 1994) or Markov Chain Monte Carlo (Jacquier, Polson, and Rossi, 1994 and Kim, Shephard, and Chib, 1998). However, these methods are computationally intensive, and it is difficult to estimate data in a long period of time given the sample size of high-frequency data. Also, as we will demonstrate later, the approximation error in the QMLE method is less severe in duration-based models than return-based models.

To apply QMLE to the baseline model SPD0, we start by taking the logarithm of
and approximate \( \log \eta_{i+1} \) by a normally distributed variable that has the same mean and variance. Equation (2.10) and (2.5) form the linear state space representation for SPD0, so we can use the Kalman filter to get parameter estimates and smoothed volatility estimates. See de Jong (1989) for the filtering and smoothing procedure with time-varying coefficients.

Parameter estimates yielded by QMLE are consistent and asymptotically normally distributed. The efficiency of the estimator depends on the approximation error; if the true distribution is far from normal, the estimator could be highly inefficient. Return-based estimation requires approximating the logarithm of a chi-squared distribution by a normal distribution, whereas our model approximates the logarithm of price durations as normal. Figure (2.1) plots the true distribution versus a normal distribution with the same mean and variance for both logarithmic squared returns and logarithmic price durations. As can be seen, the logarithm of price duration is better approximated by the normal distribution\(^4\). Hence, for our duration-based models, we gain computational speed from using QMLE without much loss of efficiency.

If the asset prices have a jump component, as suggested by much empirical work in the literature, the true distribution of price durations would differ. However, price durations have some natural robustness to jumps as demonstrated by Andersen, Dobrev, and Schaumburg (2009) and Tse and Yang (2012). Jumps in the price process might shorten the price duration, but the amount by which the price exceeds the threshold does

\(^4\)This feature is also shared by range-base estimators, see Alizadeh, Brandt, and Diebold (2002).
Figure 2.1: PDF of the true distributions versus their normal approximations. The left panel plots the distribution of logarithmic price durations versus a normal distribution with the same mean and variance. The right panel plots the distribution of squared returns versus its normal approximation.

not directly impact the estimation. Another complication comes from time discreteness: we do not observe price continuously in time, so the actual price change is usually slightly larger than the price threshold $\delta$. This issue can be mitigated if we replace $\delta$ by the average actual price change in the MLE. We leave the exact solution to these issues to future work.

We estimate the SPD1 model using QMLE as well. To linearize the average trade durations, we take the logarithm of equation (2.6) and approximate $\log e_{i+1}$ by a normal distribution. Since $e_{i+1}$ is distributed as $\text{Gamma}(N_{i+1}, 1/N_{i+1})$, the mean and variance of $\log e_{i+1}$ are given by $\psi(N_{i+1}) - \log(N_{i+1})$ and $\psi_1(N_{i+1})$ respectively, where $\psi(x)$ denotes the digamma function and $\psi_1(x)$ denotes the trigamma function. Finally, we have

$$
\begin{pmatrix}
\log \tau_{i+1} \\
\log a_{i+1}^a
\end{pmatrix}
= 
\begin{pmatrix}
2 \log \delta + E(\log \eta) \\
E(\log e_{i+1})
\end{pmatrix}
+ 
\begin{pmatrix}
-1 & 0 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
x_i \\
w_{i+1}
\end{pmatrix},
$$

(2.11)
where
\[
    w_{i+1} \sim N \left( 0, \begin{pmatrix} \text{Var} (\log \eta) & 0 \\ 0 & \psi_1 (N_{i+1}) \end{pmatrix} \right).
\]

Equation (2.11) and (2.9) form the linear state space representation of \( SPD_1 \) model.

An important issue in applications is to infer the stochastic volatility. We obtain volatility estimates from the smoothed latent variables. Let \( x_{i|N} \) and \( P_{i|N} \) denote the projection of \( x_i \) on all observations and its mean squared error, i.e., \( x_{i|N} = E(x_i|\mathcal{F}_N) \) and \( P_{i|N} = \text{MSE}(x_{i|N}) \), the smoothed estimate for \( V_i \) is obtained from the upper left element of \( \exp (x_{i|N} + P_{i|N}/2) \).

### 2.3.2 Simulations without Microstructure Noise

We perform simulation studies to illustrate the potential gain from using additional information from trade durations (or loss from not using trade durations). We generate log-arithmic price and trade durations assuming that conditional duration and volatility are interdependent. Specifically, we use the following parameter value: \( (\Psi_{11}, \Psi_{12}, \Psi_{21}, \Psi_{22}) = (0.01, 0.01, 0.02, 0.03), \mu_x = (-18.8, 0.5)' \) and \( (\sigma_v, \sigma_\lambda, \rho) = (0.026, 0.088, -0.5) \). The parameters are chosen such that the annualized volatility is targeted at 20%, and trades happen every 1.6 seconds on average. We then obtain price durations and average trade durations by setting the price threshold to 0.001, which corresponds to approximately 0.1% change in the price.

Figure 2.2 presents an example of the true spot volatility and its estimates. Several comments can be made regarding this figure. First, we are estimating more points when volatility is high, and less points when volatility is low. Second, the volatility estimated
from both SPD0 and SPD1 models are able to capture the main dynamics in spot volatility. Third, by utilizing trade durations, the SPD1 model outperforms SPD0 in the sense that estimated volatility from SPD1 tracks the true volatility more closely. We use the root mean squared error (RMSE) to quantity the difference. The RMSE for each model is computed by $	ext{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (V_i - \hat{V}_i)^2}{N}}$, where $V_i$ and $\hat{V}_i$ denote the true and estimated spot variance at time $t_i$, respectively. The RMSE from the baseline model SPD0 is 32% higher than the RMSE from the SPD1 model.

We also plot the true logarithmic conditional duration versus its estimates from SPD1 model in Figure 2.3. The estimates trace changes in the true conditional duration although they do not fully capture the rapid fluctuations since we are using average trade durations.
2.3.3 Simulation with Microstructure Noise

We compare the performance of SPD0 and SPD1 models to the popular realized volatility approach in the presence of market microstructure noises. Suppose that the observed logarithmic price is contaminated by i.i.d. microstructure noises,

\[ y_i^o = y_i + m_i, \tag{2.12} \]

where \( m_i \sim N(0, \sigma_m^2) \). We assume that the true prices are generated from the same stochastic process as in the last subsection. The observed \( y_i^o \) are generated with a different Noise-to-Signal Ratio (NSR). Here we define NSR = \( \sigma_m/V_{\text{mean}} \), where \( V_{\text{mean}} \) is the long run mean of spot volatility. We set NSR to (0.25, 1, 1.5), representing low, median and high noise levels.
We conduct 200 simulations, while each simulation consists of data that represents one trading day (6.5 hours or 23,400 seconds). Realized volatility is then computed by the sum of squared returns within a day, \( RV_t = \sum_{j=1}^{1/\Delta} r_{t+j\Delta}^2 \). Theoretically, \( RV_t \) converges in probability to the integrated volatility over the day, \( IV_t \), when sampling frequency goes to infinity. In the Monte Carlo experiment, we choose the sampling frequency according to the noise levels. Specifically, we sample every 3, 4 or 5 minutes for the low, median or high noise levels. For price durations, we choose price threshold such that the duration-based approach has the same number of observations as the realized volatility approach. In other words, the average price duration is calibrated to 3, 4 or 5 minutes for the low, median or high noise levels. We then estimate the spot volatility from SPD0 and SPD1 model, and compute the integrated volatility in a given day by \( \hat{IV}_t = \sum_{i\in\text{day}(t)} V_i \tau_{i+1} \). We use RMSE to compare the performance of the estimated \( IV \), with RMSE = \( \sqrt{\sum_{t=1}^{T} (\hat{IV}_t - IV_t)^2 / T} \).

Table 2.1 reports the Monte Carlo results for the estimated IV from the SPD1 model, the SPD0 model, and the RV approach. It can be seen that both SPD models outperforms the realized volatility across different noise levels. Also the SPD1 models performs better than the SPD0 model, and the gain increases when NSR increases. This is as expected since SPD0 does not utilize trade durations, and when NSR increases, prices are more contaminated while trade durations are not affected.

\[ \text{We could also use RMSE to compare the spot volatility estimates between the SPD0 and the SPD1 model as in the previous section, but the RV approach does not provide spot volatility estimates.} \]
Table 2.1: Monte Carlo results for the estimated IV

<table>
<thead>
<tr>
<th>NSR = 0.25</th>
<th>ME</th>
<th>SE</th>
<th>RMSE</th>
<th>Relative RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPD1</td>
<td>-2.88E-06</td>
<td>4.11E-05</td>
<td>1.97E-05</td>
</tr>
<tr>
<td></td>
<td>SPD0</td>
<td>1.58E-06</td>
<td>4.09E-05</td>
<td>2.01E-05</td>
</tr>
<tr>
<td></td>
<td>RV</td>
<td>1.71E-06</td>
<td>5.46E-05</td>
<td>3.65E-05</td>
</tr>
<tr>
<td>NSR = 1</td>
<td>SPD1</td>
<td>1.30E-05</td>
<td>4.45E-05</td>
<td>2.56E-05</td>
</tr>
<tr>
<td></td>
<td>SPD0</td>
<td>1.87E-05</td>
<td>4.48E-05</td>
<td>2.98E-05</td>
</tr>
<tr>
<td></td>
<td>RV</td>
<td>4.32E-06</td>
<td>5.69E-05</td>
<td>4.20E-05</td>
</tr>
<tr>
<td>NSR = 1.5</td>
<td>SPD1</td>
<td>2.60E-05</td>
<td>4.87E-05</td>
<td>3.69E-05</td>
</tr>
<tr>
<td></td>
<td>SPD0</td>
<td>3.34E-05</td>
<td>4.98E-05</td>
<td>4.33E-05</td>
</tr>
<tr>
<td></td>
<td>RV</td>
<td>-1.07E-06</td>
<td>5.79E-05</td>
<td>4.55E-05</td>
</tr>
</tbody>
</table>

Notes: ME = mean error, SE = standard deviation of sample estimates. RMSE = root mean squared error. The last column express the RMSE from the SPD1 and SPD0 model as a percentage of the RMSE from RV. We conduct 200 Monte Carlo simulations, while each simulation corresponds one trading day.
2.4 Empirical Results

2.4.1 Data

We apply our model to the milli-second time stamped IBM trade data in the US Equity Data provided by TickData. The sample period is August and September 2011 (44 trading days). We follow the cleaning procedure proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) to filter out the potentially erroneous data. First, entries with a correction indicator other than 0 are deleted. Second, we delete entries with abnormal sales conditions (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 within 0.1 second as one observation and use the mean price to alleviate possible measurement error in the transaction time.

To obtain price durations, we set the price threshold to 0.002 (roughly a 0.2% change in price) so that average price duration is roughly 5 minutes. This results in a total of 3,361 sampling points. Figure 2.4 illustrates an example of the sampling points in a day. As we can see, the sampling points are more concentrated near the beginning, when price is changing violently. We then divide each price duration by the number of trades within that price duration to obtain average trade durations. Summary statistics for the observables is given in Table 2.2. Since trades are occurring frequently (every 1.25 seconds on average), the impact from time discreteness is minimal.
Figure 2.4: Sampling points in one trading day. We sample when the change in the logarithmic price reaches or exceeds 2%.

Table 2.2: Summary statistics for IBM in 2011/08/01-2011/09/31

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price Duration</td>
<td>299</td>
<td>144</td>
<td>465</td>
</tr>
<tr>
<td>Number of Trades per price duration</td>
<td>209</td>
<td>132</td>
<td>245</td>
</tr>
<tr>
<td>Average Trade Duration</td>
<td>1.25</td>
<td>1.07</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Notes: All units reported are in seconds.
2.4.2 Diurnal Pattern

Intraday volatility and duration have well known diurnal patterns. Transactions happen more frequently near the opening and closing times, and less frequently during the middle of a day. This deterministic diurnal pattern needs to be accounted for before we specify a stochastic model for the latent variables.

We use a quadratic function\(^6\) to approximate the diurnal pattern and estimate it within the model. The level of the quadratic function is fixed by setting its minimum to 1, otherwise the mean of the latent process becomes unidentifiable. For the \(SPD1\) model (the procedure for \(SPD0\) model naturally follows), we adopt the following quadratic functions for volatility and conditional duration:

\[
g_v(t) = a_1(t + a_2)^2 + 1,
\]
\[
g_d(t) = a_3(t + a_4)^2 + 1,
\]

Letting \(V^*_i\) and \(\lambda^*_i\) denote the deseasonalized volatility and conditional duration, we have

\[
V_i = V^*_i g_v(t_i),
\]
\[
\lambda_i = \frac{\lambda^*_i}{g_d(t_i)}.
\] (2.14)

This specification produces the U-shaped pattern in volatility and the inverse U-shaped pattern in the conditional duration. After considering the diurnal effect, the observation

---

\(^6\)The choice of a quadratic function is a trade-off between better approximation and less parameters to estimate. The nonparametric estimate in Chapter 1 indicates that a quadratic function describes the main dynamics of the diurnal pattern. Higher order approximation may improve the fit, and we leave that to future work.
equation for SPD1 model becomes

\[
\begin{pmatrix}
\log \tau_{i+1} \\
\log d^a_{i+1}
\end{pmatrix}
= \begin{pmatrix}
2 \log \delta + E(\log \eta) - \log g_v(t_i) \\
E(\log e_{i+1}) - \log g_d(t_i)
\end{pmatrix} + \begin{pmatrix}
-1 & 0 \\
0 & 1
\end{pmatrix} x_i + w_{i+1},
\]

(2.15)

where \( x_i = (\log V^*_i, \log \lambda^*_i)' \) and it follows the evolution equation (2.9).

### 2.4.3 Estimation

We estimate the data in the sample period using both the \( SPD0 \) and \( SPD1 \) models. To deal with observations from different trading days, we assume that each day “starts fresh”: the latent OU process starts with its long-run mean and variance each day. The parameter estimates from the \( SPD0 \) and \( SPD1 \) models are presented in Table 2.3. In the \( SPD0 \) model, the parameter estimates indicate an annualized volatility of 25%. In the \( SPD1 \) model, the market microstructure theory from Easley and O’Hara (1987) predicts that high volatility leads to short durations, while short durations have a positive effect on volatility. In our estimate, the impact of volatility on conditional duration is profound, while the effect of conditional duration on volatility and their instantaneous correlation is not statistically significant.

We plot the diurnal patterns estimated from the SPD1 model in Figure 2.5. The diurnal pattern indicates that the volatility near the beginning of a trading day is almost 5 times as big as its minimum around noon. The conditional trading durations are less than half as long as the conditional trading durations near the middle of the day.

We compare the parametric \( SPD \) models to the nonparametric RV approach as well. Figure 2.6 presents the daily integrated volatility estimated from the 5-minute RV, \( SPD0 \) model and \( SPD1 \) model. The integrated volatility in the parametric models is obtained
Table 2.3: Parameter estimates for the SPD1 and SPD0 model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SPD1</th>
<th>std. error</th>
<th>SPD0</th>
<th>std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Psi_{11}$</td>
<td>3.91E-04</td>
<td>6.58E-03</td>
<td>4.09E-04</td>
<td>8.38E-05</td>
</tr>
<tr>
<td>$\Psi_{21}$</td>
<td>0.049</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Psi_{12}$</td>
<td>1.11E-10</td>
<td>1.51E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Psi_{22}$</td>
<td>0.115</td>
<td>0.040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu^v$</td>
<td>-19.293</td>
<td>0.162</td>
<td>-19.103</td>
<td>0.094</td>
</tr>
<tr>
<td>$\mu^d$</td>
<td>0.911</td>
<td>0.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^v$</td>
<td>0.025</td>
<td>0.005</td>
<td>0.026</td>
<td>0.003</td>
</tr>
<tr>
<td>$\sigma^d$</td>
<td>0.129</td>
<td>0.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.546</td>
<td>0.561</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>4.14E-08</td>
<td>1.13E-08</td>
<td>2.28E-08</td>
<td>3.99E-09</td>
</tr>
<tr>
<td>$a_2$</td>
<td>-1.07E+04</td>
<td>4.93E+02</td>
<td>-1.29E+04</td>
<td>5.17E+02</td>
</tr>
<tr>
<td>$a_3$</td>
<td>1.56E-08</td>
<td>1.98E-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td>-1.03E+04</td>
<td>3.10E+02</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: We estimate the $SPD1$ and $SPD0$ model using milli-second IBM data in 2011/08/01-2011/09/30.

from the smoothed estimates of spot volatility, $\hat{IV}_t = \sum_{i \in day(t)} V_i \tau_{i+1}$. As we can see, the three volatility estimates trace each other quite closely.

### 2.5 Conclusion

In this paper we present a new parametric model to estimate stochastic volatility based on price durations. This model has several advantages: first, price durations have some robustness to jumps and market microstructure noise, especially the noise from price discreteness. Second, we utilize the persistence of volatility and we can infer volatility integrated over any period of time. Third, contrary to the ACD-type models, we assume that volatility is stochastic, and we obtain the distribution of price durations from the passage theory for Brownian motions. Last, we can conduct inference easily using QMLE.
Figure 2.5: Diurnal patterns in volatility and conditional duration.

Figure 2.6: Daily volatility of IBM in 2011/08/01-2011/09/30, 44 days.
without much loss of efficiency since the logarithmic price duration is better approximated by a normal distribution than the logarithmic squared returns.

We also extend the baseline model $SPD0$ to incorporate information from trading durations, as market microstructure theory suggests that trading durations and volatility are interdependent. We call the more sophisticated model $SPD1$. We conduct Monte Carlo studies to demonstrate the performance of the price duration models. We find that $SPD1$ outperforms $SPD0$ in estimating spot volatility, and both duration-based model performs better than realized volatility in estimating the integrated volatility.

There are several interesting extensions we can explore in this class of price duration models. First, volume is another variable that could influence volatility, and hence could be incorporated into the latent process. Second, we can consider the distribution of price duration if the asset price follows a jump diffusion. Last, we can further investigate the influence of time discreteness on the distribution of price durations, especially for less liquid stocks.
Chapter 3

Backtesting Value at Risk using a Parametric Duration Model

3.1 Introduction

The importance of risk management has long been recognized by financial market participants. One essential element of managing risk is measuring risk. The 1996 Market Risk Amendment to the Basel Accord established Value-at-Risk (VaR) as the basis for determining market risk capital requirements. Since then, VaR has become a standard tool to measure risk (See Jorion, 2006 and Berkowitz and O’Brien, 2002). VaR is the maximum expected loss for a given time horizon and confidence level. For example, a one-day VaR with coverage rate 5% is the value such that the loss in the next day would be smaller than this value with a 95% probability. In other words, the probability that the loss exceeds VaR is 5%.

VaR summarizes risk in a single number, and it is easily communicated among financial institutions and their regulators. Statistically, VaR is a quantile of the conditional
distribution of returns. Despite the simple concept, the estimation of VaR is a challenging problem, because the conditional distribution of returns is generally unknown. Although several methods have been developed to estimate VaR, arguably the most popular approach is Historical Simulation. Pérgnon and Smith (2010) find that 73% of banks that disclosed their VaR method are using Historical Simulation. Historical Simulation approximates the conditional distribution of returns using a rolling window with typically one or two years of past returns. It is nonparametric and easy to implement, hence favored by practitioners, but it relies on unrealistic assumptions; it assumes that returns are independent and identically distributed. It ignores stylized facts such as volatility clustering and leverage effects, and the time-varying dynamics of returns is only accounted for by the rolling window. Moreover, Historical Simulation is under-responsive to changes in conditional risk; see Pritsker (2006).

Given the popularity of VaR and the lack of effective implementation, evaluating the performance of the VaR measure is of great importance. In practice, the evaluation of VaR is usually carried out through backtesting, which compares the VaR forecasts with realized returns. If the ex-post loss exceeds the ex-ante VaR forecast, it is referred to as a violation. Define a hit sequence, \( \{ I_t \} \), where \( I_t = 1 \) indicates there is a violation; if the VaR measure with coverage rate \( p \) is correctly specified, the hit sequence must be i.i.d. Bernoulli with parameter \( p \). Christoffersen (1998) builds a Likelihood Ratio (LR) framework for the conditional evaluation of an interval forecast such as VaR. He first formulates a LR test for the unconditional coverage of VaR, which amounts to testing if the average number of violations corresponds to its expected value, i.e, if \( E(I_t) = p \). The unconditional coverage test does not take into consideration higher-order dynamics: it assumes that the observations are independent. Violations could have correct nominal coverage while exhibiting time dependence, in particular violation clustering. The cluster-
ing in violations indicates closely grouped large losses and misspecified VaR. To examine the independence hypothesis, Christoffersen (1998) specifies a first-order Markov chain alternative for the hit sequence. Finally, he develops the test for correct conditional coverage, which is a joint test of the unconditional coverage hypothesis and the independence hypothesis.

The First-order Markov chain alternative has limited power against general forms of time-dependence in violations. Christoffersen and Pelletier (2004) develop a duration-based approach for backtesting VaR. The intuition is that if VaR is correctly specified with coverage rate $p$, the hit sequence should be i.i.d. Bernoulli with parameter $p$, and the duration between hits should have no memory and mean equal to $1/p$. The distribution of durations under the null hypothesis is approximated by the exponential distribution since it is the only continuous distribution with constant hazard rate. For the alternative hypothesis they considered a Weibull distribution with a decreasing hazard rate. Their test can also be decomposed into an unconditional coverage test and an independence test, where the unconditional coverage test checks if the mean of durations is equal to $1/p$, and the independence test checks if the hazard rate is constant. They also considered an autoregressive model for the expected conditional duration.

It is also possible to specify a discrete distribution for the durations. Haas (2005) finds that discrete distributions have better power against violation clustering than continuous distributions. Candelon, Colletaz, Hurlin, and Tokpavi (2011) propose a GMM test for duration-based backtesting. They find that discrete distributions perform as well as continuous distributions within the GMM approach. Berkowitz, Christoffersen, and Pelletier (2011) implement a discrete duration test under the LR framework, which they refer to as the geometric test. Under the null hypothesis that durations have no memory, discrete durations follow the geometric distribution, hence the name geometric test. Monte Carlo
simulations suggest that the geometric test is always more powerful than the continuous distribution based Weibull test.

Engle and Manganelli (2004) argue that requiring the hit sequence to be i.i.d. is a necessary but not sufficient condition for a correctly specified VaR; if the VaR forecast is a valid quantile measure, the expectation of hit sequence conditional on information set at time $t \minus{} 1$ must equal to the coverage rate. In other words, the violation $I_t$ should be unbiased and it should be uncorrelated with any information up to $t \minus{} 1$. They propose a dynamic quantile (DQ) test for backtesting VaR. In particular, they regress the hit sequence on a set of explanatory variables that includes the VaR forecasts and the first four lagged hits. Dumitrescu, Hurlin, and Pham (2012) extend this approach to a dynamic binary choice model which allows for non-linear dependence between the probability of violations and the explanatory variables. Gaglianone, Lima, Linton, and Smith (2011) develop evaluation methods based on quantile regressions and it has better small sample property than the DQ test.

Berkowitz, Christoffersen, and Pelletier (2011) provide a unified framework for backtesting VaR by noting that violations form a martingale difference sequence. They compare the power of existing backtesting methods using data generating processes that resemble actual Profits and Losses (P/L) from four business lines. They find that the DQ test performs the best overall but that the Geometric test also performs well in many cases.

In this paper, we propose a new duration-based test by incorporating VaR forecasts in the geometric test, hereafter geometric-VaR test. The insight is twofold: first, duration-based approaches can capture general forms of time-dependence in violations. Specifically, we choose the geometric test over the Weibull test because discrete distributions have better power against violation clustering. Second, if VaR is not correctly specified, the
probability of observing a violation would depend on past information, including the VaR forecast. Hence, specifying VaR in the distribution of duration should improve the power of a duration-based test.

The geometric-VaR test can be decomposed into three individual tests: the first test focuses on correct unconditional coverage; the second test considers the dependence structure in durations; the third test examines whether the probability of getting a violation depends on the VaR forecasts. We obtained the same dataset as Berkowitz, Christoffersen, and Pelletier (2011): desk-level daily P/L and VaR forecasts from four separate business lines of one large commercial bank. The actual P/Ls provide guidance for choosing realistic data generating processes to compare the power of the geometric-VaR test and the related duration-based tests. We find that the geometric-VaR test has better power than other duration-based tests, and it has power against various forms of misspecifications. We assess the performance of the actual VaR forecasts provided by the bank using geometric-VaR test and its component tests. Our framework not only tests whether the VaR forecast is misspecified, but also helps understand how the VaR forecast is misspecified by examining the individual hypotheses separately.

The rest of the paper is organized as follows. Section 2 reviews the geometric test and presents the geometric-VaR test. Section 3 discusses how to implement the tests. Section 4 compares the power of the newly proposed test with other duration-based tests. Section 5 applies the geometric-VaR test to the actual P/L and VaR forecasts provided by the bank. Section 6 concludes.
3.2 Duration-based Backtesting

VaR is defined with a promised coverage rate over a given time horizon. We focus on the one-day VaR horizon in the paper. A one-day VaR with coverage rate $p$ is the value such that the loss next day would exceed VaR with probability $p$. In particular, we say that the VaR forecast $VaR_t(p)$ is efficient with respect to information set $\Omega_{t-1}$ if

$$Pr(r_t < -VaR_t(p)|\Omega_{t-1}) = p.$$  \hfill (3.1)

Here we follow the convention of reporting VaR as a positive number. Given a VaR forecast $VaR_t$ conditional on information up to time $t-1$ and the realized return $r_t$ at time $t$, we can obtain the hit sequence $\{I_t\}$ by comparing the ex-post return $r_t$ and the ex-ante forecast $VaR_t$. A violation or a hit refers to the event that the loss exceeds the VaR forecast. Let $I_t$ be an indicator function such that $I_t = 1$ when there is a violation, that is,

$$I_t = \begin{cases} 
1, & \text{if } r_t < -VaR_t(p) \\
0, & \text{otherwise}
\end{cases}.$$  \hfill (3.2)

If the VaR forecast is efficient with respect to information set $\Omega_{t-1}$, the conditional distribution of the hit sequence should be i.i.d. Bernoulli with parameter $p$:

$$Pr(I_t = 1|\Omega_{t-1}) = p.$$  \hfill (3.3)

Let $t_i$ denote the day of the $i$th violation/hit, the no-hit duration $D_i$ is constructed by $D_i = t_i - t_{i-1}$. The hit sequence is transformed into the duration sequence and we can use duration modelling techniques to explore the data.\footnote{See Kiefer (1988) for an extensive review on duration modelling.}
If the hit sequence is i.i.d. Bernoulli, $D_i$ measures the number of Bernoulli trials needed to get one hit. Hence, under the null hypothesis that VaR is correctly specified, $D_i$ follows a geometric distribution with parameter $p$:

$$Pr(D_i = d) = p(1 - p)^{d-1}. \quad (3.4)$$

The geometric distribution is characterized by a flat hazard function. Hazard function $\lambda^i_d$ is defined as the ratio of pdf $f^i(d)$ over survival function $S^i(d)$,

$$\lambda^i_d = \frac{f^i(d)}{S^i(d)}. \quad (3.5)$$

For a discrete distribution, we can write

$$f^i(d) = Pr(D_i = d) = \lambda^i_d(1 - \lambda^i_{d-1})...(1 - \lambda^i_1),$$

$$S^i(d) = Pr(D_i \geq d) = (1 - \lambda^i_d)(1 - \lambda^i_{d-1})...(1 - \lambda^i_1), \quad (3.6)$$

where $f^i(d)$ is the probability of $D_i$ equaling $d$; Survival function $S^i(d)$ is the probability of duration $D_i$ being at least $d$. Hence, $\lambda_d$ measures the probability of getting a hit (a failure) on day $d$ given that the no-hit duration has survived for $d - 1$ days. In other words,

$$\lambda^i_d = Pr(I_{t_i+d} = 1|I_{t_i+d-1} = 0, ..., I_{t_i+1} = 0, I_{t_i} = 1, \Omega_{t_i+d-1}). \quad (3.7)$$

If the VaR forecast is efficient, the probability of getting a hit does not depend on any past information, so the hazard function must be a constant. Furthermore, if the VaR forecast has correct unconditional coverage, the hazard function must equal the coverage rate $p$. 

62
On the other hand, if VaR is not efficient with respect to past information, the hazard function is no longer the constant $p$. The probability of getting a hit can depend on how long the no-hit sequence has lasted and the VaR forecast. Berkowitz, Christoffersen, and Pelletier (2011) propose the geometric test in which the hazard function decreases with duration under the alternative hypothesis. We extend the geometric test by allowing the hazard function to depend on VaR forecasts as well as durations. We review the Geometric test in Section 2.1 and present the Geometric-VaR test in Section 2.2.

### 3.2.1 Geometric Test

The Geometric test allows time dependence in the hit sequence by specifying the following hazard function for durations:

$$\lambda_d = ad^{b-1},$$

with $0 \leq a < 1$ and $0 \leq b \leq 1$.\(^2\) The ordering of duration does not play a role here so we omit the superscript $i$.

We can decompose the geometric test into a test of unconditional coverage and a test of duration independence. The null hypothesis of correct unconditional coverage corresponds to $a = p$, since the percentage of violations equals the coverage rate $p$ and the average duration equals $1/p$. As in Kupiec (1995) and Christoffersen (1998), we conduct the unconditional coverage test assuming independence. However, if violations are clustered together, the risk of bankruptcy would increase even if violations have correct unconditional coverage. Hence, it is important to examine the time dependence of vio-

\(^2\)This corresponds to the discrete Weibull distribution specified in Stein and Dattero (1984). Note that if $b > 1$, the distribution has finite support.
lations. Under the null hypothesis, duration has “no memory,” and the hazard function is flat, i.e. $b = 1$. Under the alternative hypothesis, violations are clustered, we would observe an excessive number of short durations and long durations compared to the geometric distribution. This corresponds to a decreasing hazard function\(^3\): if the no-hit duration has not lasted for long, the probability of getting a hit is high, hence the excessive amount of short durations; if the no-hit duration has survived for a long time, the probability of getting a hit is low, hence the excessive amount of long durations. In other words, the alternative hypothesis is specified as $b < 1$. By testing the unconditional coverage and duration independence hypotheses jointly, we obtain the geometric conditional coverage test.

### 3.2.2 Geometric-VaR Test

If the VaR forecast $VaR_t$ is not efficient with respect to the information up to time $t - 1$, the conditional probability of getting a violation can depend on how long the no-hit duration has lasted as well as the VaR forecast. We utilize the hazard function to specify the dependence structure. Specifically,

$$
\lambda_d^t = ad^{b-1}e^{-cVaR_t+d},
$$

(3.9)

with $0 \leq a < 1$, $0 \leq b \leq 1$ and $c \geq 0$. Under the null hypothesis that VaR is correctly specified, durations follow a geometric distribution with parameter $p$, so the null corresponds to $a = p$, $b = 1$ and $c = 0$.

As with the geometric test, the parameter $a$ in the hazard rate specifies the unconditional coverage. The second part in the hazard function, $d^{b-1}$, describes duration

\(^3\)Decreasing hazard function is also referred to as negative duration dependence.
dependence or the time-dependence in violations. Under the alternative hypothesis that violations are clustered, we would observe negative duration dependence, and \( b < 1 \). The last part in the hazard function examines the impact of \( \text{VaR}_{t_i+d} \) on the probability of getting a hit. Under the alternative hypothesis, the probability of getting a hit becomes lower if the forecasted \( \text{VaR}_{t_i+d} \) takes a large absolute value, so we have \( c > 0 \).

The geometric-VaR test can be decomposed into three tests of individual hypotheses under one unified framework. Specifically, we consider a test of unconditional coverage, a test of duration independence and a test of VaR independence. Combining the unconditional coverage test and the duration independence test we have the geometric test. For comparison, we also consider a joint test of unconditional coverage and VaR independence. In summary, we will compare three individual tests and three joint tests:

1. Unconditional Coverage Test (under the maintained assumption that \( b = 1 \) and \( c = 0 \)):

   \[
   H_0 : a = p \\
   H_a : a \neq p
   \]

2. Duration independence test (under the maintained assumption that \( c = 0 \)):

   \[
   H_0 : b = 1 \\
   H_a : b \neq 1
   \]
3. VaR independence test:

\[ H_0 : c = 0 \]
\[ H_a : c \neq 0 \]

4. Geometric test: unconditional coverage and duration independence (under the maintained assumption that \( c = 0 \))

\[ H_0 : a = p \text{ and } b = 1 \]
\[ H_a : a \neq p \text{ or } b \neq 1 \]

5. VaR test: unconditional coverage and VaR independence (under the maintained assumption that \( b = 0 \))

\[ H_0 : a = p \text{ and } c = 0 \]
\[ H_a : a \neq p \text{ or } c \neq 0 \]

6. Geometric-VaR test: unconditional coverage, duration independence and VaR independence

\[ H_0 : a = p, b = 1 \text{ and } c = 0 \]
\[ H_a : a \neq p \text{ or } b \neq 1 \text{ or } c \neq 0 \]

This unified framework allows us to not only test whether VaR forecasts are misspecified overall, but also understand how they are misspecified by looking into which individual
hypothesis is rejected. The three individual hypotheses have different economic impacts: financial institutions focus on the unconditional coverage since regulators use the number of violations to determine the penalties that banks might incur (see Annex 10a in Basel Committee on Banking Supervision, 2006). The duration independence test considers whether VaR is capturing the time-varying nature of risk, in particular closely grouped high risk since that might increase the probability of bankruptcy. The rejection of VaR independence hypothesis indicates that the VaR forecast does not fully represent the return dynamics. Note that banks use an internal VaR model to determine the capital requirements. From the practitioners’ point of view, they face a trade-off between correctly specified VaR and smooth capital requirements. Examining the three hypotheses separately can help internal model builders to decide the trade-off.

3.3 Test Implementation

A hit sequence does not usually start with a violation. In that case, the first duration measures the number of days with no violation rather than the number of days between two violations. In other words, the first duration is left-censored. Similarly, if the hit sequence does not end with a violation, the last duration is right-censored. To implement the test, we generate a binary series \( \{C_i\}_{i=1}^N \) along with the duration series \( \{D_i\}_{i=1}^N \), while \( C_i = 1 \) indicates that \( D_i \) is censored. If a duration \( D_i \) is not censored, its contribution to the likelihood is the probability \( f^i(D_i) \). On the other hand, if a duration is censored or incomplete, we only know that the duration has lasted for at least \( D_i \) days. Hence, its contribution to the likelihood is the survival function \( S^i(D_i) \). When the hit sequence is converted to the duration sequence, only the first and the last duration could be censored.
The combined log-likelihood function is given by

$$
\log L(D | \Theta) = C_1 \log S^1(D_1) + (1 - C_1) \log f^1(D_1) + \sum_{i=2}^{N-1} \log f^i(D_i) \\
+ C_N \log S^N(D_N) + (1 - C_N) \log f^N(D_N).
$$

We estimate $\Theta$ using maximum likelihood. Following Christoffersen (1998), we utilize likelihood ratio tests so that the individual hypothesis testing and joint hypothesis testing can be conveniently implemented under a unified framework. The standard LR statistics can be formulated by $LR = -2(\log L(D | \hat{\Theta}_R) - \log L(D | \hat{\Theta}_{UR}))$ for each of the tests in section 2, where $\hat{\Theta}_R$ denotes the maximum likelihood estimate of $\Theta$ when parameters are restricted by the null hypothesis, and $\hat{\Theta}_{UR}$ is the unrestricted maximum likelihood estimate (although the parameter space could be restricted by the maintained assumptions). Specifically, the LR test statistic for the unconditional coverage test is given by

$$
LR^{UC} = -2 \left[ \log L(D | a = p, b = 1, c = 0) - \log L(D | \hat{a}, b = 1, c = 0) \right]. \quad (3.10)
$$

For the test of duration independence, we have

$$
LR^{Dind} = -2 \left[ \log L(D | \hat{a}, b = 1, c = 0) - \log L(D | \hat{a}, \hat{b}, c = 0) \right]. \quad (3.11)
$$

The duration independence test does not depend on the true coverage $p$; it captures the time-dependence of violations while maintaining the assumption of VaR independence (by imposing $c = 0$). Next, for the VaR independence test, the LR test statistic is formulated after taking duration dependence into consideration. In other words, the VaR independence test captures whether the probability of violation depends on the VaR
forecast when we allow for the time-dependence of violations. Specifically,

$$LR^{Vind} = -2 \left[ \log L(D|\hat{a}, \hat{b}, \hat{c}) - \log L(D|\hat{a}, \hat{b}, c = 0) \right].$$ \hspace{1cm} (3.12)$$

The geometric test jointly tests the unconditional coverage and duration independence; the test statistic is given by:

$$LR^G = -2 \left[ \log L(D|\hat{a}, \hat{b}, c = 0) - \log L(D|a = p, b = 1, c = 0) \right].$$ \hspace{1cm} (3.13)$$

Note that $LR^G$ equals the sum of $LR^{UC}$ and $LR^{Dind}$. Next, the VaR test jointly tests the unconditional coverage and VaR independence. Since duration dependence and VaR dependence might be capturing similar dynamics in the data, we assume duration independence ($b = 1$) when forming the test statistic:

$$LR^V = -2 \left[ \log L(D|\hat{a}, b = 1, \hat{c}) - \log L(D|a = p, b = 1, c = 0) \right].$$ \hspace{1cm} (3.14)$$

Last, for the geometric-VaR test,

$$LR^{GV} = -2 \left[ \log L(D|\hat{a}, \hat{b}, \hat{c}) - \log L(D|a = p, b = 1, c = 0) \right].$$ \hspace{1cm} (3.15)$$

The geometric-VaR test can be decomposed into three individual tests: the test of unconditional coverage, the test of duration independence and the test of VaR independence. In particular, $LR^{GV}$ is equal to the sum of the three test statistics:

$$LR^{GV} = LR^{UC} + LR^{Dind} + LR^{Vind}.$$ \hspace{1cm} (3.16)$$
Once we obtain the test statistics from the sample, we could use their asymptotic distribution to calculate the p-value. However, the sample size of the duration sequence is usually small. For example, if we have one year of VaR forecast with a 1% coverage rate, the expected number of durations is 2.5. Even with 10 years of data, the average sample size for durations is 25. The small sample size also results in a nontrivial sample selection issue since we cannot run the backtest unless a minimum number of violations are observed. Another complication is that we are testing parameter values at the boundary of the parameter space and this might affect the asymptotic distribution. Also, since we are working with binary data, the distribution of LR statistics is not continuous. Hence, we use Monte Carlo techniques in Dufour (2006) to get the simulated distribution of the test statistics and reliable p-values, as in Christoffersen and Pelletier (2004).

To implement the Monte Carlo technique, we first generate $N$ realizations of data under the null hypothesis. For each test described above, we obtain the test statistics $LR_i$ for the $i$th realization of data. Then $LR_i, i = 1, \ldots, N$, form the simulated distribution of test statistics under the null. Let $LR_0$ denote the test statistics from the sample. We can obtain p-values by comparing $LR_0$ to its simulated distribution under the null.

Since we are working with discrete distributions, $LR_0$ can be equal to $LR_i$. To break the ties, we draw $U_i$ from a uniform distribution on $[0, 1]$ for each of test statistic, $LR_i$. Then the Monte Carlo p-value is given by

$$\hat{p}_N(LR_0) = \frac{N \hat{G}(LR_0) + 1}{N + 1},$$

(3.17)

and

$$\hat{G}_N(LR_0) = 1 - \frac{1}{N} \sum_{i=1}^{N} I(LR_i \leq LR_0) + \frac{1}{N} \sum_{i=1}^{N} I(LR_i = LR_0) I(U_i \geq U_0).$$
3.4 Simulation Studies

3.4.1 Data Generating Process

The most prominent characteristics of financial returns include volatility clustering, leverage effects and fat tails. Volatility clustering refers to the observation that volatility is persistent: periods of high volatility tend to cluster together. Leverage effect concerns the asymmetric response to innovations. In financial data, a negative innovation tend to have different impacts on future volatility than a positive innovation. In particular, a negative shock tends to increase future volatility more than a positive shock of the same size. Fat tails explain the large returns that would be very unlikely given a normal distribution.

The most popular approach characterizing these features is to use GARCH-type models, which allow volatility to depend on past returns and other observables. We adopt the Nonlinear asymmetric GARCH, or NGARCH process with student-t innovations. The NGARCH model accounts for leverage effects by allowing the impact of a negative shock to depend on the past conditional volatility (See Engle and Ng, 1993). Specifically, we simulate from the following process:

\[ R_{t+1} = \sigma_{t+1} \left( \frac{(d-2)}{d} \right)^{1/2} z_{t+1} \]

\[ \sigma_{t+1}^2 = \omega + \alpha \sigma_t^2 \left( \left( \frac{(d-2)}{d} \right) z_t - \theta \right)^2 + \beta \sigma_t^2, \]  

(3.18)

where \( z_t \) is drawn from a student \( t(d) \) distribution. The leverage effect is represented by \( \theta \).

If \( \theta \) is positive, volatility tends to increase more after a large negative shock than a large positive shock. The unconditional variance of returns is given by \( \omega^{-1} (1 - \alpha (1 + \theta^2) - \beta) \), while \( (\alpha (1 + \theta^2) + \beta) \) indicates the volatility persistence.

To choose realistic parameters for their simulations, Berkowitz, Christoffersen, and
Table 3.1: Parameter Estimates of NGARCH-t(d) Model for Four Business Lines

<table>
<thead>
<tr>
<th></th>
<th>Business Line 1</th>
<th>Business Line 2</th>
<th>Business Line 3</th>
<th>Business Line 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>3.808</td>
<td>3.318</td>
<td>6.912</td>
<td>4.702</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.245</td>
<td>0.503</td>
<td>-0.962</td>
<td>0.093</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.749</td>
<td>0.928</td>
<td>0.873</td>
<td>0.915</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.155</td>
<td>0.052</td>
<td>0.026</td>
<td>0.072</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.550</td>
<td>0.215</td>
<td>0.213</td>
<td>1.653</td>
</tr>
</tbody>
</table>

Pelletier (2011) estimate the volatility model in 3.18 with desk level P/Ls from four business lines of a large commercial bank. Table 3.1 reproduces their parameter estimates. The four business lines display very different dynamics. The negative $\theta$ in business line 1 and 3 indicates that positive returns have a larger impact on volatility than negative returns, the opposite of leverage effects. Business line 2 is characterized by highly persistent volatility and fat tails. Business line 4 has very high unconditional volatility. Using the parameter estimates in Table 3.1, we can generate returns that have similar dynamics to the P/L series from the four business lines.

### 3.4.2 Backtesting VaR

We compute VaR forecasts using the Historical Simulation approach. In general, VaR with coverage rate $p$ is the $p$th quantile of the conditional distribution of return $R_{t+1}$:

$$VaR_{t+1}(p) = F_{t+1}^{-1}(p),$$

(3.19)

where $F_{t+1}$ denotes the conditional distribution of $R_{t+1}$. The Historical Simulation approach uses the past one year or two years of returns to construct an empirical estimate
of $F_{t+1}$:

$$\text{VaR}_{t+1}^{\text{HS}}(p) = \text{percentile}(\{R_s\}_{s=t-T_e+1}^t, p),$$  \hspace{1cm} (3.20)

where $T_e$ is the size of the rolling window that is used to approximate the conditional distribution. Following industry practice, we choose $T_e$ to be 250, which roughly corresponds to one year of trading days.

Historical Simulation is widely adopted by financial institutions because it is nonparametric and easy to implement. However, this approach is problematic since it assumes that returns from the past $T_e$ days are independently and identically distributed. It does not take into consideration the predictability of volatility; the time varying nature of volatility is only reflected through the rolling window. Moreover, Pritsker (2006) points out that Historical Simulation is under-responsive to changes in risk. For example, suppose that the market crashed yesterday. The VaR forecast for today should reflect this large increase in risk. However, when computing the 1% VaR with Historical Simulation the magnitude of this very negative return does not directly impact the forecast. A good backtesting procedure should be able to detect the inadequacy of Historical Simulation when the volatility is changing.

We generate P/L from (3.18) using the parameters estimated from the four business lines, then we compute VaR forecasts using (3.20). By comparing the \textit{ex-post} return with the \textit{ex-ante} VaR we obtain the hit sequence for backtesting. After the test statistics $LR_0$ are computed for each test in Section 2, we can compare $LR_0$ to its distribution under the null hypothesis to get the p-value. As previously mentioned, the small sample size and discrete nature of durations means the asymptotic distribution may not be good approximations. Instead, we use the Monte Carlo technique described in Section 3 to
obtain a simulated distribution of test statistics under the null. Specifically, we generate the hit sequence according a Bernoulli distribution, and generate VaR from a GARCH process, and then we compute \( \text{LR}_i \), \( i = 1, \ldots, N \) for each test. In this setting, VaR is time varying and it does not provide information on the probability of getting a hit. We choose \( N \) to be 9,999 for the simulated distribution. The p-value is obtained by comparing \( \text{LR}_0 \) to \( \text{LR}_i \), \( i = 1, \ldots, N \) using equation (3.17).

### 3.4.3 Finite Sample Power of the tests

We assess the power of each test by computing the rejection frequency from 5000 replications of the backtesting procedure described in the previous section. The significance level is chosen to be 10\% for all tests, i.e., a hypothesis is rejected if the p-value computed using the Monte Carlo technique is smaller than 10\%. The sample sizes vary from 250 to 1500, which roughly correspond to one year through six years of trading days. We conduct the power exercise using both 1\% and 5\% VaR and the simulated power of each test is reported in Table 3.2 and Table 3.3. The different Business lines in Table 3.2 and Table 3.3 indicate that the simulated P/Ls mimic the dynamics of that business line, i.e., the P/Ls are generated using parameters estimated from the corresponding business line.

The power of the unconditional coverage test is generally low for all business lines since the unconditional coverage test does not consider higher order dynamics of violations. Also, we expect the unconditional coverage test based on duration to have similar performance as the unconditional test of Kupiec which compares the actual number of violations with the expected number (See Kupiec, 1995). The Kupiec test has been documented to have low power, see for example Pritsker (2006); Pérignon and Smith (2008).
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>UC</th>
<th>Duration ind.</th>
<th>VaR ind.</th>
<th>Geometric</th>
<th>VaR</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1% VaR in Four Business Lines</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Line 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.117</td>
<td>0.390</td>
<td>0.049</td>
<td>0.357</td>
<td>0.127</td>
<td>0.384</td>
</tr>
<tr>
<td>500</td>
<td>0.069</td>
<td>0.443</td>
<td>0.151</td>
<td>0.234</td>
<td>0.083</td>
<td>0.370</td>
</tr>
<tr>
<td>750</td>
<td>0.043</td>
<td>0.526</td>
<td>0.222</td>
<td>0.321</td>
<td>0.105</td>
<td>0.340</td>
</tr>
<tr>
<td>1000</td>
<td>0.027</td>
<td>0.569</td>
<td>0.292</td>
<td>0.331</td>
<td>0.138</td>
<td>0.365</td>
</tr>
<tr>
<td>1250</td>
<td>0.017</td>
<td>0.625</td>
<td>0.351</td>
<td>0.385</td>
<td>0.214</td>
<td>0.502</td>
</tr>
<tr>
<td>1500</td>
<td>0.021</td>
<td>0.661</td>
<td>0.422</td>
<td>0.422</td>
<td>0.333</td>
<td>0.598</td>
</tr>
<tr>
<td>Business Line 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.200</td>
<td>0.337</td>
<td>0.047</td>
<td>0.350</td>
<td>0.194</td>
<td>0.386</td>
</tr>
<tr>
<td>500</td>
<td>0.107</td>
<td>0.446</td>
<td>0.138</td>
<td>0.249</td>
<td>0.118</td>
<td>0.392</td>
</tr>
<tr>
<td>750</td>
<td>0.069</td>
<td>0.568</td>
<td>0.197</td>
<td>0.364</td>
<td>0.134</td>
<td>0.383</td>
</tr>
<tr>
<td>1000</td>
<td>0.070</td>
<td>0.612</td>
<td>0.246</td>
<td>0.377</td>
<td>0.162</td>
<td>0.413</td>
</tr>
<tr>
<td>1250</td>
<td>0.073</td>
<td>0.684</td>
<td>0.274</td>
<td>0.444</td>
<td>0.215</td>
<td>0.536</td>
</tr>
<tr>
<td>1500</td>
<td>0.060</td>
<td>0.721</td>
<td>0.340</td>
<td>0.491</td>
<td>0.300</td>
<td>0.621</td>
</tr>
<tr>
<td>Business Line 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.082</td>
<td>0.140</td>
<td>0.067</td>
<td>0.123</td>
<td>0.065</td>
<td>0.113</td>
</tr>
<tr>
<td>500</td>
<td>0.048</td>
<td>0.150</td>
<td>0.203</td>
<td>0.068</td>
<td>0.052</td>
<td>0.133</td>
</tr>
<tr>
<td>750</td>
<td>0.029</td>
<td>0.161</td>
<td>0.329</td>
<td>0.060</td>
<td>0.053</td>
<td>0.079</td>
</tr>
<tr>
<td>1000</td>
<td>0.017</td>
<td>0.140</td>
<td>0.465</td>
<td>0.040</td>
<td>0.063</td>
<td>0.060</td>
</tr>
<tr>
<td>1250</td>
<td>0.011</td>
<td>0.138</td>
<td>0.587</td>
<td>0.038</td>
<td>0.109</td>
<td>0.112</td>
</tr>
<tr>
<td>1500</td>
<td>0.012</td>
<td>0.138</td>
<td>0.679</td>
<td>0.041</td>
<td>0.212</td>
<td>0.184</td>
</tr>
<tr>
<td>Business Line 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.214</td>
<td>0.382</td>
<td>0.010</td>
<td>0.391</td>
<td>0.183</td>
<td>0.430</td>
</tr>
<tr>
<td>500</td>
<td>0.115</td>
<td>0.505</td>
<td>0.044</td>
<td>0.293</td>
<td>0.103</td>
<td>0.401</td>
</tr>
<tr>
<td>750</td>
<td>0.081</td>
<td>0.621</td>
<td>0.068</td>
<td>0.415</td>
<td>0.100</td>
<td>0.400</td>
</tr>
<tr>
<td>1000</td>
<td>0.081</td>
<td>0.685</td>
<td>0.087</td>
<td>0.447</td>
<td>0.102</td>
<td>0.425</td>
</tr>
<tr>
<td>1250</td>
<td>0.080</td>
<td>0.739</td>
<td>0.105</td>
<td>0.512</td>
<td>0.125</td>
<td>0.527</td>
</tr>
<tr>
<td>1500</td>
<td>0.080</td>
<td>0.789</td>
<td>0.124</td>
<td>0.577</td>
<td>0.163</td>
<td>0.597</td>
</tr>
</tbody>
</table>

Notes: We simulate P/Ls using NGARCH-t(d) models that have the same parameters as the four business line and then compute VaR using Historical Simulation with a rolling window of size 250. The simulated power of each test is the rejection frequency from 5000 replications. UC stands for unconditional coverage test and GV stands for geometric-VaR test. See the text for details on each test.
Table 3.3: Power of 10% Duration-based Tests on 5% VaR in Four Business Lines

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>UC</th>
<th>Duration ind.</th>
<th>VaR ind.</th>
<th>Geometric</th>
<th>VaR</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business Line 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.154</td>
<td>0.479</td>
<td>0.130</td>
<td>0.360</td>
<td>0.178</td>
<td>0.365</td>
</tr>
<tr>
<td>500</td>
<td>0.062</td>
<td>0.668</td>
<td>0.210</td>
<td>0.486</td>
<td>0.183</td>
<td>0.529</td>
</tr>
<tr>
<td>750</td>
<td>0.026</td>
<td>0.780</td>
<td>0.333</td>
<td>0.585</td>
<td>0.282</td>
<td>0.674</td>
</tr>
<tr>
<td>1000</td>
<td>0.018</td>
<td>0.852</td>
<td>0.440</td>
<td>0.692</td>
<td>0.406</td>
<td>0.796</td>
</tr>
<tr>
<td>1250</td>
<td>0.010</td>
<td>0.901</td>
<td>0.533</td>
<td>0.762</td>
<td>0.564</td>
<td>0.884</td>
</tr>
<tr>
<td>1500</td>
<td>0.009</td>
<td>0.927</td>
<td>0.627</td>
<td>0.819</td>
<td>0.691</td>
<td>0.935</td>
</tr>
<tr>
<td><strong>Business Line 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.334</td>
<td>0.451</td>
<td>0.118</td>
<td>0.496</td>
<td>0.353</td>
<td>0.492</td>
</tr>
<tr>
<td>500</td>
<td>0.240</td>
<td>0.707</td>
<td>0.174</td>
<td>0.625</td>
<td>0.370</td>
<td>0.655</td>
</tr>
<tr>
<td>750</td>
<td>0.170</td>
<td>0.823</td>
<td>0.248</td>
<td>0.734</td>
<td>0.443</td>
<td>0.788</td>
</tr>
<tr>
<td>1000</td>
<td>0.139</td>
<td>0.891</td>
<td>0.328</td>
<td>0.796</td>
<td>0.531</td>
<td>0.855</td>
</tr>
<tr>
<td>1250</td>
<td>0.123</td>
<td>0.936</td>
<td>0.355</td>
<td>0.863</td>
<td>0.618</td>
<td>0.920</td>
</tr>
<tr>
<td>1500</td>
<td>0.120</td>
<td>0.963</td>
<td>0.407</td>
<td>0.907</td>
<td>0.698</td>
<td>0.954</td>
</tr>
<tr>
<td><strong>Business Line 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.061</td>
<td>0.136</td>
<td>0.159</td>
<td>0.078</td>
<td>0.069</td>
<td>0.083</td>
</tr>
<tr>
<td>500</td>
<td>0.017</td>
<td>0.144</td>
<td>0.366</td>
<td>0.045</td>
<td>0.052</td>
<td>0.075</td>
</tr>
<tr>
<td>750</td>
<td>0.005</td>
<td>0.157</td>
<td>0.626</td>
<td>0.046</td>
<td>0.084</td>
<td>0.102</td>
</tr>
<tr>
<td>1000</td>
<td>0.003</td>
<td>0.158</td>
<td>0.799</td>
<td>0.045</td>
<td>0.171</td>
<td>0.157</td>
</tr>
<tr>
<td>1250</td>
<td>0.001</td>
<td>0.160</td>
<td>0.895</td>
<td>0.049</td>
<td>0.334</td>
<td>0.267</td>
</tr>
<tr>
<td>1500</td>
<td>0.002</td>
<td>0.162</td>
<td>0.932</td>
<td>0.051</td>
<td>0.490</td>
<td>0.388</td>
</tr>
<tr>
<td><strong>Business Line 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>0.328</td>
<td>0.507</td>
<td>0.036</td>
<td>0.506</td>
<td>0.322</td>
<td>0.492</td>
</tr>
<tr>
<td>500</td>
<td>0.220</td>
<td>0.754</td>
<td>0.048</td>
<td>0.664</td>
<td>0.273</td>
<td>0.670</td>
</tr>
<tr>
<td>750</td>
<td>0.146</td>
<td>0.868</td>
<td>0.068</td>
<td>0.763</td>
<td>0.290</td>
<td>0.780</td>
</tr>
<tr>
<td>1000</td>
<td>0.124</td>
<td>0.931</td>
<td>0.083</td>
<td>0.851</td>
<td>0.335</td>
<td>0.864</td>
</tr>
<tr>
<td>1250</td>
<td>0.105</td>
<td>0.958</td>
<td>0.108</td>
<td>0.909</td>
<td>0.423</td>
<td>0.925</td>
</tr>
<tr>
<td>1500</td>
<td>0.102</td>
<td>0.980</td>
<td>0.134</td>
<td>0.939</td>
<td>0.503</td>
<td>0.957</td>
</tr>
</tbody>
</table>

Notes: We simulate P/Ls using NGARCH-t(d) models that have the same parameters as the four business line and then compute VaR using Historical Simulation with a rolling window of size 250. The simulated power of each test is the rejection frequency from 5000 replications. UC stands for unconditional coverage test and GV stands for geometric-VaR test. See the text for details on each test.
and Berkowitz, Christoffersen, and Pelletier (2011). Moreover, Escanciano and Pei (2012) show that the unconditional test is always inconsistent for backtesting Historical Simulation models, and the problem is less severe when the volatility is more persistent. This is consistent with what we find in business line 1 and 3; in both cases the negative $\theta$ and relatively small volatility persistence lead to less dependent left tail and hence the very low power in unconditional coverage test.

The duration independence test performs well expect for Business line 3. Business Line 3 is characterized by a large negative $\theta$, so a negative shock tends to be followed by low volatility, and violation clustering is less likely to happen. Since the duration independence test deals with the time dependence of violations, it is not surprising that the test has low power against this line of specification (opposite leverage effects), while the VaR independence test still has good power properties in this business line. The VaR independence test picks up the dependence between violation and VaR forecast after the time dependence is accounted for. Its power is lower than the duration independence test in business lines 1, 2 and 4. The duration independence test has the highest power in the these business lines, but it does not consider unconditional coverage, which determines the penalty for banks under current regulation.

The geometric test is a combination of the unconditional coverage test and duration independence tests; its likelihood ratio test statistic is the sum of the two individual test, and its power is closely related to the power of the two individual tests. The VaR test jointly tests the unconditional coverage hypothesis and VaR independence hypothesis while assuming duration independence. The VaR test has lower power than the geometric test in business line 1, 2 and 4 but it has better power against the misspecifications in business line 3.

Finally, the geometric-VaR test has the highest power in most cases when uncondi-
tional coverage is included. It combines the strength of three individual hypothesis tests, namely the unconditional coverage test and duration independence test and the VaR independence test, and it has good power properties against various types of misspecification.

To conduct the geometric-VaR test, we need to have more than three durations, or three durations with at least one of them uncensored. Roughly speaking, this selection criterion corresponds to having at least three violations. The VaR independence test also requires the estimation of all three parameters hence it has the same feasible fraction as the geometric-VaR test. The duration independence test, geometric test and VaR test all require the estimation of two parameters, hence we need three durations, or two durations with at least one of them uncensored. Table 3.4 reports the fraction of samples where the test is feasible.

3.5 Empirical Results

We apply our tests to the actual business line P/Ls and VaRs from Berkowitz, Christoffersen, and Pelletier (2011) to assess the performance of the VaR forecasts provided by the bank. The left panel of Figure 3.1 displays the time series of P/Ls and VaR forecasts for the four business lines. The right panel of Figure 3.1 displays the time series of violations, while the magnitude is the loss on the day of violation. Table 3.5 reports descriptive statistics from the four business lines. Business Line 3 only has one violation and none of the duration-based testing could be conducted.

The test statistics are computed using equations (3.10) through (3.15) and the test values are reported in Table 3.6. To get reliable p-values, we utilize the Monte Carlo

---

4Unless both the first and the last hit sequence is a violation, in which case three violations corresponds to two uncensored durations.
<table>
<thead>
<tr>
<th>VaR coverage</th>
<th>Sample Size</th>
<th>Geometric</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td>%1</td>
<td>250</td>
<td>0.7117</td>
<td>0.4708</td>
</tr>
<tr>
<td>%1</td>
<td>500</td>
<td>0.9796</td>
<td>0.9205</td>
</tr>
<tr>
<td>%1</td>
<td>750</td>
<td>0.9996</td>
<td>0.9982</td>
</tr>
<tr>
<td>%5</td>
<td>250</td>
<td>0.9988</td>
<td>0.9964</td>
</tr>
<tr>
<td>%1</td>
<td>250</td>
<td>0.6817</td>
<td>0.4822</td>
</tr>
<tr>
<td>%1</td>
<td>500</td>
<td>0.9682</td>
<td>0.9093</td>
</tr>
<tr>
<td>%1</td>
<td>750</td>
<td>0.9986</td>
<td>0.9907</td>
</tr>
<tr>
<td>%5</td>
<td>250</td>
<td>0.9866</td>
<td>0.9724</td>
</tr>
<tr>
<td>%1</td>
<td>250</td>
<td>0.7403</td>
<td>0.4272</td>
</tr>
<tr>
<td>%1</td>
<td>500</td>
<td>0.9917</td>
<td>0.9459</td>
</tr>
<tr>
<td>%1</td>
<td>750</td>
<td>1.0000</td>
<td>0.9994</td>
</tr>
<tr>
<td>%5</td>
<td>250</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>%1</td>
<td>250</td>
<td>0.6962</td>
<td>0.4946</td>
</tr>
<tr>
<td>%1</td>
<td>500</td>
<td>0.9690</td>
<td>0.9088</td>
</tr>
<tr>
<td>%1</td>
<td>750</td>
<td>0.9986</td>
<td>0.9936</td>
</tr>
<tr>
<td>%5</td>
<td>250</td>
<td>0.9878</td>
<td>0.9764</td>
</tr>
</tbody>
</table>

Notes: We report the fraction of simulations where we can compute geometric and geometric-VaR (GV) test.

<table>
<thead>
<tr>
<th>Business Line 1</th>
<th>Business Line 2</th>
<th>Business Line 3</th>
<th>Business Line 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P/Ls</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of OBS</td>
<td>873</td>
<td>811</td>
<td>623</td>
</tr>
<tr>
<td>Mean</td>
<td>0.1922</td>
<td>1.5578</td>
<td>1.8740</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>2.6777</td>
<td>5.2536</td>
<td>1.6706</td>
</tr>
<tr>
<td><strong>VaRs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of OBS</td>
<td>873</td>
<td>811</td>
<td>623</td>
</tr>
<tr>
<td>Mean</td>
<td>-7.2822</td>
<td>-16.3449</td>
<td>-3.2922</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>3.1321</td>
<td>10.5446</td>
<td>1.1901</td>
</tr>
<tr>
<td>Observed number</td>
<td>9</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>of hits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expected number</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 3.5: P/Ls and VaRs for Four Business Lines: Descriptive Statistics
Figure 3.1: Time Series of P/Ls, VaRs and Violations from Four Business Lines. The left panel displays P/Ls (dashed lines) and one-day 1% VaRs (solid lines). The right panel displays violations, while the magnitude is the P/L on the day of violation.
<table>
<thead>
<tr>
<th>Business Line 1</th>
<th>UC</th>
<th>Duration ind.</th>
<th>VaR ind.</th>
<th>Geometric</th>
<th>VaR</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Value</td>
<td>0.062</td>
<td>1.236</td>
<td>0.529</td>
<td>1.297</td>
<td>1.498</td>
<td>1.826</td>
</tr>
<tr>
<td>p-value</td>
<td>0.796</td>
<td><strong>0.077</strong></td>
<td>0.204</td>
<td>0.370</td>
<td>0.375</td>
<td>0.386</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Business Line 2</th>
<th>UC</th>
<th>Duration ind.</th>
<th>VaR ind.</th>
<th>Geometric</th>
<th>VaR</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Value</td>
<td>2.576</td>
<td>1.277</td>
<td>3.771</td>
<td>3.853</td>
<td>7.624</td>
<td>7.624</td>
</tr>
<tr>
<td>p-value</td>
<td>0.110</td>
<td><strong>0.075</strong></td>
<td><strong>0.003</strong></td>
<td>0.119</td>
<td><strong>0.012</strong></td>
<td><strong>0.011</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Business Line 4</th>
<th>UC</th>
<th>Duration ind.</th>
<th>VaR ind.</th>
<th>Geometric</th>
<th>VaR</th>
<th>GV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Value</td>
<td>2.082</td>
<td>2.930</td>
<td>0.000</td>
<td>5.012</td>
<td>2.334</td>
<td>5.012</td>
</tr>
<tr>
<td>p-value</td>
<td>0.160</td>
<td><strong>0.021</strong></td>
<td>0.497</td>
<td><strong>0.059</strong></td>
<td>0.208</td>
<td><strong>0.029</strong></td>
</tr>
</tbody>
</table>

Notes: We report the test statistics and p-value from actual P/Ls and VaRs. The bold numbers are p-values less than 10%.

Our geometric-VaR test rejects the VaR models in both business line 2 and business line 4 at a 5% significance level. The unconditional coverage test fails to reject any of the business line considered here, while the null of duration independence is rejected for all three business lines at a 10% significance level. This is consistent with casual observations from Figure 3.1: for business line 1, 2 and 4, there are signs of violation clustering, yet the average number of violation are not far from the expected value. Business line 1 has the exact number of expected violations, which explains why all the joint tests fail to reject. For business line 2, the null of VaR independence is rejected at a 1% significance level, as well as the joint test of unconditional coverage and VaR independence at a 5% level. Business line 3 shows evidence of violation clustering, as both the duration independence test and the geometric test rejects the null that VaR is correctly specified at a 10% significant level.

The empirical results show that the geometric-VaR test has power against different forms of alternatives. There are cases when the geometric test performs well but the VaR
tests fails to reject. There are other cases where it is the opposite. The geometric-VaR test combines strength from the two tests and it always does well without losing much power from estimating three parameters.

3.6 Conclusions

VaR has become the standard risk measure for financial institutions and regulators. The estimation of VaR is challenging and popular implementations of VaR estimation are based on unrealistic assumptions. We propose the geometric-VaR test to evaluate the performance of VaR forecasts. This test draws strength from a duration-based approach which captures general forms of time dependence in violation, as well as the strength of a regression-based approach, which captures the dependence between VaR forecasts and violations. It can be decomposed into a test of unconditional coverage, a test of duration independence and a test of VaR independence.

We conduct a Monte Carlo study to assess the power of the geometric-VaR test as well as its duration-based components. The dynamics of the P/L series are based on actual desk-level data. We find that the geometric-VaR test has higher power than other duration-based approaches against various alternatives. We apply the tests to the actual business line P/Ls and VaRs. The geometric-VaR is able to reject the efficiency of the VaR forecasts in two of the three business lines we considered. In particular, the geometric-VaR test is able to detect VaR dependence when the geometric test fails to reject in business line 2, and it is able to capture duration dependence when the VaR test fails to reject in business line 4. Our finding suggests that the geometric-VaR test offer a good alternative to detect various forms of VaR misspecifications.
REFERENCES


APPENDIX
Appendix A

MCMC algorithm for SVSDCJ model

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

1. Notations:

\[
\begin{align*}
\Phi_{i+1} &= e^{-\Psi_{D_{i+1}}} \\
\mu_{i+1} &= (I_2 - e^{-\Psi_{D_{i+1}}}) \mu^x + e^{-\Psi_{D_{i+1}}} X_i \\
\hat{r}_{i+1} &= r_{i+1}^o - \xi_{i+1} J_{i+1} \\
\hat{\mu}_{i+1} &= r_{i+1}^o - (m_{i+1} - m_i) \\
\hat{X}_{i+1} &= X_{i+1} - (\Phi_{i+1}) \mu^x - \Phi_{i+1} X_i, \\
\hat{r}_{i+1}^e &= r_{i+1}^e - \xi_{i+1}^y J_{i+1} \\
\hat{X}_{i+1}^e &= X_{i+1}^e - \xi_{i+1}^y J_{i+1} \\
\end{align*}
\]
2. Updating $V$ and $\lambda$:

$$p(V_i|\text{rest}) \propto \frac{1}{V_i} \exp \left( -\frac{1}{2} \frac{(\log V_i - \mu_{vi})^2}{\sigma_{vi}^2} \right) \times V_i^{-\frac{1}{2}} \exp \left( \frac{\left( \tilde{r}_{i+1} - \mu_y D_{i+1} \right)^2}{2V_i D_{i+1}} \right)$$

$$\Sigma_i^* = \left( \Sigma_i^{-1} + \Phi_{i+1}^{-1} \Sigma_{i+1}^{-1} \Phi_{i+1} \right)^{-1}$$

$$\mu_i^* = \Sigma_i^* \left( \Sigma_i^{-1} (\mu_i + \xi_i J_i) + \Phi_{i+1}^{-1} \Sigma_{i+1}^{-1} (X_{i+1} - (I_2 - \Phi_{i+1}) \mu_x - \xi_{i+1} J_{i+1}) \right)$$

$$\mu_{vi} = \mu_{i,1} + \frac{\Sigma_i^*(1, 2)}{\Sigma_i^*(2, 2)} (\log \lambda_i - \mu_{i,2})$$

$$\sigma_{vi}^2 = \Sigma_i^*(1, 1) - \frac{(\Sigma_i^*(1, 2))^2}{\Sigma_i^*(2, 2)}$$

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:

$$q(V_i) \sim IG(c_1 + 0.5, c_2 + 0.5 \frac{(\tilde{r}_{i+1} - \mu_y D_{i+1})^2}{D_{i+1}})$$

$$c_1 = 1 - 2 \exp(\sigma_{vi}^2)$$

$$c_2 = (c_1 - 1) \exp(\mu_{vi} + 0.5\sigma_{vi}^2)$$

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

$$\alpha = \min \left\{ \frac{f(V_i^{(g+1)}) q(V_i^{(g)})}{f(V_i^{(g)}) q(V_i^{(g+1)})}, 1 \right\}.$$
The updating of \( \lambda_i \) follows the same procedure:

\[
p(\lambda_i|\text{rest}) \propto \frac{1}{\lambda_i} \exp \left( -\frac{(\log \lambda_i - \mu_{\lambda i})^2}{2\sigma^2_{\lambda i}} \right) \frac{1}{\lambda_i} \exp \left( -\frac{D_{i+1}}{\lambda_i} \right)
\]

\[
\mu_{\lambda i} = \mu^*_i + \frac{\Sigma^*_i(1,2)}{\Sigma^*_i(1,1)} (\log V_i - \mu^*_i)
\]

\[
\sigma^2_{\lambda i} = (\sigma^*_2)^2 - \frac{(\Sigma^*_i(1,2))^2}{\Sigma^*_i(1,1)}
\]

Proposal density:

\[
q(\lambda_i) \sim IG(d_1 + 1, d_2 + D_{i+1})
\]

\[
d_1 = \frac{1 - 2\exp(\sigma^*_{\lambda i})}{1 - \exp(\sigma^2_{\lambda i})}
\]

\[
d_2 = (d_1 - 1)\exp(\mu_{\lambda i} + 0.5\sigma^2_{\lambda i})
\]

generate \( \lambda_i \) from the proposal density, then accept it using Metropolis-Hastings principle.

3. Updating jump times \( J_i \):

\[
p(J_{i+1} = 1|\text{rest}) \propto \exp \left( \frac{1}{2} \begin{pmatrix} \hat{r}_{i+1} - \mathbf{e}_i Y_{i+1} \\ \hat{X}_{i+1} - \mathbf{e}_i X_{i+1} \end{pmatrix}^T \begin{pmatrix} V_i D_{i+1} & 0 \\ 0 & \Sigma_{i+1} \end{pmatrix}^{-1} \begin{pmatrix} \hat{r}_{i+1} - \mathbf{e}_i Y_{i+1} \\ \hat{X}_{i+1} - \mathbf{e}_i X_{i+1} \end{pmatrix} \gamma D_{i+1}. \right.
\]
Define odds ratio $or = \frac{p(J_{i+1} = 1 | \text{rest})}{p(J_{i+1} = 0 | \text{rest})}$, we have

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

4. Updating jump sizes $\xi$:

$$p(\xi_{i+1} | J_{i+1} = 0, \text{rest}) \sim N(\mu_J, \Sigma_J)$$

$$p(\xi_{i+1} | J_{i+1} = 1, \text{rest}) \propto N(\mu^*_J, \Sigma^*_J)$$

where

$$\mu^*_J = \Sigma^*_J \left( \left( \hat{r}_{i+1} V^{-1} D_{i+1}^{-1} \Sigma_{i+1}^{-1} \hat{X}_{i+1} \right) + \Sigma^{-1}_J \mu_J \right)$$

$$\Sigma^*_J = \left( \left( V_i^{-1} D_{i+1}^{-1} 0 \right) \Sigma_{i+1}^{-1} \left( \begin{array}{c} \Sigma^*_{i+1} \\
0 \end{array} \right) \right)^{-1}.$$ 

5. Updating microstructure noise $m$:

$$p(m_i | \text{rest}) \sim N\left( \frac{S_m}{K_m}, \frac{1}{K_m} \right)$$

$$K_m = \frac{1}{V_i D_{i+1}} + \frac{1}{V_{i-1} D_i} + \frac{1}{\sigma_m^2}$$

$$S_m = \frac{m_{i+1} - \hat{r}_{i+1}}{V_i D_{i+1}} + \frac{\hat{r}_i - m_{i-1}}{V_{i-1} D_i}$$

6. The conditional posterior of $\Psi$ with a diffusive matrix normal prior $MN(A_1, A_2, A_3)$
is given by

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

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

\[
\Psi \sim \text{Multivariate } t \left( \left( (B_2' B_2)^{-1} B_2' B_1 \right)', \left( B_2' B_2 \right)^{-1} \otimes \Sigma_x^x, 3 \right)
\]

\[
B_1 = \left( \frac{X_2 - X_1 - \xi_x J_2}{\sqrt{D_2}}, \ldots, \frac{X_N - X_{N-1} - \xi_x J_N}{\sqrt{D_N}} \right)
\]

\[
B_2 = \left( (\mu^x - X_1) \sqrt{D_2}, \ldots, (\mu^x - X_{N-1}) \sqrt{D_N} \right)
\]

We draw \(\Psi\) from a multivariate t distribution rather than Normal so that the proposal density bounds the target density, then accept \(\Psi\) using Metropolis-Hastings principle.

7. The conditional posterior for \(\Sigma_x\) is given by

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

Proposal density:

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

94
where \( E = \sum_{i=1}^{N-1} E_{i+1} \), and

\[
vec(E_{i+1}) = (I_2 - e^{-(\Psi \odot \Psi)D_{i+1}})(\Psi \odot \Psi)vec \left( (\bar{X}_{i+1} - \mu_{i+1}) (\bar{X}_{i+1} - \mu_{i+1})^T \right).
\]

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 \( \mu_x \) with a multivariate normal prior \( N(m_0, M_0) \) is given by

\[
p(\mu^x|\text{rest}) \sim N(m_\mu, M_\mu)
\]

\[
M_\mu = \left( \sum_{i=1}^{N-1} \left[ (I_2 - \Phi_i) \Sigma_i^{-1} (I_2 - \Phi_i) \right] \right)^{-1} + M_0^{-1}
\]

\[
m_\mu = M_\mu \left\{ \sum_{i=1}^{N-1} \left[ (I_2 - \Phi_i) \Sigma_i^{-1} (X_{i+1} - \Phi_i X_i - \xi_{i+1} J_{i+1}) \right] + M_0^{-1} m_0 \right\}.
\]

9. To update \( \gamma \), 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:

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

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

\[
q(\gamma|J) \propto (\gamma \bar{D})^{\sum_{i=2}^{N} J_i + a_\gamma - 1}(1 - \gamma \bar{D})^{N - 1 - \sum_{i=2}^{N} J_i + b_\gamma - 1}.
\]

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 \( \gamma \) using Metropolis-Hastings principle.
10. Posterior for $\mu_J$ with a multivariate normal prior $N(M_J, Z_J)$:

$$p(\mu_J|\xi, \Sigma_J) \sim N(M^*_J, Z^*_J)$$

$$M^*_J = Z^*_J (Z^{-1}_J M_J + N_j \Sigma^{-1}_J \bar{\xi})$$

$$Z^*_J = (Z^{-1}_J + N_j \Sigma^{-1}_J)^{-1}$$

Posterior for $\Sigma_J$ 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 $\sigma^2_m$, we use a conjugate Inverse Gamma prior $IG(f_m, W_m)$. The posterior is

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

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

96