Understanding Jump Dynamics Using Liquidity Measures

by

Fei Mo

B.Sc., University of Toronto, 2015

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2020

Approval

Name: **Fei Mo**

Degree: **Master of Science (Statistics)**

Title: **Understanding Jump Dynamics Using Liquidity Measures**

Examining Committee: **Chair:** Jinko Graham
Professor

Jean-François Bégin
Senior Supervisor
Assistant Professor

Derek Bingham
Supervisor
Professor

Tim Swartz
Internal Examiner
Professor

Date Defended: **July 15, 2020**

Abstract

Numerous past studies investigate the relationship between volatility and other relevant variables, e.g., asset jumps, liquidity factors. However, empirical studies examining the link between liquidity and jumps are almost non-existent. In this report, we investigate the possible improvement in estimating so-called jump distribution parameters computed from intraday returns by including liquidity measures. More specifically, we first calculate the jump distribution parameters by using classic jump detection techniques in the spirit of Lee and Mykland (2008) and Tauchen and Zhou (2011), and we then use them as our responses in the heterogeneous autoregressive model (e.g., Corsi, 2009). We examine the in-sample performance of our model and find out that liquidity measures do provide extra information in the estimation of the jump intensity and jump size variation. We also apply the same technique but using one-period-ahead instead of contemporaneous responses; we again find extra explanatory power when the liquidity measures are included.

Keywords: high-frequency data, realized variance, liquidity, jump detection.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my senior supervisor, Dr. Jean-François Bégin, for the continuous support during my graduate studies and research. This project would not be complete without his patient guidance and endless encouragement.

I would like to extend my gratitude to Dr. Jinko Graham, Dr. Derek Bingham, and Dr. Tim Swartz for reviewing this report, serving the examining committee, and providing valuable suggestions. Also, I would like to thank all the staff and faculty members here in our department for their help and support.

In addition, I would like to thank all my friends and fellow students. Special thanks go to Sichen Liu and Haiyang Jiang for cheering me up when I was depressed.

Last but not least, I want to express my appreciation to my family for your love, understanding and support throughout my studies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Volatility modelling is essential to asset pricing, asset allocation, and risk management. Researchers have been trying to model and forecast the volatility in both parametric and nonparametric frameworks for the past 50 years, at least. The former approach includes the well-known Heston (1993) model, the generalized autoregressive conditional heteroskedasticity (GARCH) model, and many others. Over the past decade, nonparametric modelling gained increasing attention, starting with Andersen et al. (2001b) and Barndorff-Nielsen and Shephard (2002a) that formalized the use of realized variance. The asymptotic distribution theory states that using realized variance calculated from a fine enough time grid would give a consistent estimator of volatility in the absence of jumps. The realized variance is biased upward in the presence of jumps (Barndorff-Nielsen and Shephard, 2004).

A number of empirical studies have shown that adding a discrete jump component while estimating the volatility would complete the model by explaining the sudden and large spike in the price process. The importance of including such a jump component has been explained in early studies by Bakshi et al. (1997) and Bates (2000). Despite its importance, it is commonly agreed that it is hard to identify jumps because the data is only available in discrete time while the price process is considered to be a continuous-time process. The change in price caused by a continuous rise in volatility would look like a sudden jump when the data is sampled from a coarse time grid. With the increasing availability of high-frequency data and the asymptotic distribution theory developed by Barndorff-Nielsen and Shephard (2002a,b, 2004), numerous nonparametric jump detection techniques have been developed since then. Most of these techniques rely on the bipower variation, a consistent estimator of volatility in the presence of jumps.

One of the main uses of realized variance estimates is to forecast volatility. The heterogeneous autoregressive model for realized volatility (HAR-RV) developed by Corsi (2009) is well-known in this area. The model retains a very simple structure by only using the realized volatility at three different time horizons, but provide a decent prediction of the future volatility.

Liquidity is another factor that plays an important role in financial markets. It provides knowledge about how many orders have been placed in the market and how large is the gap between the buyers' desired prices and sellers' best offers. Empirical studies have reported that liquidity adds value to the explanation of the volatility. For instance, Ramos and Righi (2020) concluded that implied volatility and liquidity are negatively correlated.

Inspired by the connection between jumps and volatility, as well as that between volatility and liquidity, the objective of this project is to investigate whether liquidity measures provide extra value for estimating the jump distribution parameters. These parameters characterize the discrete jump component in the price process (e.g., intensity or rate of jump, mean jump size, standard deviation of the jumps). In this project, they are estimated using the results of the aforementioned jump tests as they are not directly observable. The next two chapters review classic stochastic volatility models and introduce the nonparametric volatility framework along with its estimation and the relevant jump detection techniques. In Chapter 4, we explain liquidity in more detais and present our liquidity proxies. In Chapter 5, we explain how the models are constructed. Our methodology can be seen as an extension of Corsi's HAR-RV model. Finally, we evaluate the performance of our model on ten stocks; the results are provided in Chapter 6.

Chapter 2

Stochastic Volatility Modelling

A stock price represents the market value of each outstanding share of a company. Its price, therefore, reflects the expected return of such stock, its future growth, the company's ability to make profit, and other relevant information. In this day and age, we are able to collect a massive amount of information, not only about the stock price itself, but also on the company's earning report and the market-wide news. This information provides an opportunity to construct complex portfolio positions and adjust them accordingly. Yet, the fundamental goal of trading remains quite simple: buy low and sell high to make a profit. Such selling and buying activities drive price movements and make the stock prices more volatile.

Volatility reveals the variation in price movements and indicates the risk associated with the stock. There has been some attempts in the literature to model the price of a stock and its volatility. The geometric Brownian motion (gBm) model, as used by Black and Scholes (1973), is one of the most common stochastic processes to model stock prices. Such stochastic process assumes that the stock price is continuous, but in reality we can only measure the price at discrete times. Let S_t be stock price at time t. The gBm states that the stock price satisfies the following stochastic differential equation (SDE):

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{2.1}$$

where

- μ is the constant drift term, also known as the expected return of the stock,
- σ is the constant volatility, and
- W_t is the time-t value of a standard Brownian motion.

By applying Itô's lemma on the logarithmic transformation of stock price, $Y_t \equiv \ln(S_t)$, one can show that:

$$\begin{aligned} d\,Y_t &= \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW_t \\ \Longrightarrow r_t \equiv \ln\left(\frac{S_t}{S_{t-1}}\right) &= \left(\mu - \frac{1}{2}\sigma^2\right) + \sigma\left(W_t - W_{t-1}\right), \end{aligned} \tag{2.2}$$

where r_t is the return at time t. Given S_{t-1}, the price at time t can be written as $S_t = S_{t-1}e^{(\mu-\frac{1}{2}\sigma^2)+\sigma(W_t-W_{t-1})}$.

In this model, there are two parameters that need to be estimated (i.e., μ and σ), and they are assumed to be constant.

In the real world, however, volatility is unlikely to be constant over the time. Indeed, the stock prices change every second or even every millisecond providing us with an updated view on the asset. Selling a large amount of shares, for instance, may suddenly cause a plunge in the stock price and a blast in its volatility. A sudden blast in volatility tends to cause the volatility on next day to be high as well. This is known as the long-memory property of volatility. Such impact caused by the first blast will become weaker and weaker as time goes by. Therefore, the volatility is non-constant, and it should change over time.

2.1 ARCH and GARCH Models

The autoregressive conditional heteroskedasticity model (ARCH) proposed by Engle (1982) is a popular model that explains the current variance of stock return as a linear combination of previous squared returns over certain lags. ARCH models assume that the return variance can be modelled by an autoregressive (AR) model . The rationale behind ARCH models is that the volatility is persistent, to some extent. For example, a blast in volatility today makes the volatility more likely to be high tomorrow. More specifically, the ARCH(m) model takes the following form:

$$r_t = \sigma_t z_t, \tag{2.3}$$

$$\sigma_t^2 = a_0 + \sum_{j=1}^{m} a_i r_{i-j}^2, \tag{2.4}$$

where z_t is a white noise process, $a_0 > 0$, $a_i \geq 0$ for $i > 0$, and $\sum_{i=1}^{m} a_i < 1$. Parameters a_i could be estimated by the ordinary least squares method, which makes this model easy to implement in practice.

ARCH models have some limitations such as (a) they only allow for finite lags, (b) ARCH models are relatively more restrictive than other models, and (c) positive and negative spikes in the returns have the same effect on volatility, which is in contradiction with the leverage effect (i.e., negative shocks tend to have a larger impact on volatility than the positive shocks).

The GARCH model was introduced by Bollerslev (1986). The GARCH(m,n) model is defined as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{m} a_i r_{t-i} + \sum_{j=1}^{n} b_j \sigma_{t-j}^2, \tag{2.5}$$

where $a_0 > 0$, $a_i > 0$ for $i = 1, 2, ...m$, $b_j \geq 0$ for $j = 1, 2, ...n$, and $\sum_{i=1}^{m} a_i + \sum_{j=1}^{n} b_j < 1$. It can be shown that a GARCH(1,1) is equivalent to an ARCH(∞) model as all σ_i^2 are included recursively. The ARCH and GARCH models have similar limitations.

2.2 The Heston Model

Heston (1993) proposes a continuous-time stochastic volatility (SV) model that allows the volatility to follow a mean-reverting stochastic process. More precisely, the instantaneous variance of the stock follows Feller's square-root process. This model provides a semi closed-form solution for European option prices, which makes it very popular among practitioners. Let v_t be the value at time t of the instantaneous variance. The Heston model dynamics are given as follows:

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^S, \\ dv_t &= \kappa(\theta - v_t)dt + \xi \sqrt{v_t} dW_t^v, \end{aligned} \tag{2.6}$$

where

- μ is the drift term,
- θ is the long-term level of variance (i.e., the expected value of v_t is θ when $t \to \infty$),
- κ is the rate of mean reversion,
- ξ is the volatility of the variance process v_t,
- W_t^S and W_t^v are time-t values of two standard Brownian motions, and
- $d\ [W^s, W^v]_t = \rho dt$.

The Heston model first relaxes one of the most restrictive assumptions used in the Black and Scholes (1973) model (i.e., constant volatility). Second, the mean-reverting process keeps the volatility at a reasonable level in long run. Finally, it allows for instantaneous correlation between the stock price and its volatility, which is another well-known stylized fact related to the leverage effect mentioned above.

Chapter 3

Nonparametric Estimation of Volatility

ARCH and GARCH models, as well as the Heston model, provide a parametric framework for volatility based on specific distributional assumptions. The validity of such assumptions is difficult to justify, and the parameter estimation under these models often requires intensive computational methods.

In the recent years, the realized measures have caught finance researchers' attention due to its neat statistical properties and the availability of high-frequency return data. Andersen et al. (2001b), among others, advocated for a model-free (so-called nonparametric) approach which utilizes the intraday returns to estimate realized volatility measures. Barndorff-Nielsen and Shephard (2002a,b) studied the properties of realized volatility measures and established their asymptotic distributional theory which became the foundation for other realized measures and jump-detection techniques. Since then, it has been shown empirically that the realized volatility-based method outperforms GARCH-based models and SV models in terms of forecasting (Andersen et al., 2004).

3.1 Theoretical Framework

The evolution of the logarithm of the stock price at time t, Y_t, is assumed to be a semimartingale, and it can be described as a continuous jump-diffusion process:

$$dY_t = \mu_{t^-} dt + \sigma_{t^-} dW_t + d\left(\sum_{k=1}^{N_t} c_k\right), \tag{3.1}$$

where

- μ_{t-} is a locally bounded and predictable process,
- σ_{t-} is a *càdlàg* process,
- W_t is the time-t value of a standard Brownian motion,
- c_k is the size of discrete price jump, and
- N_t is a Poisson counting process with time-varying intensity λ_t.

In practice, we observe Y_t on a discrete-time grid $\{t_0, t_1, ..., t_M\}$ where $0 = t_0 < t_1 < \cdots t_M = T$, over a fixed time period $[0, T]$. In this report, we assume the time grid is equally spaced. For simplicity, we assume that $\Delta = t_i - t_{i-1} = \frac{T}{M}$. Time T is usually one day, and M is the total number of returns observed during a day.

The quadratic variation (QV) process over the given time grid is defined as:

$$QV_t = [Y, Y]_t = \underset{M\to\infty}{\text{p-lim}} \sum_{t=1}^{M} (Y_t - Y_{t-1})^2 , \tag{3.2}$$

where p-lim is the probability limit. Under the framework of Equation (3.1), QV equals to the integrated volatility (IV) of the continuous component of Y_t, plus the summation of squared jumps that occurred up to time t:

$$QV_t = \int_0^t \sigma_{s-}^2 \ ds + \sum_{k=N_0+1}^{N_t} c_k^2. \tag{3.3}$$

Several studies (e.g., Huang and Tauchen, 2005) have shown that jumps make significant contributions to the total price volatility. Nonetheless, the detection of jumps remains challenging as it is difficult to distinguish between a *real* jump and the blast of volatility. Figure 3.1 of this report illustrates this issue (see Christensen et al., 2014, for more details). The sampling frequency plays a key role in the extraction of jumps, while the occurrence of a jump is not related to it. In the absence of jumps, the jump component vanishes and QV is simply equal to the integrated volatility.

3.2 Naive Estimation

3.2.1 Realized Volatility

Let $r_{t,j} = Y_{t-1+j\Delta} - Y_{t-1+(j-1)\Delta}$, for $j = 1, 2, \cdots M$, be the intraday return over the discrete-time grid aforementioned. The daily realized variance is defined as the summation of squared

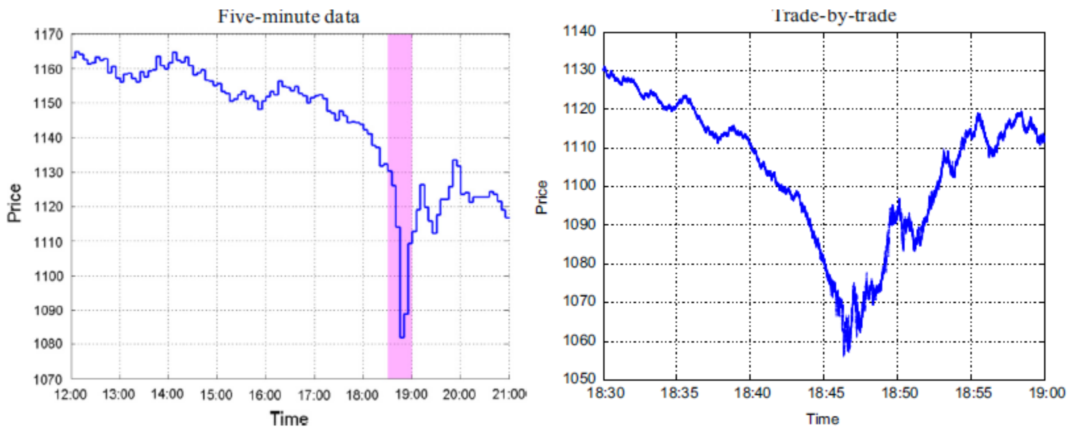

Figure 3.1: **Jump versus Blast of Volatility.**
This figure draws the price evolution of the Chicago Mercantile Exchange's E-mini S&P500 futures contract on May 6,2010. The left panel displays the data for the whole day at a five-minute frequency, while the right panel is at the trade frequency over the flash-crash episode. The reported time is Greenwich Mean Time, while local time in the US is four hours earlier (Eastern Time). This figure was taken from Christensen et al. (2014), page 579.

intraday returns over the day at sampling frequency $\frac{1}{M}$:

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2. \tag{3.4}$$

RV has been carefully studied by Andersen et al. (2001a) and Barndorff-Nielsen and Shephard (2002a,b). The probability limit of the daily RV converges to the increment of QV between $t-1$ and t as the sampling frequency goes to infinity:

$$\operatorname*{p\text{-}lim}_{M\to\infty} RV_t = \int_{t-1}^{t} \sigma_{s^-}^2 \, ds + \sum_{k=N_{t-1}+1}^{N_t} c_k^2 \tag{3.5}$$

The daily RV is a consistent estimator of daily volatility in the absence of jump, but it will be biased upward in the presence of jump as there is clear evidence that jumps make significant contribution to total volatility. Huang and Tauchen (2005) showed that the jumps normally account for about 7% of the daily volatility for the S&P 500 index.

This simple estimator has also been argued to be upward biased in the presence of microstructure noise at high sampling frequencies (e.g., bid-ask bounce). Few methods have been developed in the literature to overcome this. Hansen and Lunde (2006) showed that a kernel-based RV reveals the importance of microstructure noise. Zhang et al. (2005) suggested using subsampled RV to reduce the variance of the estimator, and the bias that comes from microstructure noise. Additionally, Liu et al. (2015) compared over 400 different estimators of QV, and they concluded that there is no strong evidence that any estimators beat the naive five-minute RV significantly. Therefore, we use the naive five-minute RV for this report.

3.2.2 Realized Bipower Variation

In the presence of jumps, we need to find a way to breakup QV into a continuous and a jump part. We follow the approach suggested by Barndorff-Nielsen and Shephard (2004). They studied two generalized realized measures: power and bipower variation (BV). The latter is now a popular estimator of IV:

$$BV_t = \phi_1^{-2}\left(\frac{M}{M-1}\right)\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}| = \frac{\pi}{2}\left(\frac{M}{M-1}\right)\sum_{j=2}^{M}|r_{t,j-1}||r_{t,j}|, \tag{3.6}$$

where $\phi_a = \mathbb{E}\left[|Z|^a\right] = 2^{a/2}\frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma(1/2)}$, $Z \sim \mathcal{N}(0,1)$, and $a > 0$. The bipower variation is robust to jumps as it converges to the integrated variance when $M \to \infty$. Indeed, if jumps are rare events, there will be at most one jump in two adjacent returns when M is sufficiently large. In other words, a non-jump adjacent return will be small enough to offset the impact of a large jump. However, BV still tends to be biased upward in practice as the non-jump return does not vanish sufficiently depending on the sampling frequency used.

The difference between RV and BV converges to the sum of the squared jumps from time $t-1$ to t:

$$\underset{M\to\infty}{\text{p-lim}}\,(RV_t - BV_t) = \sum_{N_{t-1}+1}^{N_t} c_k^2 \tag{3.7}$$

The difference vanishes when there are no jumps. No restrictions were imposed on the left-hand side of the equation, so nothing prevents the observed difference to be negative in practice. One possible simple correction to force the observed difference to be non-negative is:

$$JV_t \equiv \max\left[RV_t - BV_t, 0\right]. \tag{3.8}$$

3.2.3 Multipower Variation

Multipower variation (MPV) is an extension of BV. It was developed to overcome the upward bias in the presence of jumps when M is not sufficiently large. The statistical properties of this estimation were examined by Barndorff-Nielsen et al. (2006). The multipower variation is defined as a sum of products of n adjacent absolute returns, each raised to the power of $\frac{p}{n}$, where n is a positive integer that represents the number of adjacent returns for each term in the summation, and p is total power of the products:

$$MPV_t(n,p) = \phi_{p/n}^{-n} \left(\frac{M}{M-n+1} \right) M^{\frac{p}{2}-1} \sum_{j=n}^{M} \prod_{k=1}^{n} |r_{t,j-k+1}|^{\frac{p}{n}}. \tag{3.9}$$

It is possible to show that:

$$\operatorname*{p-lim}_{M\to\infty} MPV_t(n,p) = \int_{t-1}^{t} \sigma_{s^-}^{p}\, ds,$$

Both RV and BV are special cases of MPV. For instance, $RV_t = MPV_t(1,2)$, and $BV_t = MPV_t(2,2)$. Higher order power variation such as tripower variation (TP) can be written as $MPV_t(3,4)$, and quadpower variation (QP) is $MPV_t(4,4)$. These two higher order multi-power variation estimate the forth power of the volatility, the so-called integrated quarticity (IQ, i.e., $\int_{t-1}^{t} \sigma_s^4\, ds$). IQ is essential for developing the asymptotic distribution of the test statistic of some jump detection tests (more details in Section 3.4).

3.3 Robust Estimation

Although higher-order MPV provides some robustness to the jumps, in practice, its performance and consistency highly depend on the sampling frequency. Corsi et al. (2010) proposed to use a combination of BV and threshold estimation to resolve the jump robustness and small sample bias issues. Coincidentally, Andersen et al. (2012) developed a set of estimators for IV and IQ that use nearest neighbour truncation. These estimators were claimed to have additional robustness to jumps and allow for an asymptotic limit theory that is needed to develop jump tests.

3.3.1 Threshold Bipower Variation

The threshold bipower variation (TBV) is defined as:

$$\begin{aligned} TBV_t &= \phi_1^{-2} \sum_{j=2}^{M} |r_{t,j-1}||r_{t,j}| \mathbf{1}_{\{|r_{t,j-1}|^2 \le \omega_{j-1}\}} \mathbf{1}_{\{|r_{t,j}|^2 \le \omega_j\}}, \\ \omega_t &= g_\omega^2 \cdot \hat{V}_t, \end{aligned} \tag{3.10}$$

where $\hat{V}_t$ is an auxiliary estimator of σ_t^2 and g_ω is a scale-free constant, which can be used to adjust the threshold. The rationale for TBV is the following: if there is a jump in $r_{t,j}$, then the adjacent return $r_{t,j-1}$ should vanish. However, they do not really vanish under finite sample, and this will cause the estimator to be positively biased. With the introduction of a threshold ω_j, we can force the indicator function to vanish in the presence of a jump, thus correct bias. The simulation study has shown that the choice of the auxiliary estimator is immaterial when $g_\omega \geq 3$. We choose $\hat{V}_t = BV_t$ and $g_\omega = 3$ in this report.

3.3.2 Nearest Neighbour Truncation

The nearest neighbour truncation estimator utilizes the use of minimum and median operators to eliminate the returns that are contaminated by the jumps. That is, if there is a jump in any two or three adjacent returns, only the one without a jump will be included in the estimator, so that the estimators converge to IV. The $minRV$ and $medRV$ estimators of IV are defined as:

$$minRV_t = \frac{\pi}{\pi - 2}\left(\frac{M}{M-1}\right)\sum_{j=1}^{M-1} \min\left(|r_{t,j}|, |r_{t,j+1}|\right)^2, \tag{3.11}$$

$$medRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi}\left(\frac{M}{M-2}\right)\sum_{i=2}^{M-1} \operatorname{med}\left(|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|\right)^2, \tag{3.12}$$

where min gives the minimum of a set, and med gives the median of a set.

The idea of the nearest neighbour truncation can be easily extended to the estimation of the IQ:

$$minRQ_t = \frac{\pi M}{3\pi - 8}\left(\frac{M}{M-1}\right)\sum_{j=1}^{M-1} \min\left(|r_{t,j}|, |r_{t,j+1}|\right)^4, \tag{3.13}$$

$$medRQ_t = \frac{3\pi M}{9\pi + 72 - 52\sqrt{3}}\left(\frac{M}{M-2}\right)\sum_{j=2}^{M-1} \operatorname{med}\left(|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|\right)^4. \tag{3.14}$$

3.4 Jump Detection

3.4.1 The BNS Jump Test

As the jump component is embedded in QV, the main challenge in volatility forecasting is to separate the jump and the continuous part of the variance. One simple approach is to take the difference between RV and BV as shown in Section 3.2.2. This difference provides a consistent estimation of the daily jump contribution, and it serves as the basis of the linear BNS (Barndorff-Nielsen and Shephard, 2004, 2006) jump test.

Based on the joint asymptotic distribution of RV_t and BV_t (Barndorff-Nielsen and Shephard, 2006), the daily linear BNS test statistic is defined as:

$$Z_{TP,t} = \frac{RV_t - BV_t}{\sqrt{\left[\left(\frac{\pi}{2}\right)^2 + \pi - 5\right] \frac{1}{M} TP_t}}. \tag{3.15}$$

Under the null hypothesis of no jumps, $Z_{TP,t} \sim \mathcal{N}(0,1)$ as $M \to \infty$. When we reject the null hypothesis, we can conclude that at least one jump occurred during that day.

According to Andersen et al. (2001a), and Barndorff-Nielsen and Shephard (2006), it is natural to work with the logarithmic transformation of the realized measures as it may yield better finite sample performance. The logarithmic BNS daily test is defined as:

$$Z_{TP,l,t} = \frac{\log RV_t - \log BV_t}{\sqrt{\left[\left(\frac{\pi}{2}\right)^2 + \pi - 5\right] \frac{1}{M} \frac{TP_t}{BV_t^2}}}, \tag{3.16}$$

where $\frac{TP_t}{BV_t^2}$ is the estimator of $\frac{\int_{t-1}^{t} \sigma_s^4 ds}{\left(\int_{t-1}^{t} \sigma_s^2 ds\right)^2}$. In theory, $\frac{\int_{t-1}^{t} \sigma_s^4 ds}{\left(\int_{t-1}^{t} \sigma_s^2 ds\right)^2}$ should be less than one, but, in practice, $\frac{TP_t}{BV_t^2}$ could be greater than one due to bias and noise. Barndorff-Nielsen and Shephard (2005) proposed to use the following adjustment:

$$Z_{TP,lm,t} = \frac{\log RV_t - \log BV_t}{\sqrt{\left[\left(\frac{\pi}{2}\right)^2 + \pi - 5\right] \frac{1}{M} \max\left(1, \frac{TP_t}{BV_t^2}\right)}}. \tag{3.17}$$

Another jump measure is the so-called relative jump, which estimates the percentage of jump contribution to the total daily variance:

$$RJ_t = \frac{RV_t - BV_t}{RV_t}.$$

We therefore include two more modifications of the classic BNS jump test based on the relative measure:

$$Z_{TP,r,t} = \frac{RJ_t}{\sqrt{\left[\left(\frac{\pi}{2}\right)^2 + \pi - 5\right] \frac{1}{M} \frac{TP_t}{BV_t^2}}}, \tag{3.18}$$

$$Z_{TP,rm,t} = \frac{RJ_t}{\sqrt{\left[\left(\frac{\pi}{2}\right)^2 + \pi - 5\right] \frac{1}{M} \max\left(1, \frac{TP_t}{BV_t^2}\right)}}. \tag{3.19}$$

3.4.2 Extended BNS Jump Tests

The estimators of IV or IQ could be distorted by the presence of jumps, and jumps would further distort the jump test that uses this estimator. Therefore, it is intuitive to use jump robust estimator for our tests.

We first consider using TBV introduced in Equation (3.10). The threshold bipower variation removes the positive bias in the presence of a jump but it may introduces negative bias at the same time because it forces $r_{t,j}$ to be zero when it is suspiciously large. However, when we observe that $|r_{t,j}|^2 > \omega_j$, it does not mean it is caused by a jump necessarily; it could also be a blast of volatility. Therefore, we correct the negative bias issue by replacing $|r_{t,j}|^2$ with its expected value when $|r_{t,j}|^2 > \omega_j$. The conditional expected value of the power of an absolute return is given by:

$$\mathbb{E}\left[|r_{t,j}|^{\gamma}\middle|(r_{t,j})^2 > \Omega\right] = \frac{\Gamma\left(\frac{\gamma+1}{2}, \frac{\Omega}{2\sigma^2}\right)}{2\Phi\left(-\frac{\sqrt{p}}{\sigma}\right)\sqrt{\pi}}\left(2\sigma^2\right)^{\frac{1}{2}\gamma},$$

where $\Gamma(\alpha, x)$ is the upper incomplete gamma function, $\Omega = g_{\Omega}^2\sigma^2$ is true value of the threshold function defined in Equation (3.10), and Φ is the standard normal cumulative function. The general form of corrected realized threshold multipower variation is then defined as:

$$\text{C-}TMV_t(n,p) = \left(\frac{1}{M}\right)^{1-p/2}\sum_{j=n}^{M}\prod_{k=1}^{n} Z_{p/n}(r_{t,j-k+1}, \omega_{j-k+1}), \tag{3.20}$$

where

$$Z_{\gamma}(x,y) = \begin{cases} |x|^{\gamma}, & \text{if } x^2 \leq y \\ \frac{1}{2\Phi(-g_{\Omega})\sqrt{\pi}}\left(\frac{2}{g_{\Omega}^2}y\right)^{\gamma/2}\Gamma\left(\frac{\gamma+1}{2}, \frac{g_{\Omega}^2}{2}\right), & \text{if } x^2 > y \end{cases}.$$

The C-TBV is a special case of C-TMV: C-$TBV_t = \phi_1^{-2} \cdot$ C-$TMV_t(2,2)$. The corrected threshold BNS test is then defined as:

$$Z_{\text{C-}TBV,rm,t} = \frac{(RV_t - \text{C-}TBV_t)\cdot RV_t^{-1}}{\sqrt{\left(\frac{\pi^2}{4}+\pi-5\right)\frac{1}{M}\max\left\{1, \frac{\text{C-}TrriV_t}{(\text{C-}TBV_t)^2}\right\}}}, \tag{3.21}$$

where C-$TrriV_t$ is the corrected threshold estimator of the IQ:

$$\text{C-}TrriV_t = \phi_{-4/3}^{-3}\cdot\text{C-}TMV_t(3,4).$$

The nearest neighbour truncation modified BNS tests are also included in this report for comparison. To construct such test statistics we simply replace the IV and IQ estimators in $\{Z_{TP,t}, Z_{TP,l,t}, Z_{TP,lm,t}, Z_{TP,r,t}$ and $Z_{TP,rm,t}\}$ with those defined in Equations (3.11) to (3.14), respectively. The notations for this set of tests are listed in Table 6.2 (items 4 to 13).

3.4.3 The Lee and Mykland Test

The BNS test and all its modifications only tell us whether or not jumps occurred during a day, but not how many happened and their occurrence times. Lee and Mykland (2008, henceforth LM) proposed a nonparametric approach that provides insights on how many jumps occurred during a given day and when they happened.

A suspiciously high return could be caused by either a jump or high volatility. To distinguish these two cases, Lee and Mykland (2008) standardized the returns by a realized measure of volatility. Indeed, for each return, a single test is constructed by comparing a realized return at any given time to the volatility over a certain window K. The statistic $\mathcal{L}_{t,j}$ tests whether a jump happened between time $t-1+(j-1)\Delta$ and $t-1+j\Delta$ and is defined as:

$$\mathcal{L}_i \equiv \frac{r_{t,j}}{\hat{\sigma}_{t,j}}, \tag{3.22}$$

where

$$\hat{\sigma}_{t,j} \equiv \frac{1}{K-2} \sum_{k=j-K+2}^{j-1} |r_{t,k-1}||r_{t,k}|\,. \tag{3.23}$$

The $\hat{\sigma}_{t,j}$ is essentially BV over a window of K observations, which is a estimator of the local variance. The rejection region is then given by:

$$\begin{aligned} \frac{\mathcal{L}_{t,j} - C_n}{S_n} &> -\log(-\log(1-\alpha)), \\ C_n &= \frac{(2\log n)^{1/2}}{c} - \frac{\log \pi + \log(\log n)}{2c(2\log n)^{1/2}}, \\ S_n &= \frac{1}{c(2\log n)^{1/2}}, \end{aligned} \tag{3.24}$$

where α is the significance level, n is the total number of observations in the whole sample, and $c = \phi^{-1} = \sqrt{2/\pi}$. In this report, we choose $K = 270$ as suggested by the authors for five-minutes returns.

Chapter 4

Liquidity Measures

Liquidity is receiving increasing attention among practitioners and researchers. Despite the large number of studies on this subject in the literature, giving a definition of *liquidity* is not an easy task as most see it as a multi-dimensional measure instead of being a unidimensional measure. These dimensions vary from asset class to asset class; it depends on the asset type and its features, such as the outstanding amount, term to maturity, or investor risk aversion (Díaz and Escribano, 2020). Liquidity is not directly measurable or observable and, as a result, we use proxies that all serve different purposes.

Traditionally, liquidity is assumed to have three dimensions (e.g., Black, 1971; Kyle, 1985):

1. *Tightness* refers to the transaction cost; it is related to the bid-ask spread (the difference between the bid and ask prices).
2. *Depth* represents the existence of orders (both buying and selling) with prices around the current trading price. A market is deep when numerous orders with different ask and bid prices exist.
3. *Resilience* describes the market's or the stock's ability to absorb a sudden price shock. In case of a price shock, if there are enough new orders to correct the price and bring it back to the normal level, then the it is considered to be resilient.

Some other researchers proposed the use of two additional dimensions: *breadth* and *immediacy* (see Sarr and Lybek, 2002). *Breadth* represents the volume of the existing orders that are described in *depth*. A market is broad if such trading volume is large. *Immediacy* refers to the speed of order execution. Figure 4.1 illustrates the four out of five dimensions of liquidity.

In general, the liquidity of a stock reflects one's ability to buy or sell the stock on the market at a stable price. In other words, it measures how many buyers and sellers are available and

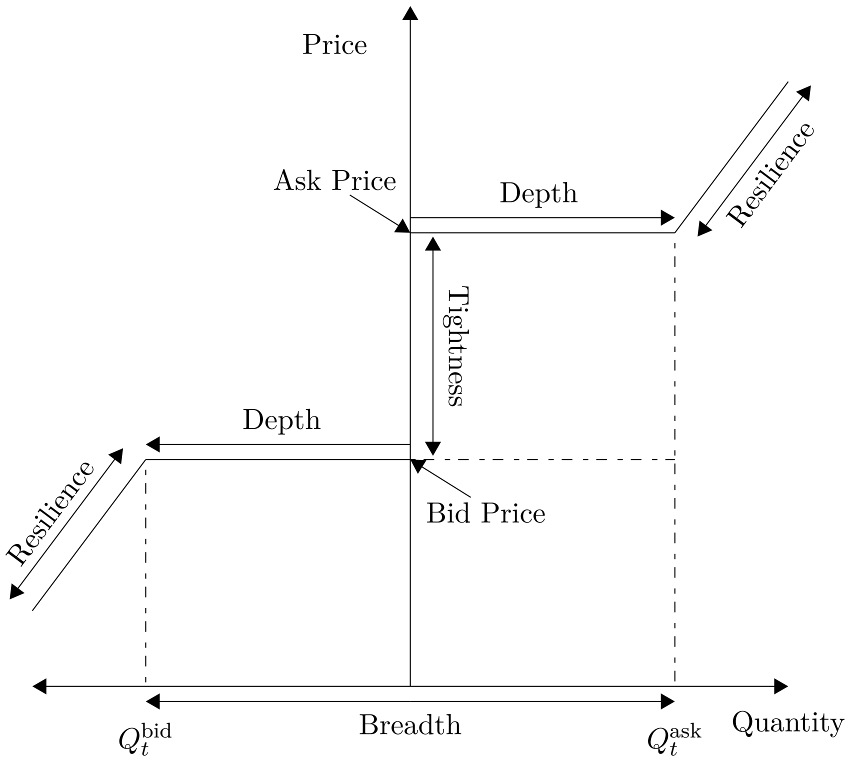

Figure 4.1: **The Various Dimensions of Liquidity**.
This figure illustrates the four dimensions of liquidity: tightness, depth, breadth, and resilience. The fifth dimension (immediacy) is not shown as it represents the time difference between Order 1 and 2, and cannot be shown in such a figure. The first dimension, tightness, represents the bid-ask spread. As such spread becomes small, the actual trading approaches the underlying value, the midquote. Second, depth represents the existence of orders (both buying and selling) with prices around the current trading price. For example, if the current asset price is \$48, and if there exist orders at \$46, \$47, \$48, \$49 and \$50, then such market is deep. Third, breadth represents the trading volume of those orders. Finally, resilience determines the market's ability to absorb a price shock.

how easily the stock can be traded. Investors can close their position easily at a desirable price in a liquid market. There is no need for the seller to cut the price to attract the buyer as enough buyers are present in the market. In contrast, in an illiquid market, an investor tends to be stuck as there are fewer participants.

4.1 Specific Liquidity Measures

In this report, we choose to measure liquidity from two perspectives: the bid-ask spread and the trading volume. The following measures are selected:

1. Quoted spread (QSPR):

$$\mathrm{QSPR}_t = S_t^{\mathrm{ask}} - S_t^{\mathrm{bid}}. \tag{4.1}$$

The quoted spread directly measures the bid-ask spread in dollar. It represents the difference between the price a buyer would pay and the price a seller would receive.

2. Effective spread (ESPR):

$$\mathrm{ESPR}_t = 2D_t \left(S_t - S_t^{\mathrm{mid}}\right), \tag{4.2}$$

where

- S_t is the current trading price,
- $D_t = 1$ if the trade is buy and $D_t = -1$ if the trade is sell. The term D_t is determined using the Lee and Ready algorithm (Lee and Ready, 1991), and
- $S_t^{\mathrm{mid}} = \frac{S_t^{\mathrm{ask}}+S_t^{\mathrm{bid}}}{2}$ is the midquote price.

The effective spread measures the same cost but instead uses the difference between the actual execution price and the midquote price. The midquote price can be viewed as the *true* underlying price for a stock at time t.

3. Proportional quoted spread (PQSPR):

$$\mathrm{PQSPR}_t = \frac{\mathrm{QSPR}_t}{S_t^{\mathrm{mid}}}. \tag{4.3}$$

4. Proportional effective spread (PESPR):

$$\mathrm{PESPR}_t = \frac{\mathrm{ESPR}_t}{S_t^{\mathrm{mid}}}. \tag{4.4}$$

Both proportional quoted and proportional effective spreads express the two preceding spread measures in the percentage of midquote price. Such percentage is assumed to be small in a liquid market as the spread is small. It makes the spread comparable across different stocks as it does not depend on the price level anymore.

5. Trading volume in number of shares (VLMN):

$$\mathrm{VLMN}_t = Q_t, \tag{4.5}$$

where Q_t is the total number of shares traded at time t.

6. Trading volume in dollars (VLMD):

$$\mathrm{VLMD}_t = S_t\, Q_t. \tag{4.6}$$

7. Proportional volume in number of shares (PVLMN):

$$\mathrm{PVLMN} = \frac{\mathrm{VLMN}_t}{O_t}, \tag{4.7}$$

where O_t is the number of outstanding shares.

8. Proportional volume in number of shares (PVLMD):

$$\text{PVLMD} = \frac{\text{VLMD}_t}{O_t\, S^{\text{close}}}, \tag{4.8}$$

   where S^{close} is the daily closing price.

   The proportional volume measures are created using the same logic as that used for proportional spread measures.

All four spread measures are first calculated at each timestamp and then aggregated into daily statistics by taking the daily average. The four volume measures are aggregated into daily statistics by summing over the intraday measures. These liquidity measures, except the two proportional volume measures, have been widely used and been compared to some newly developed proxies in the literature (e.g., Goyenko et al., 2009; Díaz and Escribano, 2020).

4.2 Liquidity, Volatility, and Jumps

It has long been recognized that liquidity and volatility are two correlated aspects of the financial market. Epps and Epps (1976) studied the relationship between the log price and the trading volume with a mixture distribution hypothesis. Their study concluded that the variance of log price changes depends on the volume. Domowitz et al. (2001) examined the relationship between turnover, transaction cost, and volatility, and investigated their impact on the equity returns. They showed that increased volatility, acting through transaction cost, reduces expected return. Chordia et al. (2000) studied the commonality in liquidity while controlling for individual stock's volatility, volume, and price. More recently, Ma et al. (2018) analyzed the relationship between market volatility, liquidity shocks, and stock returns in 41 countries. They concluded that liquidity plays a key role in the market volatility, which impacts stock returns. Nguyen et al. (2020) modelled the joint dynamic of intraday liquidity, volume, and volatility in the US Treasury market. They found out that market liquidity and volume explain volatility dynamics, but not vice versa. On the stock market side, Ramos and Righi (2020) examined the correlation between liquidity and the implied volatility for NYSE-listed stocks. They showed that increases in implied volatility reduces liquidity.

Strong evidence suggests that liquidity is related to the volatility. Meanwhile, the realized measures introduced in Chapter 3 are clearly related to the occurrence of jumps. This feature inspired us to consider if the liquidity are also related to jumps. The stock prices in an illiquid market are more likely to be volatile than that in a liquid market, as mentioned above. This increase in volatility will imply changes in the stock returns, either via the continuous component or the jump part. On any given day, stock returns are most likely to be dominated by the jump component if a jump occurred.

Intuitively, a jump is more likely to occur in an illiquid market (i.e., lack of any of the five dimensions mentioned above). For example, in a thin and shallow market, an executed order with transaction price far from the current price would cause the price to jump. Meanwhile, the size of jump might also relate to the bid-ask spread since the spread impacts the final execution price. Therefore, we conjecture that the jump intensity, jump size, and volatility attributed to the jump component can be explained by the combination of realized and liquidity measures. The next chapters investigate this relationship.

Chapter 5

Methodology

In the past, liquidity and volatility have been studied together, as discussed in Chapter 4. However, empirical studies examining the link between liquidity and jumps are lacking. In this chapter, we propose a simple model to investigate this relationship.

5.1 The Heterogeneous Autoregressive Model

We are interested in exploring the relationship between jump, volatility, and liquidity. Our approach is inspired by the heterogeneous autoregressive model for realized volatility (HAR-RV) proposed by Corsi (2009). The HAR-RV model predicts the future daily volatility by using the realized volatility over the previous day, week, and month:

$$RV_{t+1} = \beta_0 + \beta_{(1)} RV_t^{(1)} + \beta_{(5)} RV_t^{(5)} + \beta_{(22)} RV_t^{(22)} + \epsilon_{t+1}, \tag{5.1}$$

where ϵ_{t+1} is a white noise and

$$\begin{aligned} RV_t^{(1)} &= RV_t, \\ RV_t^{(5)} &= \frac{1}{5}\left(RV_t^{(1)} + \ldots + RV_{t-4}^{(1)}\right), \\ RV_t^{(22)} &= \frac{1}{22}\left(RV_t^{(1)} + \ldots + RV_{t-21}^{(1)}\right). \end{aligned}$$

The coefficients are estimated by ordinary least squares (OLS). The HAR-RV model is based on the concept of heterogeneous market hypothesis (HMH) proposed by Müller et al. (1993). The HMH states that the market consists of three types of investors with different time horizons and trading frequencies (short-, medium-, and long-term). More specifically, short-term investors, such as daily traders and speculators, trade frequently through the day while mid-term investors adjust their positions weekly, and long-term investor such as pension fund may adjust their strategies monthly. Each tenor has its own impact on the next day's volatility. Empirical results shows that the simple HAR-RV model is able to capture the long-memory property of volatility and provides moderate to strong predicting power while maintaining a very simple structure. The model has since then been extended to include jumps, e.g., HAR-RV-J and LHAR-RV-CJ models (see Andersen et al., 2007; Corsi and Renò, 2012, for more details).

By construction, the error terms in the HAR-RV model may be serially correlated. Therefore, in this study, the Newey-West covariance correction is used to adjust the standard error for estimated $\boldsymbol{\beta}$ parameters. In this report, we use the automatic lag selection of $h = 4\,(N/100)^{2/9}$ for the Newey-West correction (Newey and West, 1994), where N is the total sample size of daily RV series.

5.2 Estimating the Jump Distribution

The jump component is not directly observable as explained in Chapter 3. We therefore determine if jumps happened on each trading day first. Once a daily jump indicator is constructed from our jump statistics of Chapter 3, we then filter out the the daily realized jumps in the following way:

$$\hat{J}_t = \operatorname{sign}(r_t)\sqrt{(RV_t - BV_t)\ \mathbf{1}_{\{Z_t > \Phi_\alpha^{-1}\}}}, \tag{5.2}$$

where r_t is the open-to-close daily return, Z_t is a jump statistic value as introduced in Chapter 3, and $\alpha = 0.01$ in this report.

Recall from Equation (3.1) that the jump component is a compound Poisson process. We can therefore estimate the daily jump intensity (λ_t), the mean of jump size (η_t), and its variance (δ_t) once the jumps are identified. We follow the approach used by Tauchen and Zhou (2011) to estimate the jump parameters:

$$\begin{aligned}
\hat{\lambda}_t &= \begin{cases} \frac{\text{Number of Realized Jump Days}}{\text{Window size } K_J}, & \text{if a BNS-like test is used,} \\ \frac{\text{Number of Realized Jumps}}{\text{Window size } K_J}, & \text{if the LM test is used,} \end{cases} \\
\hat{\eta}_t &= \text{Mean of Realized Jumps}, \\
\hat{\delta}_t &= \text{Standard Deviation of Realized Jumps}, \\
\hat{\tau}_t &= \sqrt{\hat{\lambda}_t(\hat{\eta}_t^2 + \hat{\delta}_t^2)},\ \text{Standard Deviation of the Compound Poisson Process},
\end{aligned}$$

where K_J is the window size in days.

The last value is the standard deviation of the compound Poisson process under the assumption of jump sizes are independent and identically distributed; this quantity is also called the standard deviation attributed to the jump component in this report.

The value of first three parameters changes over time, but they have to be estimated over a certain moving window K_J. In other words, the parameters at time t is assumed to be the average from time $t - K_J - 1$ to t. Clearly, the estimation highly depends on the choice of K_J. Tauchen and Zhou (2011) chose a two-year rolling estimation in their study. However, we believe that K_J should not be too long or too short. If it is too short, such as a day or a week, we may only detect zero-jump days which leads the $\hat{\lambda}_t$ equal to zero; on the other hand, the persistency is unlikely to last for a long period, such as two years. We use $K_J = 252$ trading days (i.e., one year) in this report as it gives smooth estimated paths for the time-dependent parameters. The eighteen jump tests presented in Chapter 3 are included in this study for comparison purposes.

5.3 The J-HAR-*RV*-LIQ Model

Our hypothesis is that liquidity measures provide useful information on the parameters of the jump distributions when used in concert with realized measures. The model is designed to establish the relationship among these factors.

Unlike the HAR-RV model of Corsi (2009) that uses one-day ahead RV as the response, we build a model for estimating the current jump distribution first, where we replace the left-hand side of Equation (5.1) with the current-time estimated values, i.e., $\hat{\lambda}_t$, $\hat{\delta}_t$, and $\hat{\tau}_t$. We also modify a few things with respect to the classic Corsi case. First, since jumps are closely related to RV and BV, this motivates us to split the right-hand side into three subsets: RV-based, BV-based, and RV-BV-based models. The RV-based model only includes the logarithm of RV as well as its weekly and monthly lag terms. In contrast, the BV-based model only uses the logarithm of BV as well as its weekly and monthly lag terms. The RV-BV based model includes both realized measures. This step allows us to separately investigate the impact of RV and BV in our model. Including both terms might cause multicollinearity issues between RV and BV by construction. In our study, we choose the logarithmic transformation of daily five-minute RV because the logarithmic daily RV and BV are approximately normally distributed according to Andersen et al. (2003). Second, we add the liquidity measures introduced in Equations (4.1) to (4.8) to all subset models (e.g., RV-based, BV-based, and RV-BV-based) as additional predictors. Daily liquidity measures are also aggregated into weekly and monthly lag terms as done for RV and BV.

The λ-HAR-RV-LIQ model is defined as

$$\begin{aligned}\lambda_t = \beta_0 &+ \beta_{RV^{(1)}} \log\left(RV_t^{(1)}\right) + \beta_{RV^{(5)}} \log\left(RV_t^{(5)}\right) + \beta_{RV^{(22)}} \log\left(RV_t^{(22)}\right) \\ &+ \boldsymbol{\zeta}_{(1)}\boldsymbol{\ell}_t^{(1)} + \boldsymbol{\zeta}_{(5)}\boldsymbol{\ell}_t^{(5)} + \boldsymbol{\zeta}_{(22)}\boldsymbol{\ell}_t^{(22)} + \epsilon_t,\end{aligned} \tag{5.3}$$

where ϵ_t is a white noise and $\boldsymbol{\ell}_t^{(i)}$ contains all the liquidity measures of Chapter 3 with a lag of i, $i \in \{1, 5, 22\}$.

The λ-HAR-BV-LIQ model is defined as:

$$\begin{aligned}\lambda_t = \beta_0 &+ \beta_{BV^{(1)}} \log\left(BV_t^{(1)}\right) + \beta_{BV^{(5)}} \log\left(BV_t^{(5)}\right) + \beta_{BV^{(22)}} \log\left(BV_t^{(22)}\right) \\ &+ \boldsymbol{\zeta}_{(1)}\boldsymbol{\ell}_t^{(1)} + \boldsymbol{\zeta}_{(5)}\boldsymbol{\ell}_t^{(5)} + \boldsymbol{\zeta}_{(22)}\boldsymbol{\ell}_t^{(22)} + \epsilon_t,\end{aligned} \tag{5.4}$$

and the λ-HAR-$RVBV$-LIQ model is defined as

$$\begin{aligned}\lambda_t = \beta_0 &+ \beta_{RV^{(1)}} \log\left(RV_t^{(1)}\right) + \beta_{RV^{(5)}} \log\left(RV_t^{(5)}\right) + +\beta_{RV^{(22)}} \log\left(RV_t^{(22)}\right) \\ &+ \beta_{BV^{(1)}} \log\left(BV_t^{(1)}\right) + \beta_{BV^{(5)}} \log\left(BV_t^{(5)}\right) + \beta_{BV^{(22)}} \log\left(BV_t^{(22)}\right) \\ &+ \boldsymbol{\zeta}_{(1)}\boldsymbol{\ell}_t^{(1)} + \boldsymbol{\zeta}_{(5)}\boldsymbol{\ell}_t^{(5)} + \boldsymbol{\zeta}_{(22)}\boldsymbol{\ell}_t^{(22)} + \epsilon_t.\end{aligned} \tag{5.5}$$

The scale of the VLMD and VLMN measures are relatively large when compared to the other liquidity measures. Therefore, we standardize these two measures to avoid non-invertible matrix problem. The models we use for δ_t and τ_t as responses are defined similarly to Equations (5.3), (5.4), and (5.5). The models without any liquidity measures are taken as benchmark models, which provides us with a guideline to test if liquidity measures contain additional information.

The one-period-ahead forecasting model is built on same idea, with one-period-ahead estimated jump parameter used as the response. For instance, a one-period-ahead λ-HAR-RV-LIQ model is defined as:

$$\begin{aligned}\lambda_{t+K_J} = \beta_0 &+ \beta_{RV^{(1)}} \log\left(RV_t^{(1)}\right) + \beta_{RV^{(5)}} \log\left(RV_t^{(5)}\right) + \beta_{RV^{(22)}} \log\left(RV_t^{(22)}\right) \\ &+ \boldsymbol{\zeta}_{(1)}\boldsymbol{\ell}_t^{(1)} + \boldsymbol{\zeta}_{(5)}\boldsymbol{\ell}_t^{(5)} + \boldsymbol{\zeta}_{(22)}\boldsymbol{\ell}_t^{(22)} + \epsilon_{t+K_J},\end{aligned} \tag{5.6}$$

where ϵ_{t+K_J} is a white noise. The BV- and RV-BV-based models are defined in a similar way.

Using $\hat{\lambda}_t$ as response estimates the number of jumps that would occur during a given period of time. We also expect that the liquidity measures provide additional value information on the jump sizes variation. By simply switching the response to $\hat{\delta}_t$ and $\hat{\tau}_t$ we model the variation of each jump and the variation of the jump component, respectively. These three responses investigate the relationship between jumps and liquidity from two different perspectives: rate or intensity (λ_t) and variation (δ_t and τ_t).

5.4 Variable Selection

Unlike the traditional variable selection methods such as principal component analysis and LASSO, we decide to select variable based on the in-sample model performance across all tickers. Specifically, we use the adjusted R^2, the root-mean-square error, and the Newey-West adjusted t-test. The selection step is described as follows:

1. For each jump test, every liquidity measure enters into the benchmark model individually. The model performance is then compared to that of benchmark model.
2. For a given response (either $\hat{\lambda}_t$, $\hat{\delta}_t$, or $\hat{\tau}_t$), we choose the top four predictors that appear to be the most significant (determined by t-test), and the one that increases the adjusted R^2 the most when compared to the benchmark model.
3. The selected predictors are then added to the benchmark model simultaneously.

Chapter 6

Data and Results

6.1 Data

6.1.1 Company Selection

The millisecond level quotes and trades data first became available in the TAQ database on September 10, 2003. Our sample period, therefore, begins on September 10, 2003 and finishes on December 31, 2016. Ten companies were randomly selected from the constituents of the S&P 500 index. Historical data indicate that the average number of stocks being part of the index is approximately 500, with immaterial fluctuation. Figure 6.1 shows the historical number of stocks that were included in the S&P 500 index. To get an uninterrupted series for each company, we only picked the companies that were part of the index on every single trading day during the whole sample period. There were 251 stocks that fit this criterion.

We first sort these 251 companies based on their market capitalization and divide them into ten bins. We then randomly choose one in each bin. The selected company and their market capitalization are given in Table 6.1. Our selection covers a wide range of sectors.

6.1.2 The Data Cleaning Procedure

The data cleaning procedure plays an important role in our econometric analysis. An improper cleaning process would include redundant information and remove useful pieces. We follow the procedure used in Barndorff-Nielsen et al. (2009). Their cleaning procedure has now become the standard in high-frequency data analysis. The steps are as follows:

- For all data:

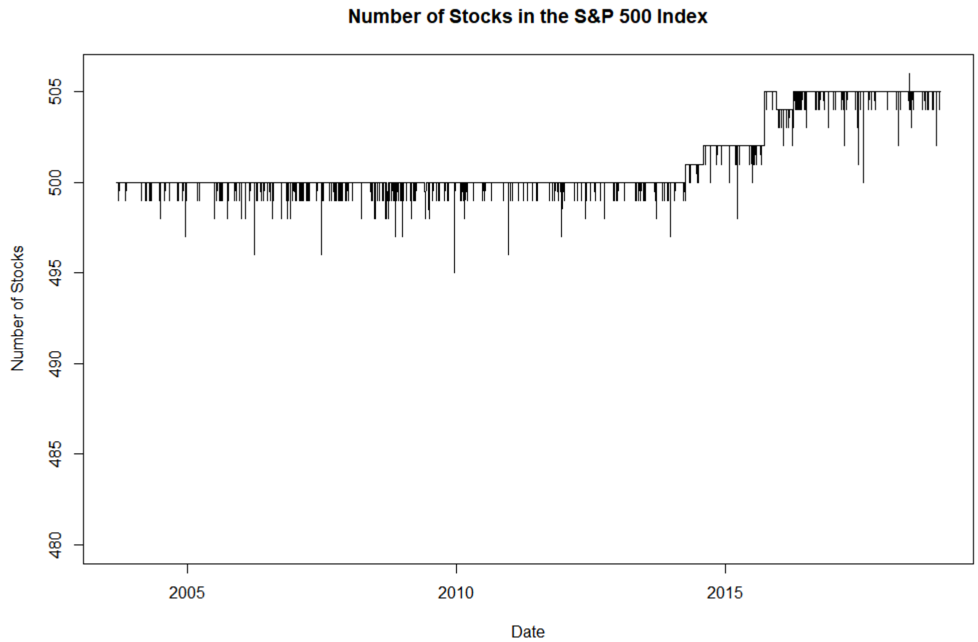

Figure 6.1: **Daily Number of Stocks in the S&P 500 Index.**
This information was obtained from the Index Consistuents dataset in Compustat via Wharton Research Data Services (WRDS).

1. Only entries with a timestamp between 9:30 am to 4:00 pm Eastern Time are retained.
2. We delete the entries with a bid, ask, or transaction price equal to zero.
3. We retain the entries originating from a single exchange centre. We delete other entries. We choose NYSE for all our stocks, except for SYMC as it is listed on the NASDAQ.

- For quotes data only:
  1. When multiple quotes have the same timestamp, we replace them with a single entry with the median bid and ask prices.
  2. We delete the entries for which the spread is negative.
  3. We delete the entries for which the spread is more that 50 times the median spread on that day.
  4. We delete the entries for which the midquote deviated by more than ten mean absolute deviations from a rolling centred median of 50 observations (excluding the observation under consideration).
- For trades data only:
  1. We delete the entries with corrected trades. (i.e., trades with a correction indicator, CORR, not equal to zero).

Table 6.1: **Market Capitalization of the Selected Companies.**

Bin	Company	Market Cap	Sector	Ticker
1	Walt Disney Co.	$ 163,232,775,150	Media	DIS
2	United Technologies Corp.	$ 91,933,128,320	Manufacturing	UTX
3	Cigna Corp.	$ 72,317,357,760	Insurance	CI
4	Deere & Co.	$ 47,521,236,070	Manufacturing	DE
5	Progressive Corp-Ohio	$ 35,184,456,000	Insurance	PGR
6	Consolidated Edison Inc.	$ 23,815,454,960	Energy	ED
7	Edison International	$ 18,496,290,470	Utility	EIX
8	Key Corp.	$ 15,290,353,400	Finance	KEY
9	Robert Half Intl Inc.	$ 12,071,788,760	Consulting	RHI
10	Symantec Corp.	$ 6,948,370,000	Software	SYMC

The market capitalization is calculated as of December 31st, 2016 by multiplying the number of outstanding shares by daily closing prices. The data were obtained from the Security Daily dataset in Compustat via WRDS. We sort the market capitalization and divide all companies into ten bins. One company in each bin is randomly selected.

2. We delete the entries with abnormal sale condition, that is we only retain entries with the COND flag equal to any of 'E', 'F', '@', or 'NA'.
3. If multiple transactions have the same timestamp, we use the median of these prices.
4. We delete the entries with prices that are above the ask plus the bid-ask spread and the entries with prices below the bid minus the bid-ask spread.

6.1.3 Data Overview

Figure 6.2 reports the time series plots of daily RV for each firm. The plots are consistent with the well-known long-memory property of volatility. The impact of a volatility shock tends to last for a long period before it disappears. The figure also shows that some volatility spikes happened over the 13 years under investigation in this study. For example, we observe that the volatility rose around September 2008 at the same period when Lehman Brothers announced their bankruptcy. Most of them reverted back to long-run level by mid 2009, with the exception of CI and KEY. The volatility of these two firms increased as early as late 2007. Both CI and KEY belong to the insurance and finance sector, which means they were in the front line during the 2008 Financial Crisis. While the crisis happened in late 2008, banks and other financial institutions experienced the impact earlier. This phenomenon suggests that a systemic event would impact volatility and create price shocks across all sectors.

The path of proportional spread measures are given in Figures 6.3 and 6.4. PQSPR appears to be stable over time, except during the period of the 2008 Financial Crisis. Similar to RV, there was an obvious spike in this spread during the financial crisis, while the volatility of

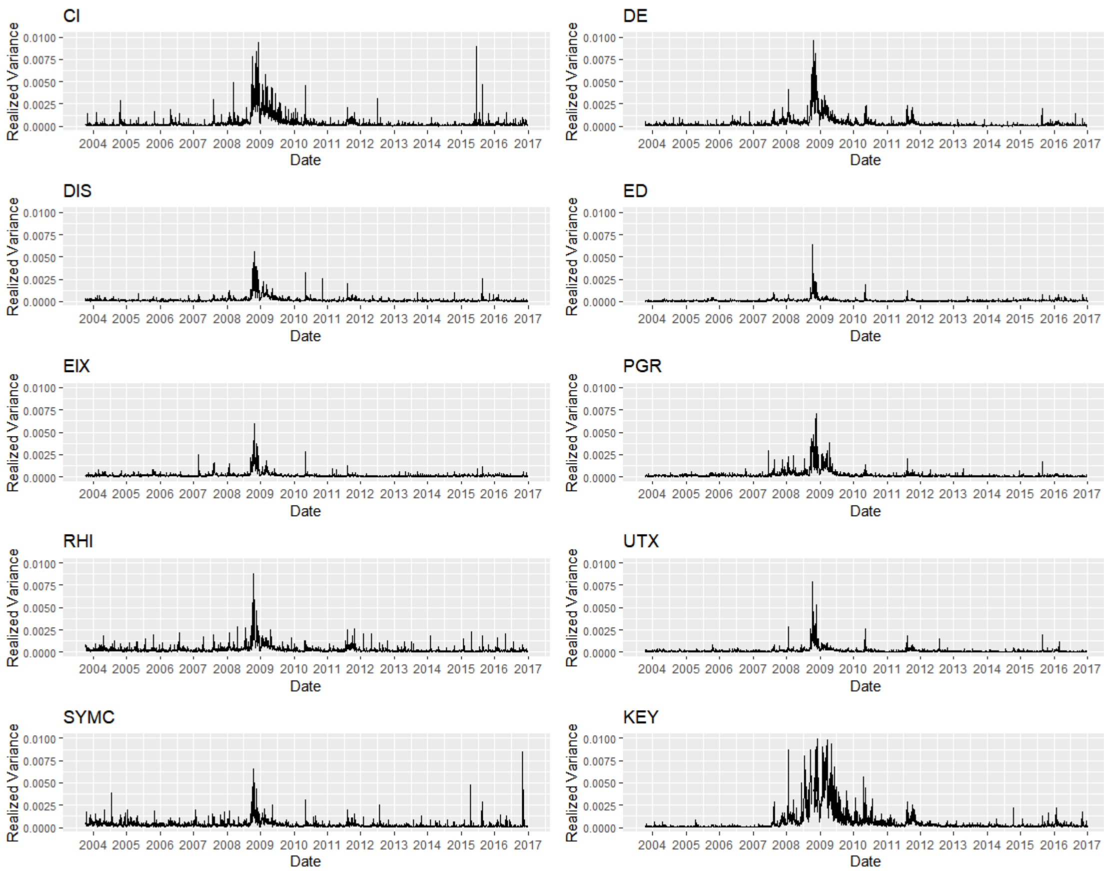

Figure 6.2: **Time Series Plot of Daily Realized Variance.**
This figure illustrates the path of daily five-minute RV between 2003 and 2016 for each stock.

the price series also peaked. This finding is consistent with some of the studies we mentioned in Chapter 4: an increase in the volatility reduces liquidity. All firms but KEY reverted back to the normal level after 2011. The volume series are given in the appendix of this report (see Figures A.1 to A.2). These two series are more volatile at the beginning of our sample period.

6.2 Jump Distribution Estimation

In this section, we present the jump detection results and the estimated jump distribution parameters. For each jump test, Table 6.2 summarizes the average annual number of days for which a jump has been filtered. The tests that use nearest neighbour truncation appear to be more conservative than the other tests, with $Z_{minRV RQ,rm,t,}$ to be the most conservative test. On average, it only captures 9 to 15 jumps per year. It is unclear to us which stock has the most jumps during the sample period as the results change for each test.

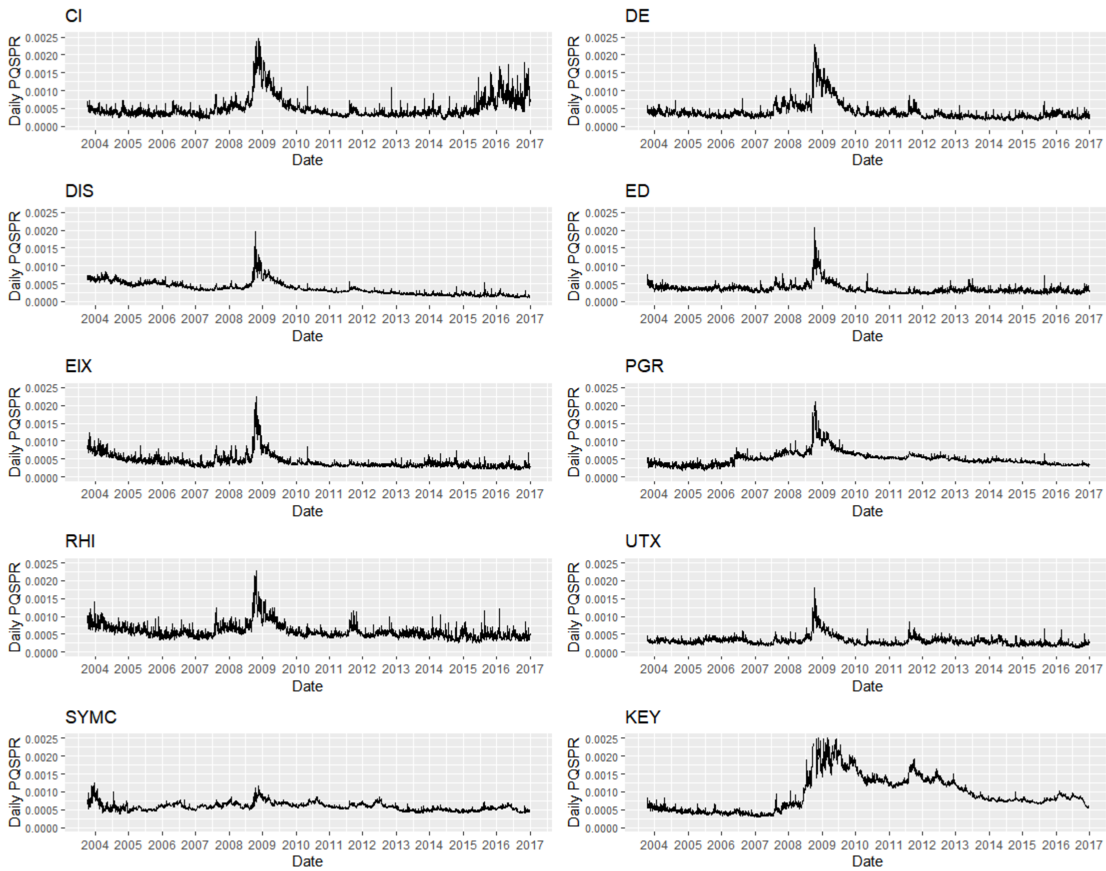

Figure 6.3: **Time Series Plot of Daily Proportional Quoted Spread.**
This figure illustrates the path of daily PQSPR between 2003 and 2016 for each selected stock. The daily measure is calculated by taking the mean of the intraday PQSPR.

We then estimate the jump distribution parameters using the method discussed in Section 5.2. The estimated paths for jump intensity using $Z_{TP,t}$ are given in Figure 6.5 for illustrative purposes. The jump intensity rates vary between 15% and 30% for most selected firms. We expect $\hat{\lambda}_t$ to spike when there is a systematic risk, such as the 2008 Financial Crisis. Not all intensity rate paths surged during the crisis period. In fact, some of them actually reached a local minima at the beginning of 2009.

There are other times where the intensity moves in the same direction. For example, CI, EIX, RHI, DE, and PGR all reached their own local minima near mid 2012. This suggests that the jump intensity for individual stocks may be affected by systemic events, which would cause co-jumps to occur across sectors. We also notice that the intensity for KEY continued to rise from 2008 to 2014 which may be related to the high spread level we observe.

The path for $\hat{\delta}_t$ and $\hat{\tau}_t$ are presented in Figures 6.6 and A.3, respectively. These two parameters share similar dynamics: both series experienced a large surge during the 2008 Financial

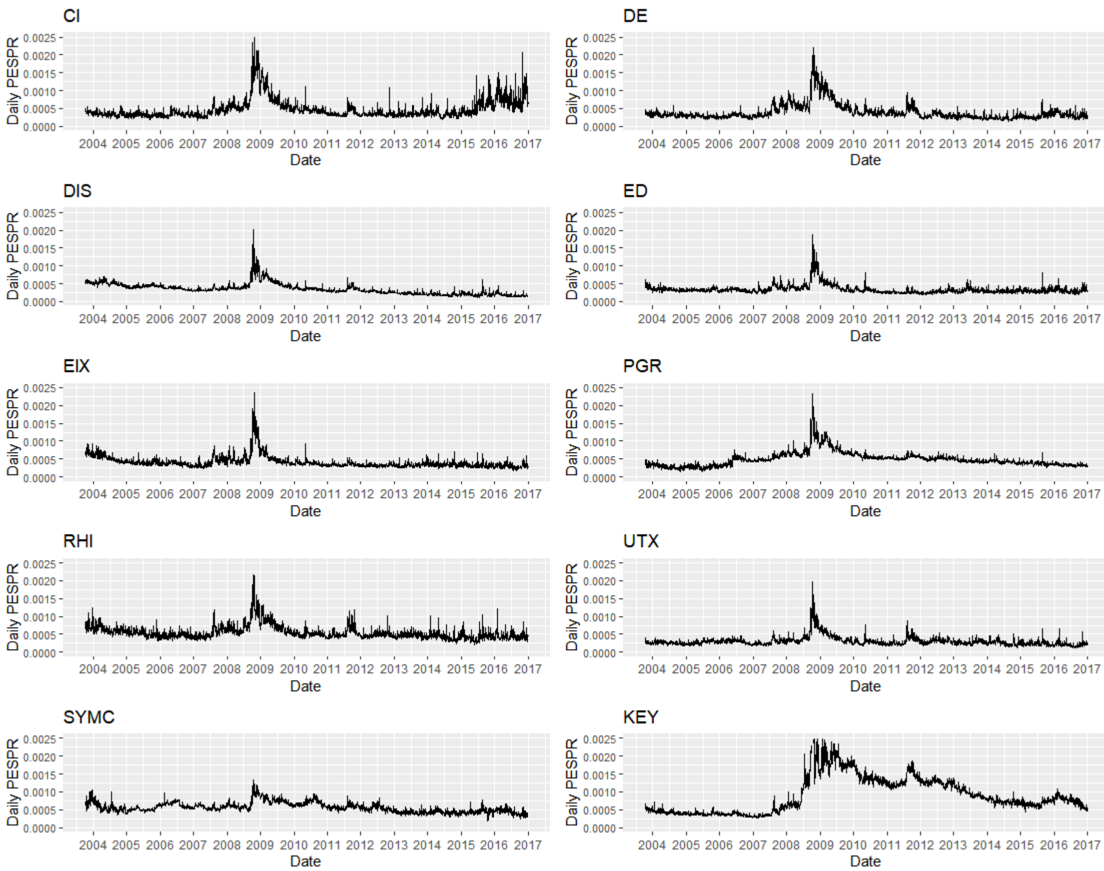

Figure 6.4: **Time Series Plot of Daily Proportional Effective Spread.**
This figure illustrates the path of daily PESPR between 2003 and 2016 for each selected stock. The daily measure is calculated by taking the mean of the intraday PESPR.

Crisis. By construction, we know that λ_t, η_t, and δ_t all contribute to τ_t, but the similar behaviour of δ_t and τ_t suggests that the jump volatility dominates the variation of the jump component. The counting process, or by extension the jump intensity, has a minimal impact on τ.

6.3 Correlation Analysis and Regressions

6.3.1 Correlation Analysis

A preliminary correlation analysis is conducted to examine the relationship between the daily liquidity measures and the different responses (i.e., $\hat{\lambda}_t$, $\hat{\delta}_t$, and $\hat{\tau}_t$). Since there are eighteen jump tests in our study, and they all lead to their own jump parameters, we only report the correlation based on the average of these eighteen values for each company. The correlation matrices are reported in Table 6.3.

Table 6.2: **Estimated Annual Average Number of Days with Jumps.**

	Jump Statistics	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
1	LM	54	33	32	41	42	40	46	29	44	32
2	$Z_{\text{C-}TBV,lm,t}$	49	34	40	48	47	60	55	50	54	78
3	$Z_{\text{C-}TBV,rm,t}$	40	27	32	39	38	50	44	41	44	66
4	$Z_{medRVRQ,l,t}$	28	22	24	26	27	27	32	31	27	27
5	$Z_{medRVRQ,lm,t}$	28	22	23	25	26	27	32	30	26	27
6	$Z_{medRVRQ,r,t,}$	20	15	15	17	17	19	22	21	18	17
7	$Z_{medRVRQ,rm,t}$	20	15	15	17	17	19	22	21	18	17
8	$Z_{medRVRQ,t}$	37	30	31	34	35	36	41	39	36	35
9	$Z_{minRVRQ,l,t}$	26	20	20	24	24	25	29	28	25	24
10	$Z_{minRVRQ,lm,t}$	25	20	19	23	23	24	28	27	23	23
11	$Z_{minRVRQ,r,t}$	15	12	10	11	13	13	16	15	12	12
12	$Z_{minRVRQ,rm,t,}$	14	10	9	11	11	13	15	14	11	10
13	$Z_{minRVRQ,t}$	37	29	28	33	34	35	40	40	35	35
14	$Z_{TP,l,t}$	36	27	31	36	36	49	42	41	43	63
15	$Z_{TP,lm,t}$	38	29	32	38	38	51	43	43	44	65
16	$Z_{TP,r,t}$	30	23	25	30	29	39	35	34	35	54
17	$Z_{TP,rm,t}$	28	21	24	28	27	36	32	32	33	52
18	$Z_{TP,t}$	46	36	41	46	47	60	53	51	53	74

The daily jump indicators are first computed by each jump test. Then, annual average number of days with indicators equal to one are reported in this table.

PVLMD and PVLMN appear to have similar explanatory power as their correlations are identical. All four volume-based measures are slightly negatively correlated with $\hat{\lambda}_t$ except for RHI and SYMC. The correlation matrices for $\hat{\delta}$ and $\hat{\tau}$ are similar in terms of their sign and magnitude. This is consistent with Figures 6.6 and A.3 as they share similar dynamics. The two proportional spread measures have the strongest correlations with $\hat{\delta}$ and $\hat{\tau}$, and the highest correlation coefficient is 80% (between PQSPR and $\hat{\tau}$ for DE).

6.3.2 The Univariate Regressions

The univariate regressions, where one liquidity measure is added to the benchmark models at a time, are used for variable selection purposes. We record the number of times the coefficient associated with a liquidity measure is being significant and the average adjusted R^2 improvement across all companies and jump tests. The significance is determined by the Newey-West corrected t-statistic with a significance level of 0.05. For each response and each subset model, we list the top five liquidity measures determined as the most significant and the top five liquidity measures increasing the adjusted R^2 the most.[1] Moreover, we only select one version of the spread and volume measures at each lag to avoid multicollinearity

[1] The increase in the adjusted R^2 is computed as the difference between the R^2 of the liquidity-based model and the R^2 of the benchmark model.

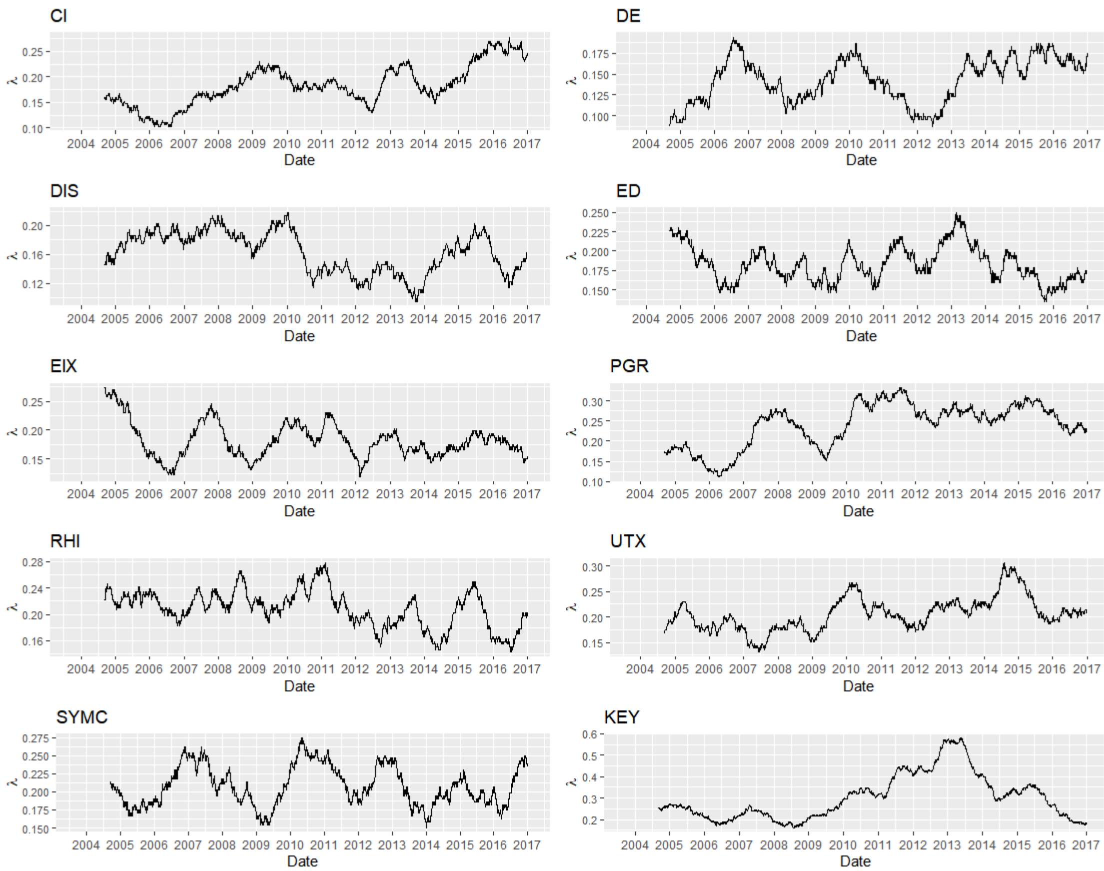

Figure 6.5: **The Estimated Path of Jump Intensity Using $Z_{TP,t}$.**
This figure illustrates the path of estimated jump intensity ($\hat{\lambda}_t$) using the basic BNS test (i.e., $Z_{TP,t}$). The daily jump indicators are first calculated using $Z_{TP,t}$. The jump intensity rate is then estimated using a window size of $K_J = 252$.

issues; for example, if ESPR$^{(1)}$ is selected then PESPR$^{(1)}$, QSPR$^{(1)}$ and PQSPR$^{(1)}$ will not be selected. Table 6.4 lists the top five variables under different model settings.

The top five variables do not vary much among RV-based, BV-based and RV-BV-based models. For $\hat{\lambda}_t$, the volume measure VLMD$^{(1)}$ is the one that shows the strongest significance. The selection for $\hat{\delta}_t$ and $\hat{\tau}_t$ is almost identical. Moreover, VLMD$^{(22)}$ tends to improve the adjusted R^2 the most across all settings.

To provide a better understanding of the model-specific results, the estimated coefficients and their corresponding p-values for liquidity measures under univariate $\hat{\lambda}_t$-RV-based models (based on $Z_{TP,t}$) are reported in Tables B.1 to B.3. The daily liquidity measures appear to be more significant than weekly and monthly measures; in other words, the significance decreases as the lag rises. A similar trend is observed while using other jump tests. This indicates that the daily measures of liquidity is better at explaining the dynamic jump intensity rates. However, no single measure is determined to be significant for all companies,

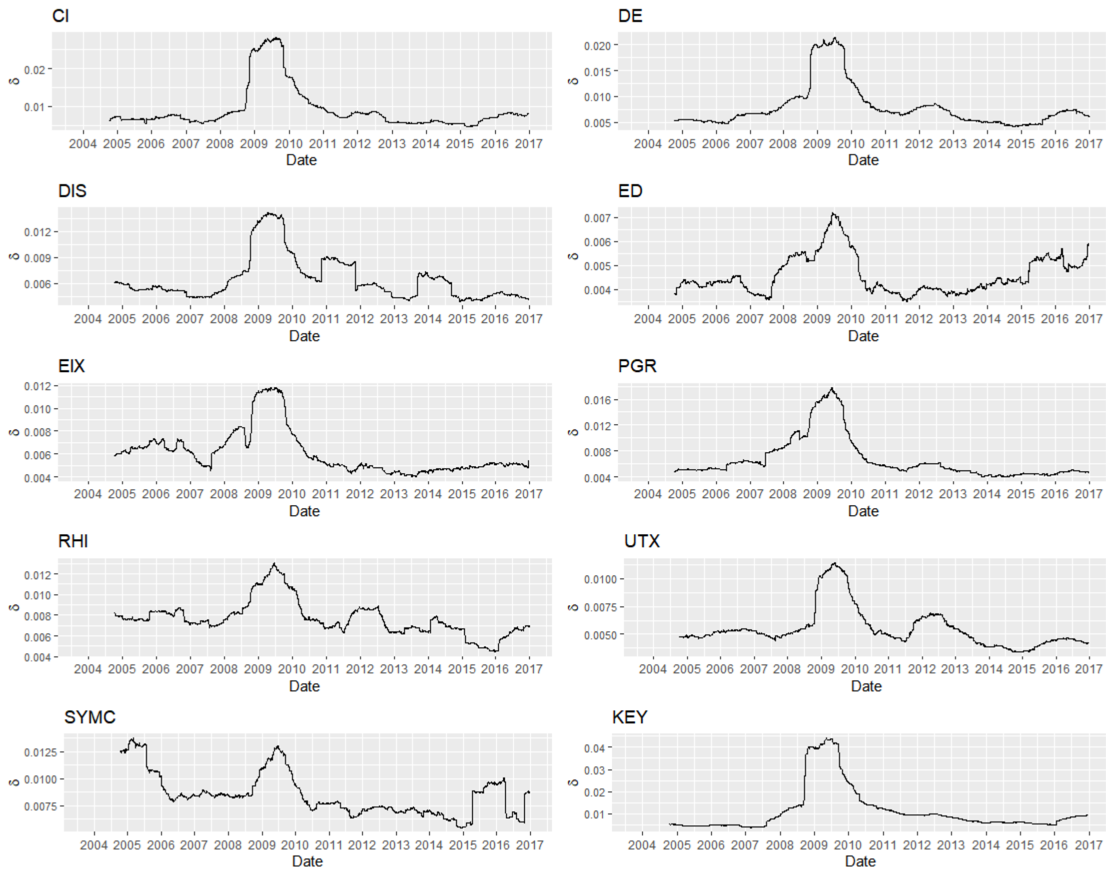

Figure 6.6: **The Estimated Path of Jump Standard Deviation Using $Z_{TP,t}$.**
This figure illustrates the path of estimated jump standard deviation ($\hat{\delta}_t$) using the basic BNS test (i.e., $Z_{TP,t}$). The daily jump indicators are first calculated using $Z_{TP,t}$. The daily jumps are then filtered using Equation (5.2). The $\hat{\delta}_t$ is finally estimated with the filtered jumps by using a window size of $K_J = 252$.

which means the selected liquidity measures may not be robust. For example, all daily liquidity measures are determined to be significant for CI while none of them are for DE and ED.

6.3.3 The Multivariate Regressions

Adding a single liquidity measure to the benchmark model is already providing us with some additional explanatory power about the jump dynamics. We, therefore, expect that having multiple liquidity measures would create more powerful models. In this section, we report the results of multivariate regressions.

As described in Section 5.4, and for parsimony's sake, we only select four variables among those being the most significant, plus the one that increases the adjusted R^2 the most for

Table 6.3: **Correlation Between Daily Liquidity Measures and Estimated Jump Intensity, Jump Standard Deviation, and Standard Deviation of Jump Component.**

Panel A: Correlation for Estimated Jump Intensity ($\hat{\lambda}_t$).

	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
QSPR$^{(1)}$	0.27	-0.10	0.06	-0.12	-0.06	-0.38	-0.06	0.01	0.01	-0.09
ESPR$^{(1)}$	0.29	-0.07	0.03	-0.14	-0.10	-0.38	-0.06	0.03	0.01	-0.09
PQSPR$^{(1)}$	0.23	-0.16	0.04	-0.13	0.10	0.03	-0.09	-0.15	-0.12	0.12
PESPR$^{(1)}$	0.25	-0.14	0.03	-0.15	0.06	0.04	-0.09	-0.13	-0.11	0.11
VLMN$^{(1)}$	-0.28	-0.26	-0.01	-0.14	0.13	0.04	0.03	-0.25	0.02	-0.05
VLMD$^{(1)}$	-0.37	-0.19	0.00	-0.16	0.01	-0.26	0.03	-0.21	0.04	-0.21
PVLMD$^{(1)}$	-0.37	-0.22	-0.01	-0.13	0.13	-0.23	0.01	-0.25	0.04	-0.17
PVLMN$^{(1)}$	-0.37	-0.22	-0.01	-0.13	0.13	-0.23	0.01	-0.25	0.04	-0.17

Panel B: Correlation for Estimated Jump Standard Deviation ($\hat{\delta}_t$).

	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
QSPR$^{(1)}$	-0.08	0.40	0.09	0.32	0.35	-0.04	0.04	0.21	0.35	0.43
ESPR$^{(1)}$	-0.09	0.37	0.04	0.32	0.28	-0.02	0.01	0.22	0.42	0.48
PQSPR$^{(1)}$	0.68	0.77	0.53	0.36	0.54	0.69	0.42	0.57	0.29	0.75
PESPR$^{(1)}$	0.68	0.77	0.59	0.38	0.52	0.68	0.42	0.57	0.35	0.73
VLMN$^{(1)}$	0.27	0.06	0.01	-0.02	0.20	0.49	0.17	0.14	0.15	0.26
VLMD$^{(1)}$	-0.14	-0.08	-0.05	-0.05	0.08	0.01	0.06	-0.00	0.23	-0.19
PVLMD$^{(1)}$	0.07	-0.06	0.02	-0.03	0.20	0.20	0.17	0.09	0.26	0.32
PVLMN$^{(1)}$	0.07	-0.06	0.02	-0.03	0.20	0.20	0.17	0.09	0.26	0.32

Panel C: Correlation for Estimated Standard Deviation of Jump Component ($\hat{\tau}_t$).

	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
QSPR$^{(1)}$	-0.03	0.42	0.16	0.35	0.36	-0.13	0.02	0.24	0.38	0.43
ESPR$^{(1)}$	-0.04	0.39	0.10	0.34	0.28	-0.12	-0.01	0.25	0.45	0.48
PQSPR$^{(1)}$	0.73	0.80	0.56	0.40	0.59	0.74	0.44	0.60	0.22	0.76
PESPR$^{(1)}$	0.72	0.79	0.62	0.41	0.56	0.73	0.44	0.60	0.28	0.74
VLMN$^{(1)}$	0.22	0.02	-0.00	-0.05	0.26	0.51	0.18	0.08	0.17	0.22
VLMD$^{(1)}$	-0.19	-0.12	-0.06	-0.09	0.11	-0.05	0.06	-0.07	0.25	-0.24
PVLMD$^{(1)}$	0.01	-0.09	0.00	-0.06	0.26	0.15	0.18	0.03	0.30	0.25
PVLMN$^{(1)}$	0.01	-0.09	0.00	-0.06	0.26	0.15	0.18	0.03	0.30	0.25

This table reports the average correlation between daily liquidity measures and three estimated jump distribution parameters. The the average correlation is taken over the eighteen jump tests presented in this report.

further analyses. If the one with the best R^2 improvement has already been included under the significance criterion, then the second best is added.

The RV- and BV-based models have similar performance under all model settings. We believe this is a consequence of the difference between RV and BV being negligible in the absence of jumps. The RV-BV-based models have the highest adjusted R^2, but they also suffer the most from multicollinearity issues. Therefore, we only report the result for RV-based models as RV is the most widely-used measure for HAR-like models.

Table 6.4: **Top Variables Under Two Criteria for Estimated Jump Parameters.**

Panel A: Top Five Variables for Jump Intensity ($\hat{\lambda}_t$).

	Significance	Adjusted R^2
RV-Based Model	VLMD$^{(1)}$	VLMD$^{(22)}$
	QSPR$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	VLMD$^{(1)}$
	ESPR$^{(5)}$	ESPR$^{(22)}$
	VLMD$^{(22)}$	ESPR$^{(5)}$
BV-Based Model	VLMD$^{(1)}$	VLMD$^{(22)}$
	QSPR$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	VLMN$^{(1)}$
	ESPR$^{(5)}$	ESPR$^{(22)}$
	VLMD$^{(5)}$	ESPR$^{(5)}$
RV-BV-Based Model	VLMD$^{(1)}$	VLMD$^{(22)}$
	ESPR$^{(1)}$	VLMD$^{(5)}$
	PVLMD$^{(5)}$	VLMN$^{(1)}$
	ESPR$^{(5)}$	ESPR$^{(22)}$
	VLMD$^{(22)}$	ESPR$^{(5)}$

Panel B: Top Five Variables for Jump Standard Deviation ($\hat{\delta}_t$).

	Significance	Adjusted R^2
RV-Based Model	ESPR$^{(1)}$	VLMD$^{(22)}$
	VLMD$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	PESPR$^{(22)}$
	ESPR$^{(5)}$	VLMD$^{(1)}$
	VLMD$^{(22)}$	ESPR$^{(5)}$
BV-Based Model	ESPR$^{(1)}$	VLMD$^{(22)}$
	VLMD$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	PESPR$^{(22)}$
	ESPR$^{(5)}$	VLMD$^{(1)}$
	PVLMD$^{(5)}$	ESPR$^{(22)}$
RV-BV-Based Model	ESPR$^{(1)}$	VLMD$^{(22)}$
	VLMD$^{(1)}$	VLMD$^{(5)}$
	ESPR$^{(5)}$	PESPR$^{(22)}$
	VLMN$^{(5)}$	VLMD$^{(1)}$
	QSPR$^{(22)}$	ESPR$^{(5)}$

Panel C: Top Five Variables for Jump Standard Deviation ($\hat{\tau}_t$).

	Significance	Adjusted R^2
RV-Based Model	VLMD$^{(1)}$	VLMD$^{(22)}$
	ESPR$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	VLMD$^{(1)}$
	ESPR$^{(5)}$	PQSPR$^{(22)}$
	QSPR$^{(22)}$	ESPR$^{(5)}$
BV-Based Model	VLMD$^{(1)}$	VLMD$^{(22)}$
	ESPR$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	VLMD$^{(1)}$
	ESPR$^{(5)}$	PESPR$^{(22)}$
	QSPR$^{(22)}$	PESPR$^{(5)}$
RV-BV-Based Model	VLMD$^{(1)}$	VLMD$^{(22)}$
	ESPR$^{(1)}$	VLMD$^{(5)}$
	VLMD$^{(5)}$	VLMD$^{(1)}$
	ESPR$^{(5)}$	PQSPR$^{(22)}$
	QSPR$^{(22)}$	ESPR$^{(5)}$

There are 180 regression models for each liquidity measure (10 companies and 18 jump tests). We count the number of times each statistics is significant, and report the average improvement on the adjusted R^2 when comparing to the benchmark model for each measure. The significance is determined by the Newey-West corrected t-statistic with $\alpha = 0.05$. The top five measures under these two criteria are presented in this table.

Table 6.5: **Estimated Coefficients and p-values for Selected Liquidity Measures under the λ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressors	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
$VLMD^{(1)}$	-0.0010	0.0003	-0.0003	0.0005	-0.0002	-0.0011	-0.0000	-0.0007	-0.0008	**-0.0095**
	(0.1949)	(0.7476)	(0.3187)	(0.3980)	(0.8669)	(0.4390)	(0.9525)	(0.2969)	(0.2649)	(0.0000)
$QSPR^{(1)}$	0.0536	0.0511	**3.4598**	**1.0330**	-1.0109	-0.1558	0.0394	-0.4977	**-3.4767**	**-10.9574**
	(0.5136)	(0.9261)	(0.0000)	(0.0011)	(0.0653)	(0.7545)	(0.8561)	(0.3399)	(0.0326)	(0.0116)
$VLMD^{(5)}$	0.0020	0.0011	0.0000	0.0003	0.0010	-0.0042	-0.0014	0.0038	-0.0023	-0.0070
	(0.6400)	(0.8065)	(0.9871)	(0.8984)	(0.8558)	(0.5973)	(0.6710)	(0.3811)	(0.5690)	(0.3588)
$ESPR^{(5)}$	**0.6413**	-0.2432	**-3.8463**	-1.2643	-0.1576	-2.7218	-0.8877	**1.9022**	**3.6822**	8.0736
	(0.0033)	(0.7806)	(0.0000)	(0.0773)	(0.8905)	(0.2713)	(0.4120)	(0.0010)	(0.0241)	(0.0672)
$VLMD^{(22)}$	**-0.0261**	-0.0044	0.0060	-0.0035	0.0038	-0.0161	0.0069	-0.0112	0.0029	**-0.0579**
	(0.0001)	(0.7259)	(0.0795)	(0.4653)	(0.6576)	(0.4771)	(0.3305)	(0.1387)	(0.6837)	(0.0006)
Adj. R^2 Increase	67%	2%	11%	2%	6%	5%	5%	16%	2%	50%

This table reports the estimated regression coefficients for selected variables under the λ-HAR-RV-LIQ model using $Z_{TP,t}$ for each company. The p-values of Newey-West corrected t-statistics are given in brackets. The coefficients are in bold when they are significant at 95% confidence level. The same liquidity measures are selected for all companies. The changes in the adjusted R^2 as compared to the benchmark models are given in the last row.

Figure 6.7 shows that adding multiple liquidity measures increase the adjusted R^2. It reports the results of the average adjusted R^2 for each jump test (Panel A) and for each company (Panel B) of the jump intensity multivariate regressions. Liquidity measures appear to contain relevant information on the jump intensity, with an average increase of the adjusted R^2 of about 20%. All models explain about 22% to 30% of the jump intensity variation under different jump tests. This improvement obtained by adding liquidity measures is very impressive because the benchmark models only explain 4% to 10%. The results vary from company to company, with CI being the one for which we observe the largest increase (61%) when having the liquidity measures. The variation among jump tests is relatively small.

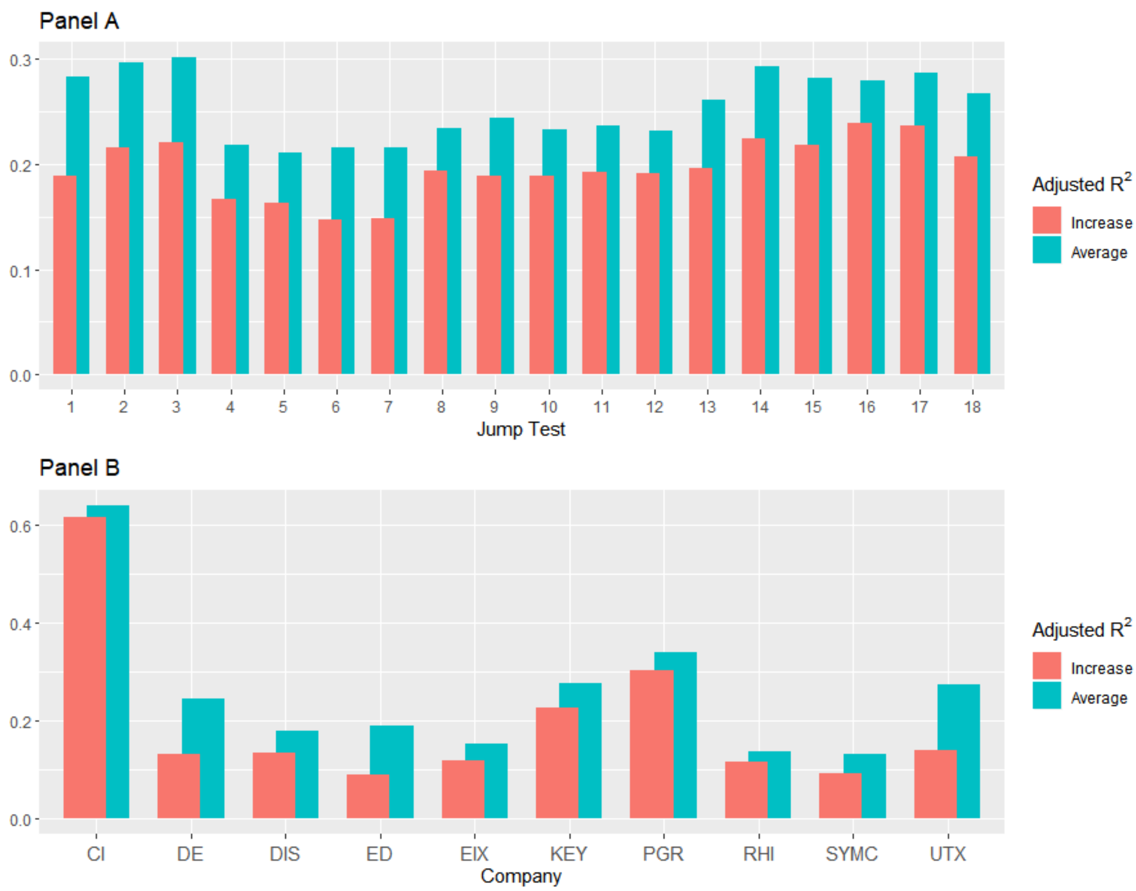

Figure 6.7: **Performance Comparison for the RV-Based Jump Intensity Model.** Panel A illustrates average adjusted R^2s, and the increase in the adjusted R^2s for RV-based jump intensity ($\hat{\lambda}_t$) model for each jump test. Panel B illustrates that same items, but for each company. See Table 6.2 for the names of jump tests that associated to the numbers in this figure.

The estimated coefficients and the p-values for the λ-HAR-RV-LIQ models using $Z_{TP,t}$ are presented in Table 6.5. At least one liquidity measure is determined to be statistically significant for all companies, except for DE, EIX, PGR, and RHI. We mentioned in the previous section that measures with monthly lag are the least significant among all lags under univariate regressions, but under the multivariate model these might become significant when they are combined with other liquidity measures (e.g., see VLMD$^{(22)}$ for CI and KEY). Moreover, the coefficients for VLMD$^{(1)}$ are all negative except for ED. This indicates that the daily volume measures are negatively related to the occurrence of jumps: higher the volume, lower the probability for having jumps. However, this is not true for VLMD$^{(5)}$ and VLMD$^{(22)}$. As the weekly and monthly measures describe the trading activity of investors over longer horizons, it is possible that their response to jumps are in the opposite direction of those of the day traders. On the other hand, half of the coefficients associated with the daily spread measure QSPR$^{(1)}$ are positive and the other half is negative. Therefore, we cannot find a general relationship between the spread measures and the jump intensity rate. These relationships are firm-specific.

The benchmark models explain about 42% of the variation in the data while estimating the jump standard deviation. A 7% increase on the adjusted R^2 is observed after adding the liquidity measures, on average. Unlike the fluctuation we see for jump intensity, the jump standard deviation models have relatively stable R^2 increases. Figure 6.8 reports the results of the average adjusted R^2 for each jump test (Panel A) and for each company (Panel B) of the multivariate regressions based on the standard deviation.

The results for the jump component standard deviation are similar to those reported for the jump standard deviation. We see an increase of about 12% when including the liquidity measures. Figure C.1 in Appendix C reports the results of the average adjusted R^2 for each jump test test (Panel A) and for each company (Panel B) of the regressions involving the standard deviation of the jump component.

The estimated coefficients and p-values for δ_t and τ_t using Z_{TP_t} as the jump test are given in Tables C.1 and C.2 of Appendix C. In contrast to the intensity model, we observed that only half the companies have significant liquidity measures when using $\hat{\delta}_t$ as the response, and this number drops to four for $\hat{\tau}_t$. This phenomenon implies that the the benchmark variables (i.e., RV at different time horizons) are more useful in understanding the jump size variation, while the liquidity measure are less relevant. In fact, the filtered jumps are essentially a combination of RV, BV and the jump test result. Therefore, we would expect that the variation in the jump size can be captured by just using the benchmark models. Moreover, despite the variable significance of liquidity measures, we do observe that they add extra explanatory power on top of the benchmark models. It is worth investigating in the future if a transformation of these spread measures or other liquidity proxies can provide better results. We leave this question for future research.

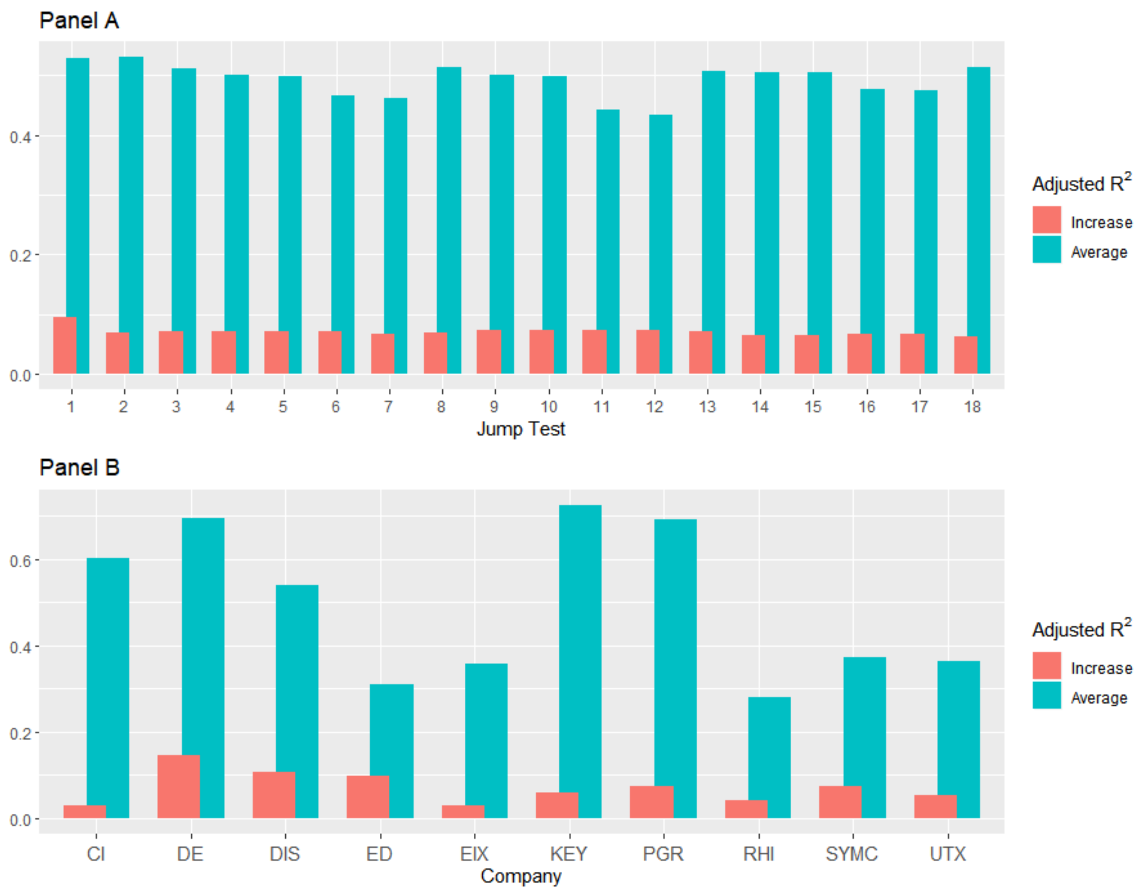

Figure 6.8: **Performance Comparison for the *RV*-Based Jump Standard Deviation Model.**
Panel A illustrates average adjusted R^2s, and the increase in the adjusted R^2s for RV-based jump standard deviation ($\hat{\delta}_t$) model for each jump test. Panel B illustrates that same items, but for each company. See Table 6.2 for the names of jump tests that associated to the numbers in this figure.

6.3.4 One-Period-Ahead Model

We apply the same methodology to one-period-ahead forecast, where the response is set to the one-period-ahead jump parameters (252 trading days in our case as per Equation 5.6). The liquidity measures again provide us significant explanatory power for all models. The average increase in the adjusted R^2 (across all jump tests and companies) is 28%, 13%, and 16% for $\hat{\lambda}_{t+252}$, $\hat{\delta}_{t+252}$, and $\hat{\tau}_{t+252}$, respectively. The forecast performance while using $Z_{TP,t}$ as the jump test is reported in Table 6.6. The comparison is based on the changes in the root mean square error (RMSE), the changes in the mean absolute error (MAE), and the increase in the adjusted R^2. The changes in RMSE and MAE are computed using the following formula:

$$\text{Relative Change} = 100\% \times \left(\text{Error}_{\text{with liquidity}} - \text{Error}_{\text{benchmark}}\right) / \text{Error}_{\text{benchmark}}. \quad (6.1)$$

Table 6.6: **One-Period-Ahead In-Sample Performance.**

Panel A: Jump Intensity Model ($\hat{\lambda}_{t+252}$).										
	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
RMSE	-31%	-8%	-13%	-1%	-7%	-30%	-4%	-25%	-8%	-29%
MAE	-30%	-11%	-17%	-1%	-8%	-31%	-5%	-24%	-9%	-29%
Adj. R^2 Increase	52%	15%	23%	1%	13%	51%	8%	44%	14%	46%
Panel B: Jump Standard Deviation Model ($\hat{\delta}_{t+252}$).										
	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
RMSE	-2%	-20%	-3%	-2%	-12%	-6%	-2%	-1%	-17%	-8%
MAE	0%	-26%	-3%	-3%	-9%	-8%	-2%	-2%	-21%	-6%
Adj. R^2 Increase	2%	23%	4%	3%	17%	8%	3%	2%	24%	9%
Panel C: Standard Deviation of Jump Component Model ($\hat{\tau}_{t+252}$).										
	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
RMSE	-2%	-13%	-1%	-4%	-8%	-8%	-1%	-1%	-21%	-7%
MAE	-1%	-20%	-2%	-9%	-3%	-10%	-3%	-2%	-22%	-4%
Adj. R^2 Increase	2%	17%	2%	6%	12%	9%	2%	2%	28%	9%

This table reports one-period-ahead in-sample performance of the $\hat{\lambda}_t$, $\hat{\delta}_t$, and $\hat{\tau}_t$ RV-based models in Panels A, B and C, respectively. The measures are compared to those of the benchmark models are given in percentage. A negative sign means a decrease in the RMSE or the MAE.

Similar to our multivariate regressions, the jump intensity models provide us large improvements when liquidity measures are included. Again, CI has the best fit among all companies: its RMSE decreases by 31% and its adjusted R^2 increases by 52% when comparing to the benchmark model. ED, on the other hand, has immaterial improvements among all firms: the RMSE only decreases by 1%, and the djusted R^2 increases by 1%.

The estimated coefficients and p-values for the one-period-ahead λ-HAR-RV-LIQ model using Z_{TP_t} are given in Table 6.7. Three additional companies have at least one significant liquidity measures than what we had in our multivariate regressions. The two spread measures QSPR$^{(1)}$ and ESPR$^{(5)}$ are determined to be significant seven and five times, respectively. This may imply that the spread measures is more useful in making forecast on the jump intensity. Moreover, the explanatory improvement obtained when using liquidity measures are clear across all companies and jump tests.

Three companies' adjusted R^2 improved by more than 10% for the one-period-ahead δ and τ models: DE, EXI, and SYMC. These two sets of models share a similar performance again due to how $\hat{\delta}_t$ and $\hat{\tau}_t$ were constructed. The extra explanatory power provided by the liquidity measures is not as strong as that obtained for the intensity models. Moreover, the significance of the liquidity measures is rather weak. There are two potential reasons for this. First, as explained in previous section, we expect the benchmark models to capture most of variation in the jump size variance due to the construction of our responses. Therefore,

Table 6.7: **Estimated Coefficients and p-values for the One-Period-Ahead λ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressor	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
$VLMD^{(1)}$	-0.0012	0.0002	-0.0005	0.0002	0.0002	**-0.0031**	0.0001	-0.0003	0.0006	-0.0062
	(0.1152)	(0.7241)	(0.1161)	(0.7038)	(0.7325)	(0.0191)	(0.8742)	(0.7097)	(0.3078)	(0.0314)
$QSPR^{(1)}$	**-0.3023**	**0.7980**	**6.4872**	0.4756	**1.7078**	**-1.9662**	0.1302	**1.0973**	**-3.6933**	-0.3162
	(0.0045)	(0.0000)	(0.0000)	(0.1239)	(0.0013)	(0.0000)	(0.6231)	(0.0000)	(0.0001)	(0.9345)
$VLMD^{(5)}$	0.0002	-0.0018	-0.0003	0.0011	-0.0020	-0.0024	-0.0015	0.0018	0.0012	-0.0078
	(0.9573)	(0.5956)	(0.8291)	(0.5991)	(0.5645)	(0.6994)	(0.6533)	(0.6575)	(0.7050)	(0.3562)
$ESPR^{(5)}$	**1.4326**	**-0.9821**	**-6.4335**	**0.2020**	-3.4341	-0.2409	-1.4943	0.6449	**2.5371**	-6.3219
	(0.0000)	(0.0000)	(0.0000)	(0.8255)	(0.0001)	(0.8566)	(0.1428)	(0.3673)	(0.0145)	(0.3854)
$VLMD^{(22)}$	**-0.0238**	**-0.0080**	0.0051	0.0008	-0.0061	-0.0170	0.0069	**-0.0179**	0.0035	-0.0515
	(0.0029)	(0.0320)	(0.0663)	(0.8478)	(0.2695)	(0.1091)	(0.3531)	(0.0099)	(0.4498)	(0.0560)
Adj. R^2 Increase	52%	15%	23%	1%	13%	51%	8%	4%	14%	46%

This table reports the estimated regression coefficients for selected variables under the one-period-ahead λ-HAR-RV-LIQ model using $Z_{TP,t}$. The p-values for Newey-West corrected t-statistics are given in brackets. The coefficients are in bold when they are significant at 95% confidence level. The same liquidity measures are selected for all companies. The changes in the adjusted R^2 as compared to the benchmark models are given in the last row.

the liquidity measures are relatively redundant. Second, multicollinearity issue might exist in our liquidity measures, even after our variable selection process.

Chapter 7

Conclusion

Identifying jumps from high-frequency data has always been challenging as the *true* jumps are not directly observable. It is commonly accepted that suspiciously high realized variances are caused by the discrete jumps. Since the development of the detection technique that utilizes the realized variance and bipower variation (Barndorff-Nielsen and Shephard, 2004, 2006) to identify jumps, many researchers have attempted to develop their own jump tests. Most of the literature focused on identifying jumps by using the realized variation and the bipower variation, or some robust versions of these two measures.

Given the fact that trading activities trigger price moves, and liquidity describes those activities in the market, one can intuitively connect jumps to liquidity.

The main objective of this report was to examine whether the relationship between jumps and liquidity exists. In particular, we first estimated three jump distribution parameters (the jump intensity rate, the jump standard deviation, and the standard deviation of the jump component) by using common jump tests. These estimators were then treated as responses in our J-HAR-RV-LIQ models. We added liquidity measures as predictors, in addition to the realized measures. The results provided in Chapter 6 showed that the liquidity measures were able to improve the model performance by a significant amount when recovering the dynamic jump intensity rate. More specifically, the daily volume measure tended to be negatively correlated to the intensity rate. The improvement when considering the jump standard deviation and the standard deviation of the jump component were smaller, and the significance of liquidity measures were lower than those obtained with the jump intensity case.

The limitation of this report is that we do not have the *true* response. The estimated responses may not reflect the true picture of the jump dynamics, which can distort the explanatory power given by the liquidity measures. Moreover, the liquidity measures included in this report may not be robust. Each of them is still a unidimensional measure. One can consider using proxy that measures multiple dimensions of liquidity simultaneously. We leave this question for future research.

Bibliography

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens. 2001a. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics* 61:43–76.

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

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

Andersen, T. G., T. Bollerslev, and N. Meddahi. 2004. Analytical Evaluation of Volatility Forecasts. *International Economic Review* 45:1079–1110.

Andersen, T. G., D. Dobrev, and E. Schaumburg. 2012. Jump-Robust Volatility Estimation using Nearest Neighbor Truncation. *Journal of Econometrics* 169:75–93.

Bakshi, G., C. Cao, and Z. Chen. 1997. Empirical Performance of Alternative Option Pricing Models. *Journal of Finance* 52:2003–2049.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2009. Realized kernels in practice: Trades and quotes. *Econometrics Journal* 12.

Barndorff-Nielsen, O. E., and N. Shephard. 2002a. Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:253–280.

Barndorff-Nielsen, O. E., and N. Shephard. 2002b. Estimating Quadratic Variation Using Realized Variance. *Journal of Applied Econometrics* 17:457–477.

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

Barndorff-Nielsen, O. E., and N. Shephard. 2005. How Accurate is the Asymptotic Approximation to the Distribution of Realized Variance. *Identification and Inference for Econometric Models. A Festschrift in Honour of TJ Rothenberg* pp. 306–311.

Barndorff-Nielsen, O. E., and N. Shephard. 2006. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. *Journal of Financial Econometrics* 4:1–30.

Barndorff-Nielsen, O. E., N. Shephard, and M. Winkel. 2006. Limit Theorems for Multipower Variation in the Presence of Jumps. *Stochastic Processes and Their Applications* 116:796–806.

Bates, D. S. 2000. Post-'87 Crash Fears in the S&P 500 Futures Option Market. *Journal of Econometrics* 94:181–238.

Black, F. 1971. Toward a Fully Automated Stock Exchange, part I. *Financial Analysts Journal* 27:28–35.

Black, F., and M. Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81:637–654.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31:307–327.

Chordia, T., R. Roll, and A. Subrahmanyam. 2000. Commonality in Liquidity. *Journal of Financial Economics* 56:3–28.

Christensen, K., R. C. Oomen, and M. Podolskij. 2014. Fact or Friction: Jumps at Ultra High Frequency. *Journal of Financial Economics* 114:576–599.

Corsi, F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7:174–196.

Corsi, F., D. Pirino, and R. Reno. 2010. Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting. *Journal of Econometrics* 159:276–288.

Corsi, F., and R. Renò. 2012. Discrete-time Volatility Forecasting with Persistent Leverage Effect and the Link with Continuous-time Volatility Modeling. *Journal of Business & Economic Statistics* 30:368–380.

Díaz, A., and A. Escribano. 2020. Measuring the Multi-faceted Dimension of Liquidity in Financial Markets: A literature Review. *Research in International Business and Finance* 51:101079.

Domowitz, I., J. Glen, and A. Madhavan. 2001. Liquidity, Volatility and Equity Trading Costs Across Countries and Over Time. *International Finance* 4:221–255.

Engle, R. F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50:987–1007.

Epps, T. W., and M. L. Epps. 1976. The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-distributions Hypothesis. *Econometrica* 44:305–321.

Goyenko, R., C. W. Holden, and C. Trzcinka. 2009. Do Measures of Liquidity Measure Liquidity. *Journal of Financial Economics* 92:153–181.

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

Heston, S. L. 1993. A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6:327–343.

Huang, X., and G. Tauchen. 2005. The Relative Contribution of Jumps to Total Price Variance. *Journal of Financial Econometrics* 3:456–499.

Kyle, A. S. 1985. Continuous Auctions and Insider Trading. *Econometrica* 53:1315–1335.

Lee, C. M., and M. J. Ready. 1991. Inferring Trade Direction from Intraday Data. *Journal of Finance* 46:733–746.

Lee, S. S., and P. A. Mykland. 2008. Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies* 21:2535–2563.

Liu, L. Y., A. J. Patton, and K. Sheppard. 2015. Does Anything Beat 5-minute RV? A Comparison of Realized Measures Across Multiple Asset Classes. *Journal of Econometrics* 187:293–311.

Ma, R., H. D. Anderson, and B. R. Marshall. 2018. Market Volatility, Liquidity Shocks, and Stock Returns: Worldwide Evidence. *Pacific-Basin Finance Journal* 49:164–199.

Müller, U. A., M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen, and J. R. Ward. 1993. Fractals and Intrinsic Time: A Challenge to Econometricians. Technical report, Olsen & Associates, Research Institute for Applied Economics, Zürich.

Newey, W. K., and K. D. West. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies* 61:631–653.

Nguyen, G., R. Engle, M. Fleming, and E. Ghysels. 2020. Liquidity and Volatility in the US Treasury Market. *Journal of Econometrics* 217:207–229.

Ramos, H. P., and M. B. Righi. 2020. Liquidity, Implied Volatility and Tail Risk: A Comparison of Liquidity Measures. *International Review of Financial Analysis* 69:101463.

Sarr, A., and T. Lybek. 2002. Measuring Liquidity in Financial Markets. Working paper, International Monetary Fund.

Tauchen, G., and H. Zhou. 2011. Realized Jumps on Financial Markets and Predicting Credit Spreads. *Journal of Econometrics* 160:102–118.

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

Appendix A

Data Overview and Jump Parameter Estimation

Appendix A provides additional figures reporting the sample path of two volume measures (Figures A.1 and A.2) and the estimated path of the standard deviation of jump component using $Z_{TP,t}$ (Figure A.3).

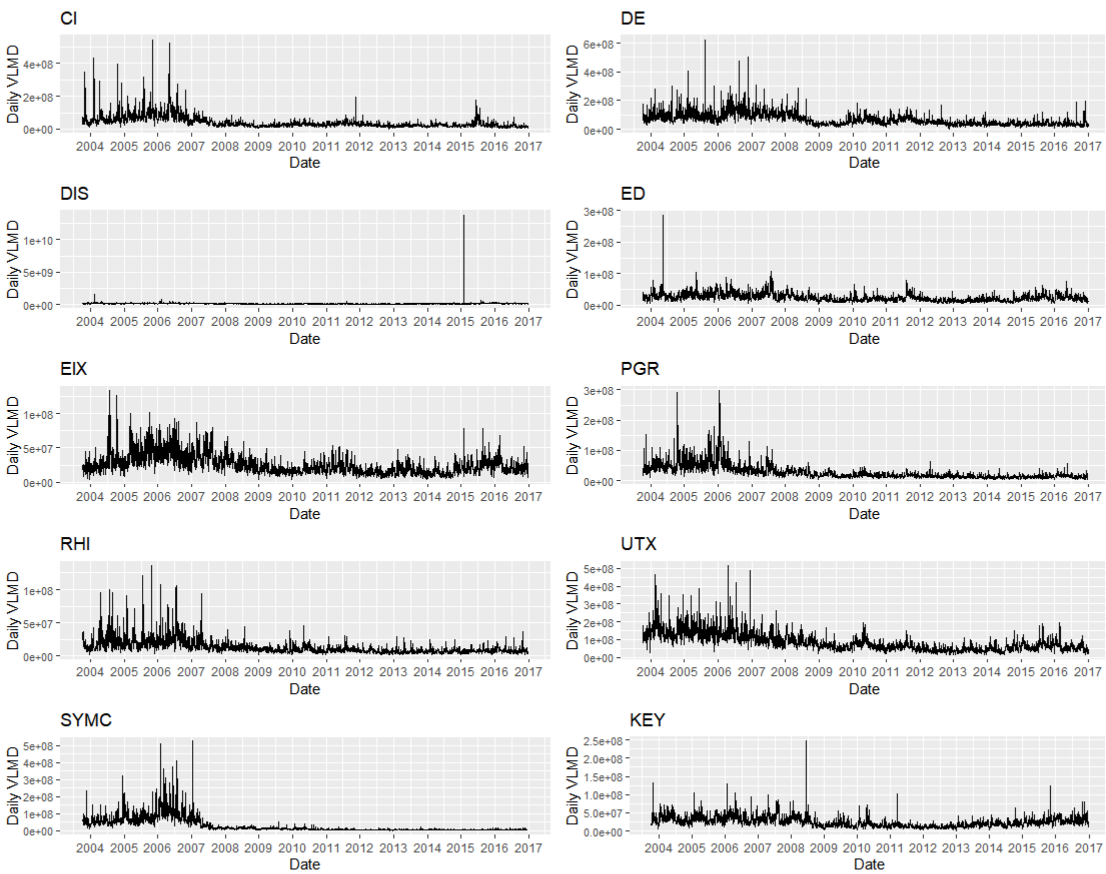

Figure A.1: **Time Series Plot of Daily Volume in Dollars.**
This figure illustrates the path of daily PVLMD between 2003 and 2016 for each selected stock. The daily measure is calculated by taking the sum of intraday PVLMD.

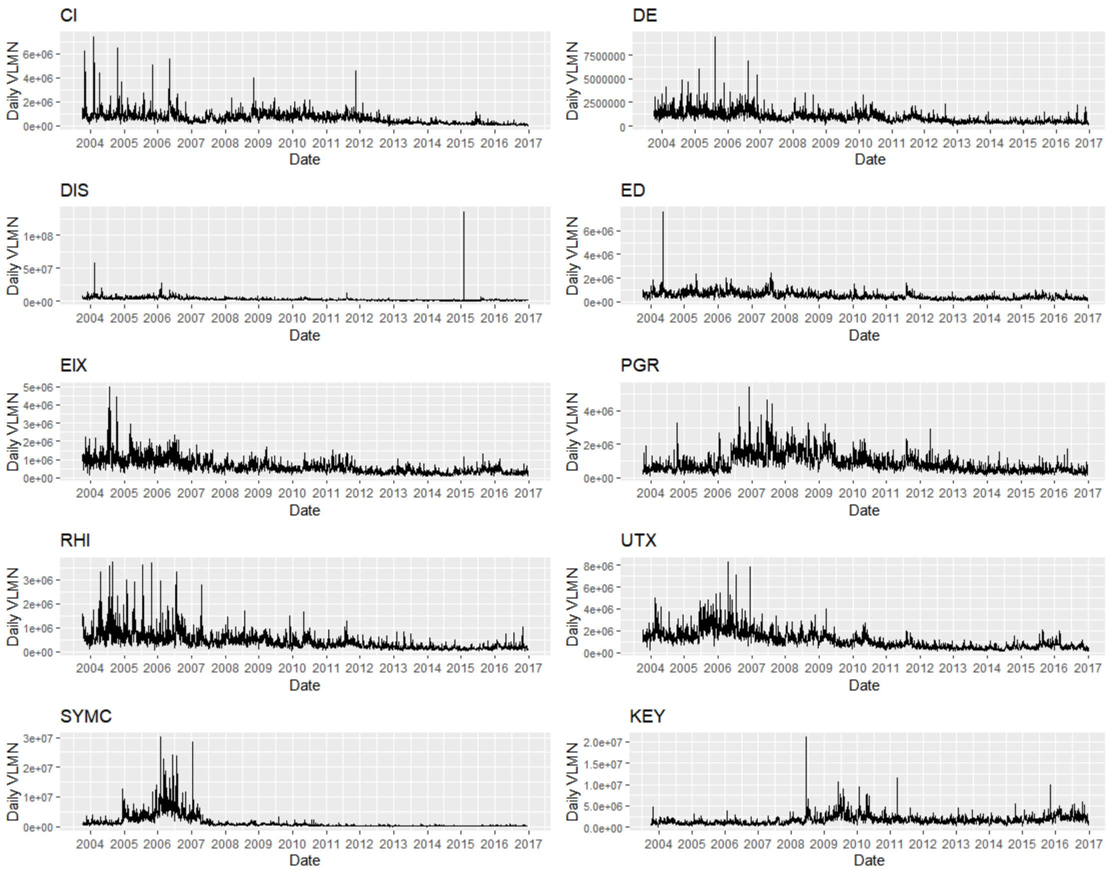

Figure A.2: **Time Series Plot of Daily Volume in Number of Shares.**
This figure illustrates the path of daily PVLMN between 2003 to 2016 for each selected stock. The daily measure is calculated by taking the sum of intraday PVLMN.

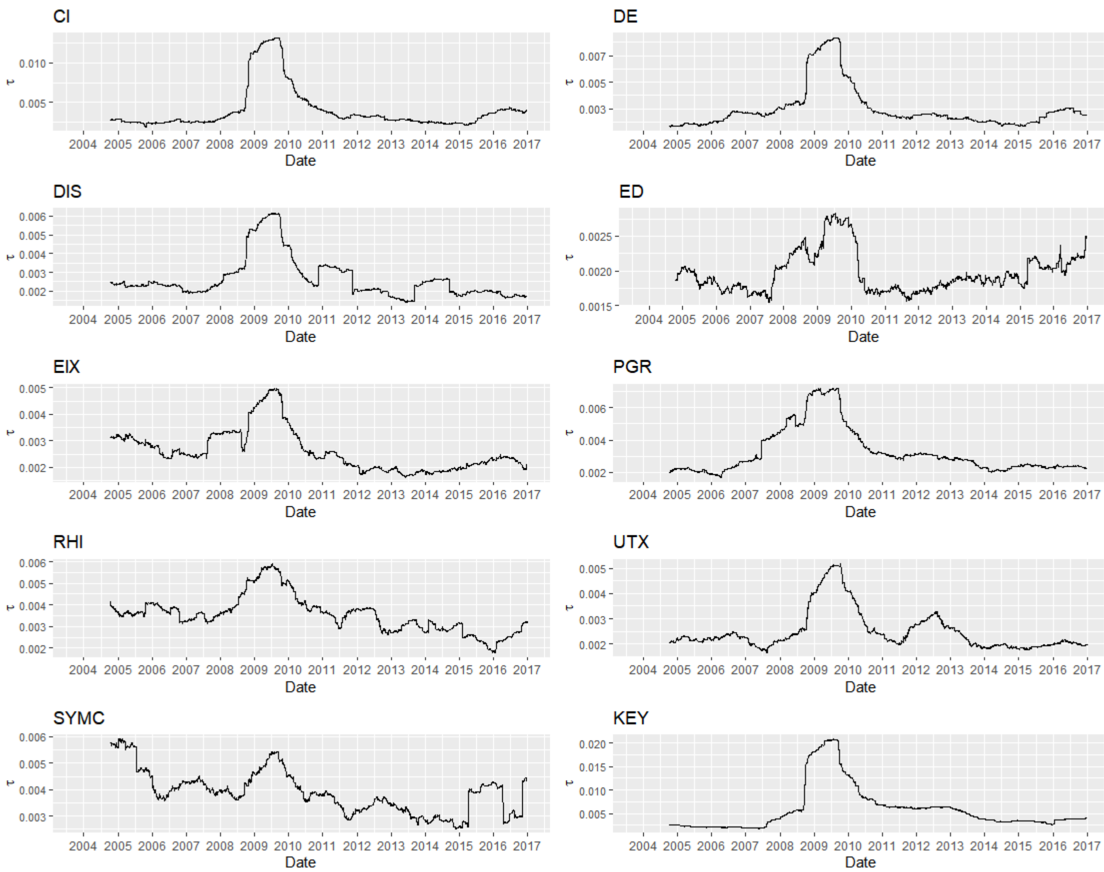

Figure A.3: **The Estimated Path of Standard Deviation of Jump Component Using $\boldsymbol{Z_{TP,t}}$.**
This figure illustrates the path of estimated standard deviation of jump component ($\hat{\tau}_t$) using the basic BNS test (i.e., $Z_{TP,t}$). The daily jump indicators are first calculated using $Z_{TP,t}$. The daily jumps are then filtered using Equation (5.2). The $\hat{\tau}_t$ is finally estimated with the filtered jumps by using a window size of $K_J = 252$.

Appendix B

Univariate Regressions

Appendix B provides additional tables for the estimated coefficients and p-values under the univariate regressions.

Table B.1 reports the estimated regression coefficients and their p-values for daily liquidity measures (lag of 1) under the univariate λ-HAR-RV-LIQ model using $Z_{TP,t}$. Table B.3 reports the estimated regression coefficients and their p-values for weekly liquidity measures (lag of 5) under the univariate λ-HAR-RV-LIQ model using $Z_{TP,t}$. Finally, Table B.3 reports the estimated regression coefficients and their p-values for monthly liquidity measures (lag of 22) under the univariate λ-HAR-RV-LIQ model using $Z_{TP,t}$.

Table B.1: **Estimated Coefficients and p-values for Daily Liquidity Measures under the Univariate λ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressor	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
QSPR$^{(1)}$	**0.6076**	-0.2163	0.2611	0.3688	**-1.1780**	**-4.3544**	-0.5622	**1.1750**	-0.3224	**-11.1635**
	(0.0000)	(0.2780)	(0.7430)	(0.3759)	(0.0000)	(0.0000)	(0.0527)	(0.0000)	(0.6870)	(0.0051)
ESPR$^{(1)}$	**0.6993**	-0.2105	-0.7315	0.0544	**-1.0094**	**-5.0250**	**-0.6831**	**1.5315**	0.2422	**-11.1694**
	(0.0000)	(0.3238)	(0.0828)	(0.8926)	(0.0094)	(0.0000)	(0.0293)	(0.0000)	(0.6573)	(0.0028)
PQSPR$^{(1)}$	**75.8911**	-2.9952	**88.6061**	28.7169	-20.6196	**106.1305**	1.5115	-18.9708	**60.9025**	**151.7052**
	(0.0000)	(0.8954)	(0.0000)	(0.1157)	(0.4268)	(0.0020)	(0.8832)	(0.5076)	(0.0065)	(0.0013)
PESPR$^{(1)}$	**91.8059**	-1.8868	**99.9620**	29.5314	-23.0836	**132.2804**	3.4308	11.6730	**56.6138**	**180.4296**
	(0.0000)	(0.9488)	(0.0000)	(0.1162)	(0.1780)	(0.0001)	(0.7567)	(0.5701)	(0.0011)	(0.0000)
VLMN$^{(1)}$	**-0.0253**	-0.0030	0.0071	-0.0002	**0.0079**	-0.0019	**0.0067**	**-0.0121**	0.0006	0.0006
	(0.0000)	(0.2414)	(0.1153)	(0.8816)	(0.0138)	(0.6391)	(0.0003)	(0.0000)	(0.6386)	(0.9759)
VLMD$^{(1)}$	**-0.0221**	-0.0026	0.0013	-0.0018	0.0043	**-0.0306**	**0.0044**	**-0.0099**	0.0004	**-0.0619**
	(0.0000)	(0.2814)	(0.3479)	(0.2046)	(0.1146)	(0.0000)	(0.0172)	(0.0000)	(0.8005)	(0.0000)
PVLMD$^{(1)}$	**-7.4997**	-0.7606	3.2101	0.0769	6.4726	**-23.1331**	**3.1142**	**-10.5492**	0.2460	**-20.8848**
	(0.0000)	(0.5223)	(0.1723)	(0.9474)	(0.0134)	(0.0000)	(0.0003)	(0.0000)	(0.6528)	(0.0019)
PVLMN$^{(1)}$	**-7.5390**	-0.7596	3.2223	0.0813	**6.4556**	**-23.1059**	**3.1125**	**-10.5501**	0.2439	**-20.9431**
	(0.0000)	(0.5226)	(0.1713)	(0.9444)	(0.0138)	(0.0000)	(0.0003)	(0.0000)	(0.657)	(0.0017)

This table reports the estimated regression coefficients for daily liquidity measures under the univariate λ-HAR-RV-LIQ model using $Z_{TP,t}$ for each company. The p-values of Newey-West corrected t-statistics are given in brackets. The coefficients are in bold when they are significant at 95% confidence level.

Table B.2: **Estimated Coefficients and p-values for Weekly Liquidity Measures under the Univariate λ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressor	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
QSPR$^{(5)}$	**0.6542**	-0.2441	0.3190	0.4742	-1.4581	**-4.6157**	-0.7550	1.4062	-0.3624	-12.1874
	(0.0002)	(0.4597)	(0.8368)	(0.6550)	(0.1567)	(0.0000)	(0.5502)	(0.1478)	(0.7003)	(0.3746)
ESPR$^{(5)}$	**0.7628**	-0.2400	-0.8213	0.0655	-1.2369	**-5.3362**	-0.9401	**1.8878**	0.2340	-12.3111
	(0.0000)	(0.4818)	(0.5408)	(0.9553)	(0.3664)	(0.0000)	(0.4683)	(0.0101)	(0.7335)	(0.3428)
PQSPR$^{(5)}$	**91.6656**	-3.4724	**93.5764**	33.9174	-24.9154	128.0913	-0.5003	-25.2438	65.2275	168.3088
	(0.0077)	(0.9117)	(0.0365)	(0.3848)	(0.5359)	(0.1581)	(0.9845)	(0.7067)	(0.5071)	(0.3306)
PESPR$^{(5)}$	**115.4767**	-2.7275	**111.9344**	37.1202	-29.5117	156.9589	1.7684	14.3139	62.0282	197.7950
	(0.0021)	(0.9473)	(0.0303)	(0.3650)	(0.3739)	(0.0605)	(0.9492)	(0.7441)	(0.4164)	(0.1773)
VLMN$^{(5)}$	**-0.0308**	-0.0038	**0.0116**	-0.0002	0.0089	-0.0024	0.0084	**-0.0134**	0.0008	0.0018
	(0.0000)	(0.5494)	(0.0137)	(0.9610)	(0.3004)	(0.8541)	(0.0522)	(0.0004)	(0.8455)	(0.9389)
VLMD$^{(5)}$	**-0.0248**	-0.0031	0.0031	-0.0021	0.0049	**-0.0339**	0.0053	**-0.0112**	0.0005	**-0.0722**
	(0.0001)	(0.5614)	(0.3692)	(0.6100)	(0.4716)	(0.0000)	(0.1391)	(0.0014)	(0.9085)	(0.0000)
PVLMD$^{(5)}$	**-9.6494**	-1.0208	**8.7958**	0.0878	8.3100	**-31.2289**	**4.9483**	**-12.5983**	0.3564	**-35.1995**
	(0.0001)	(0.7448)	(0.0458)	(0.9847)	(0.3076)	(0.0000)	(0.0375)	(0.0040)	(0.8404)	(0.0026)
PVLMN$^{(5)}$	**-9.7425**	-1.0102	**8.7629**	0.1070	8.3063	**-31.0649**	**4.9845**	**-12.6715**	0.3628	**-36.1519**
	(0.0006)	(0.7428)	(0.0474)	(0.9816)	(0.3004)	(0.0000)	(0.0447)	(0.0020)	(0.8416)	(0.0041)

This table reports the estimated regression coefficients for weekly liquidity measures under the univariate λ-HAR-RV-LIQ model using $Z_{TP,t}$ for each company. The p-values of Newey-West corrected t-statistics are given in brackets. The coefficients are in bold when they are significant at 95% confidence level.

Table B.3: **Estimated Coefficients and p-values for Monthly Liquidity Measures under the Univariate λ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressor	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
$QSPR^{(22)}$	**0.6889**	-0.2834	0.3225	0.4952	-1.7670	**-4.8608**	-1.1081	1.5713	-0.4376	-13.9181
	(0.0556)	(0.6323)	(0.8965)	(0.7942)	(0.5326)	(0.0000)	(0.7006)	(0.3007)	(0.6624)	(0.8891)
$ESPR^{(22)}$	**0.8070**	-0.2894	-0.9540	-0.0188	-1.4930	**-5.6470**	-1.3653	2.1069	0.1678	-14.2144
	(0.0489)	(0.6214)	(0.7093)	(0.9939)	(0.6543)	(0.0001)	(0.6615)	(0.1534)	(0.8268)	(0.8601)
$PQSPR^{(22)}$	102.5382	-2.7171	93.6384	32.8686	-27.5982	143.3240	-13.7391	-31.2412	67.8418	189.5213
	(0.4524)	(0.9433)	(0.2242)	(0.5274)	(0.6040)	(0.5840)	(0.7004)	(0.7579)	(0.6369)	(0.3761)
$PESPR^{(22)}$	130.4072	-3.6013	112.6725	34.7105	-35.7578	171.3107	-15.6959	7.5662	62.5181	216.1732
	(0.4072)	(0.9429)	(0.3302)	(0.5413)	(0.5312)	(0.4770)	(0.7128)	(0.9174)	(0.5642)	(0.3577)
$VLMN^{(22)}$	**-0.0345**	-0.0043	**0.0138**	-0.0005	0.0092	-0.0021	0.0094	-0.0144	0.0014	0.0033
	(0.0370)	(0.7583)	(0.0319)	(0.9587)	(0.5128)	(0.9525)	(0.3359)	(0.0723)	(0.8648)	(0.9274)
$VLMD^{(22)}$	-0.0264	-0.0035	0.0059	-0.0024	0.0049	-0.0362	0.0060	-0.0122	0.0010	**-0.0760**
	(0.0657)	(0.7751)	(0.2727)	(0.7913)	(0.7281)	(0.0848)	(0.4007)	(0.1060)	(0.9121)	(0.0320)
$PVLMD^{(22)}$	-11.4574	-1.2136	**14.1146**	-0.1897	9.2495	**-39.2434**	6.4850	-14.2538	0.6285	**-47.0377**
	(0.1428)	(0.8518)	(0.0288)	(0.9852)	(0.5147)	(0.0415)	(0.3140)	(0.1478)	(0.8674)	(0.0461)
$PVLMN^{(22)}$	-11.5138	-1.1573	**13.9847**	-0.1676	9.4197	**-38.0583**	6.6547	-14.5581	0.6837	-53.3460
	(0.1542)	(0.8599)	(0.0266)	(0.9872)	(0.5128)	(0.0294)	(0.3447)	(0.1268)	(0.8574)	(0.4610)

This table reports the estimated regression coefficients for monthly liquidity measures under the univariate λ-HAR-RV-LIQ model using $Z_{TP,t}$ for each company. The p-values of Newey-West corrected t-statistics are given in brackets. The coefficients are in bold when they are significant at 95% confidence level.

Appendix C

Multivariate Regressions

Appendix C provides additional tables and figures for the multivariate regressions.

Figure C.1 illustrates average adjusted R^2s and the increase in the adjusted R^2s for the RV-based jump component standard deviation ($\hat{\tau}_t$) model. Table C.1 reports the estimated regression coefficients and p-values for selected liquidity measures under the δ-HAR-RV-LIQ model using $Z_{TP,t}$. Finally, Table C.2 reports the estimated regression coefficients and p-values for selected liquidity measures under the τ-HAR-RV-LIQ model using $Z_{TP,t}$.

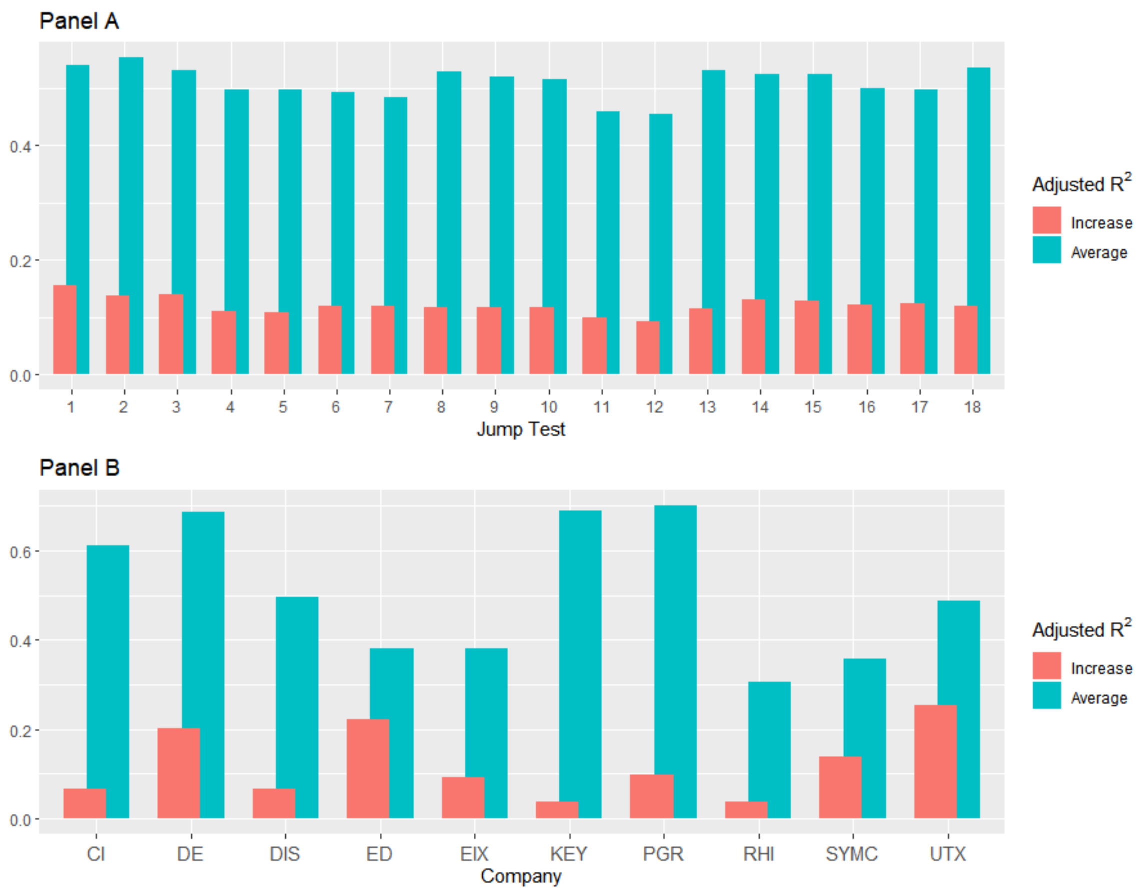

Figure C.1: **Performance Comparison for the *RV*-Based Jump Component Standard Deviation Model.**
Panel A illustrates average adjusted R^2s and the increase in adjusted R^2s for the RV-based jump component standard deviation ($\hat{\tau}_t$) model for each jump test. Panel B illustrates that same items, but for each company. See Table 6.2 for the names of jump tests that associated to the numbers in this figure.

Table C.1: **Estimated Coefficients and p-values for Selected Liquidity Measures under the δ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressor	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
$\mathrm{ESPR}^{(1)}$	-0.0082	-0.0105	-0.0072	-0.0019	**-0.0245**	-0.0090	-0.0058	**-0.0151**	-0.0196	-0.0173
	(0.2302)	(0.2605)	(0.6141)	(0.7429)	(0.0052)	(0.4824)	(0.5057)	(0.0031)	(0.6220)	(0.8736)
$\mathrm{ESPR}^{(1)}$	-0.0000	-0.0001	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0001
	(0.5969)	(0.1419)	(0.9568)	(0.5188)	(0.3153)	(0.7487)	(0.8389)	(0.4190)	(0.9286)	(0.4551)
$\mathrm{VLMD}^{(1)}$	0.0001	0.0000	-0.0000	-0.0001	-0.0001	0.0001	-0.0001	-0.0001	0.0001	0.0008
	(0.8110)	(0.9721)	(0.9347)	(0.3193)	(0.6812)	(0.4721)	(0.5713)	(0.4719)	(0.8572)	(0.4852)
$\mathrm{ESPR5}^{(5)}$	-0.0319	-0.0598	-0.1079	-0.0531	-0.0747	**-0.0125**	**-0.1134**	-0.1242	0.1953	0.0215
	(0.1158)	(0.1725)	(0.1770)	(0.3997)	(0.3767)	(0.7888)	(0.0239)	(0.0083)	(0.0267)	(0.9737)
$\mathrm{VLMD}^{(22)}$	-0.0009	**-0.0014**	-0.0006	-0.0002	-0.0001	-0.0006	0.0000	**-0.0007**	0.0003	-0.0020
	(0.1287)	(0.0175)	(0.3029)	(0.3853)	(0.7545)	(0.3103)	(0.9979)	(0.0477)	(0.6862)	(0.2582)
Adj. R^2 Increase	6%	17%	10%	12%	6%	3%	7%	22%	9%	1%

This table reports the estimated regression coefficients for selected liquidity measures under the δ-HAR-RV-LIQ model using $Z_{TP,t}$. The p-values of Newey-West corrected t-statistics are given in brackets. The same liquidity measures are selected for all companies. The coefficients are in bold when they are significant at 95% confidence level. The changes in the adjusted R^2 as compared to the benchmark models are given in the last row.

Table C.2: **Estimated Coefficients and p-values for Selected Liquidity Measures under the τ-HAR-RV-LIQ Model Using $Z_{TP,t}$.**

Regressor	CI	DE	DIS	ED	EIX	PGR	RHI	UTX	SYMC	KEY
$VLMD^{(1)}$	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000
	0.5895	0.3708	0.9301	0.2908	0.2521	0.4877	0.5420	0.4291	0.8396	0.5176
$ESPR^{(1)}$	-0.0000	-0.0000	-0.0000	-0.0000	-0.0001	-0.0000	-0.0000	**-0.0000**	-0.0001	-0.0005
	0.3899	0.2766	0.5407	0.7988	0.0128	0.7864	0.5051	0.0141	0.6869	0.7234
$VLMD^{(5)}$	0.0000	0.0000	-0.0000	-0.0000	-0.0000	0.0000	-0.0000	-0.0000	0.0000	0.0000
	0.7200	0.7071	0.9698	0.2849	0.6081	0.5576	0.5454	0.6696	0.9360	0.5386
$ESPR^{(5)}$	-0.0002	-0.0003	-0.0002	-0.0001	-0.0003	-0.0000	**-0.0004**	-0.0003	**0.0008**	-0.0009
	0.2894	0.1808	0.3715	0.3027	0.2352	0.8215	0.0220	0.1009	0.0053	0.8999
$VLMD^{(22)}$	-0.0000	**-0.0000**	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	0.0000	-0.0000
	0.1306	0.0507	0.3702	0.1905	0.5475	0.1016	0.8966	0.0886	0.6202	0.3675
Adj. R^2 Increase	5%	21%	7%	20%	10%	9%	5%	23%	13%	2%

This table reports the estimated regression coefficients for selected liquidity measures under the τ-HAR-RV-LIQ model using $Z_{TP,t}$. The p-values of Newey-West corrected t-statistics are given in brackets. The same liquidity measures are selected for all companies. The coefficients are in bold when they are significant at 95% confidence level. The changes in the adjusted R^2 as compared to the benchmark models are given in the last row.