ESSAYS ON HIGH-FREQUENCY FINANCIAL ECONOMETRICS

SHOUWEI LIU

SINGAPORE MANAGEMENT UNIVERSITY

2014

ESSAYS ON HIGH-FREQUENCY FINANCIAL ECONOMETRICS

by
Shouwei Liu

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Yiu-Kuen Tse (Supervisor/Chair)
Professor of Economics
Singapore Management University

Anthony S Tay
Associate Professor of Economics
Singapore Management University

Aurobindo Ghosh
Assistant Professor of Finance
Singapore Management University

Zhenlin Yang
Associate Professor of Economics & Statistics
Singapore Management University

Singapore Management University
2014

Abstract

Essays on High-frequency Financial Econometrics

Shouwei Liu

My dissertation consists of three essays which contribute new theoretical and empirical results to Volatility Estimation and Market Microstructure theory as well as Risk Management.

Chapter 2 extends the ACD-ICV method proposed by Tse and Yang (2012) for the estimation of intraday volatility of stocks to estimate monthly volatility. We compare the ACD-ICV estimates against the realized volatility (RV) and the generalized autoregressive conditional heteroskedasticity (GARCH) estimates. Our Monte Carlo experiments and empirical results on stock data of the New York Stock Exchange show that the ACD-ICV method performs very well against the other two methods. As a 30-day volatility predictor, the Chicago Board Options Exchange volatility index (VIX) predicts the ACD-ICV volatility estimates better than the RV estimates. While the RV method appears to dominate the literature, the GARCH method based on aggregating daily conditional variance over a month performs well against the RV method.

Chapter 3 propose to model the aggregate trade volume of stocks in a quote-driven (specialist) market using a compound Poisson distribution. Trades are assumed to be initiated by either informed or uninformed traders. Our model treats trade volume endogenously and calibrates two measures of informed trading: relative frequency of informed trading and relative volume of informed trading. Empirical analysis of daily volatility estimates of 50 NYSE stocks shows that trade volume initiated by informed traders increase volatility, while trade volume initiated by un-

informed traders reduce volatility. However, for both informed and uninformed traders, the disaggregated effect of trade frequency is to increase volatility. Our results also confirm that trade frequency dominates trade volume and trade size in affecting volatility. Yet trade volume and trade size have incremental information for volatility beyond that exhibited in trade frequency.

Chapter 4 propose to estimate the intraday Value at Risk (IVaR) for stocks using real-time transaction data. Transaction data filtered by price durations are modeled employing a two-state asymmetric autoregressive conditional duration (AACD) model, and the IVaR is computed using Monte Carlo simulation. Empirical analysis of New York Stock Exchange (NYSE) stocks show that IVaR estimated using the AACD approach track closely to those using the Dionne, Duchesne and Pacurar (2009) and Giot (2005) methods. Backtesting results show that our method performs the best among other methods.

Table of Contents

Acknowledgements

I would like to express my gratitude to all those who helped me during the writing of this thesis. I highly appreciate my supervisor Professor Tse Yiu-Kuen, who has motivated me and walked me through all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

Also, my gratitude is devoted to Professor Anthony S Tay and Professor Aurobindo Ghosh as well as Professor Zhenlin Yang, the committee members during the oral defense. Thanks for having reading a draft of this thesis and having made their precious comments and suggestions. As for the left errors, the responsibility for the text rests completely upon the author.

Last my thanks would go to my beloved parents for their loving consideration and great confidence in me all through these years. I also owe my sincere gratitude to my friends and my fellow classmates who gave me their help and time in listening to me and helping me work out my problems during the difficult course of the thesis.

Chapter 1 Introduction

Many studies in the empirical finance literature on the risk-return relationship involve the estimation of the monthly return volatility of stocks. Higher frequency or daily data are usually not used because returns are typically rather noisy, rendering the risk-return relationship in high frequency difficult to establish. Furthermore, many studies examine the effects of macroeconomic variables on asset pricing, and these variables are only available monthly or quarterly. For example, Schwert (1989) constructed vector autoregression models involving monthly data of short-term interest rates, long-term yields of high-quality and medium-quality bonds, inflation rates and industrial production to analyze the dynamic structure of stock volatility. For some recent studies requiring estimates of monthly stock volatility, see Goyal and Santa-Clara (2003), Bali, Cakici, Yan and Zhang (2005), Guo and Savickas (2008), Ludvigson and Ng (2007), Jiang and Tian (2010) and Zhang (2010).

Chapter 2 compare the performance of the RV, GARCH and ACD-ICV methods using Monte Carlo (MC) experiments and empirical data from the New York Stock Exchange (NYSE). Our MC results show that the ACD-ICV method outperforms the RV method in giving lower root mean-squared error. Indeed, it turns out that the GARCH method performs better than the RV method in the MC experiments. We also examine the use of the Chicago Board Options Exchange (CBOE) volatility index (VIX) as a predictor of the volatility for the next 30 days estimated by the ACD-ICV, RV and GARCH methods using the S&P500 index. Our results show that VIX predicts the 30-day ACD-ICV volatility estimates better than the RV estimates.

The effects of the trade volume of a stock on its return and volatility have been studied extensively in the finance literature. Wang (1993, 1994) examined the theoretical links between trade volume and return dynamics. He showed that informed trading and uninformed trading have asymmetric effects on stock returns. Andersen (1996) studied the volatility-volume relationship using the mixture-of-distribution hypothesis (MDH). He argued that his enhanced model, in which trades may originate from informed traders or liquidity traders, outperforms the standard MDH model. While Andersen (1996) assumed that liquidity trading has no effect on volatility, Li and Wu (2006) relaxed this assumption and postulated that liquidity trading can lower return volatility. They showed that the positive relationship between volume and volatility is primarily driven by informed trading.

The seminal work of Jones, Kaul and Lipson (1994) highlighted the importance of studying the effects of trade frequency and trade size on return volatility. Their results show that trade size has no information content beyond that contained in the frequency of trades. Chan and Fong (2000), however, objected to this conclusion and argued for "the significance of the size of trades, beyond that of the number of trades, in the volatility-volume relation". More recently, Chan and Fong (2006) and Xu, Chen and Wu (2006) contributed to the debate. Their results are in favor of the dominance of trade frequency over trade volume in their effects on volatility. Huang and Masulis (2003) studied transactions data from the London Stock Exchange and lent support to the general conclusion of Jones, Kaul and Lipson (1994).

Chapter 3 use transaction data of trade frequency and trade size over 30-minute intervals to estimate the aggregate-volume model. We use two approaches to estimate the model. First, we use covariates as proxies for information intensity, conditional on the covariates the likelihood function of trade frequency and trade size can be obtained, from which the parameters of the model can be estimated using the MLE method. Alternatively, we may calculate the unconditional moments and cross moments of trade volume and trade frequency, treating the moments of the information intensity as unknown parameters. The parameters of the primary and secondary distributions of the compound Poisson distribution, together with the moments of the information intensity, can then be estimated using the GMM method. For each stock we calculate the daily RFIT and RVIT measures. We then study the effects of informed trade frequency, informed trade volume, uninformed trade frequency and uninformed trade volume on volatility. Our results show that the empirical relation between volume and volatility under the MLE approach is similar to the empirical relation under the GMM approach. Both the MLE and GMM approaches show that trade frequency dominates trade volume in explaining volatility, although trade volume still has incremental information for volatility in the presence of trade frequency. Rather interestingly, while informed trading volume has positive effects on volatility, uninformed trading volume has negative effects on volatility. This result is consistent with Li and Wu (2006), who find negative correlation between volatility and trade volume due to liquidity traders. However, both the informed and uninformed trading frequency has positive effect on volatility.

The global financial crisis in 2008 highlighted the need for the banks' decision makers, such as trader and heads of desks, to have real-time access to accurate information in order to make rapid and well-informed decisions, particularly during periods of market turmoil. After the Market Access Rule (MAR) came into effect, any order sent to the market must go through pre-trade risk control. Current approaches to risk management in many large investment banks are inadequate. At present, the majority of banks are relying on risk information on a daily or, at best, intraday basis. Intraday risk reports might be regarded as adequate when used as a reporting tool. However, if a trader is required to act on that risk information, as is surely desirable, information must be provided instantaneously. The ability to react to risk events on a real-time basis would give a bank or trader serious competitive advantage. Managing risk in (near) real-time becomes increasingly important for banks and traders.

Chapter 4 apply the AACD model to a two-state point process for price movements, where the two states represent an upward or a downward price movement of a pre-determined threshold δ. Following Bauwens and Giot (2003), we allow the expected duration to vary with the lagged durations, and the lagged conditional expected durations. Using an intraday Monte Carlo simulation approach, with information for the price movements before a stated time, we simulate the price movements starting from the stated time for any horizon during the trading hours. We study all index stocks of the S&P 500 traded on the NYSE for three different periods during and after the 2008 global financial crisis. Our empirical results of 30-min IVaR backtesting show that the AACD approach outperforms other IVaR evaluation methods. IVaR can be computed for any time horizon once the AACD model has been estimated without requiring new sampling and estimation when the time horizon changes, due to the the flexibility of irregularly-spaced information. 60-min IVaR backtesting results also indicate that the AACD approach performs well against other methods.

Chapter 2 Estimation of Monthly Volatility: An Empirical Comparison of Realized Volatility, GARCH and ACD-ICV Methods

2.1 Introduction

Since the seminal work of French, Schwert and Stambaugh (1987) and Schwert (1989), researchers often use the sum of the squared daily stock returns over a month (or some modifications of it) as an estimate of the monthly volatility of the stock. Later, Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001) proposed to use the sum of the squared returns of tick data to estimate intraday volatility, and called this estimate the realized volatility (RV). Since then the literature on RV has expanded very quickly, producing asymptotic results of the properties of the RV as an estimate of the integrated volatility over daily or intraday intervals. In addition, some enhanced RV estimates dealing with problems of market microstructure noise and/or price jumps have been proposed, including the subsampling technique due to Zhang, Mykland and Aït-Sahalia (2005), the bipower variation method by Barndorff-Nielsen and Shephard (2004), the realized kernel method by Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), and the duration-based RV method by Andersen, Dobrev and Schaumburg (2008).

The estimation of intraday volatility using RV methods typically requires sampling over five-minute intervals or shorter, with sampling over one- or two-minute intervals not uncommon. Given the asymptotic theories established by Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen, Hansen, Lunde and Shephard (2008) and Zhang, Mykland and Aït-Sahalia (2005) and the availability of large numbers of return observations over short durations in a trading day, the RV methods have firm theoretic underpinning as a tool for estimating intraday volatility. In contrast, in applying the RV methods to estimate monthly volatility using daily data there are only approximately 21 return observations to compute each monthly estimate. Thus, measurement errors may be a concern and may weaken the validity of the statistical inference. Clearly, an appropriate choice of the monthly volatility estimates is important for studies that use these estimates to investigate asset pricing.

Another line of research applies the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986) to estimate monthly volatility. French, Schwert and Stambaugh (1987) estimated monthly volatility using a GARCH-in-mean model, while Fu (2009) estimated monthly idiosyncratic risk using the exponential GARCH (EGARCH) model of Nelson (1991). Generally, monthly volatility can be estimated using GARCH type of models on monthly data, or using these models on daily data from which aggregates of daily conditional variances form a monthly estimate.

Recently, Tse and Yang (2012) proposed a method to estimate high-frequency volatility using the autoregressive conditional duration (ACD) model of Engle and Russell (1998), called the ACD-ICV method. They estimate high-frequency volatility (over a day or shorter intervals) by integrating the instantaneous conditional return variance per unit time obtained from the ACD models. Unlike the RV methods, which sample data over regular intervals, the ACD-ICV method samples price events based on high-frequency transaction price changes exceeding a threshold. ACD models for the durations between sequential price events are estimated using the maximum likelihood method, and the conditional variance over a given intraday interval is computed by integrating the instantaneous conditional variance over different durations within the interval. The Monte Carlo results of Tse and Yang (2012) show that the ACD-ICV method gives lower root mean-squared error than the RV methods in estimating intraday volatility. While Tse and Yang (2012) focused on the estimation of intraday volatility, in this chapter we apply the ACD-ICV method to estimate monthly volatility.

The literature so far has little to say about the choice of the estimation method for monthly volatility. While the RV approach seems to dominate the literature, the use of the GARCH type of models is not uncommon. In addition, the ACD-ICV method may be a useful alternative, as it has been shown to perform well for the estimation of intraday volatility. In this chapter we compare the performance of the RV, GARCH and ACD-ICV methods using Monte Carlo (MC) experiments and empirical data from the New York Stock Exchange (NYSE). Our MC results show that the ACD-ICV method outperforms the RV method in giving lower root mean-squared error. Indeed, it turns out that the GARCH method performs better than the RV method in the MC experiments. We also examine the use of the Chicago Board Options Exchange (CBOE) volatility index (VIX) as a predictor of the volatility for the next 30 days estimated by the ACD-ICV, RV and GARCH methods using the S&P500 index. Our results show that VIX predicts the 30-day ACD-ICV volatility estimates better than the RV estimates.

The rest of the chapter proceeds as follows. Section 2.2 summarizes the RV and GARCH estimation methods of monthly volatility studied in this chapter. Section 2.3 describes the use of the ACD-ICV method for the estimation of monthly volatility, and Section 2.4 describes the data used in the empirical study. In Section 2.5 we report some MC results on the comparison of the RV, GARCH and ACD-ICV methods. The MC study suggests that the best results for the ACD-ICV method appear to be obtained when the range of the return for defining the price event is about 0.15% to 0.35%. It also shows that the ACD-ICV method performs very well against the RV method. Section 2.6 reports our results for the estimation of monthly volatility using some empirical data from the NYSE. In Section 2.7 we examine the use VIX as a predictor of the market volatility over the next 30 days, with market volatility estimated using the ACD-ICV, RV and GARCH methods. Finally, Section 2.8 concludes.

2.2 Review of Some Monthly Volatility Estimation Methods

Volatility estimation over monthly or quarterly intervals dated back to the 1970s. Researchers in earlier work adopted the 12-month rolling standard-deviation estimate as the volatility estimate of the centered month, as in Officer (1973), Fama (1976), and Merton (1980). Schwert (1989) employed a two-step rolling regression to construct monthly volatility, which allows the conditional mean return to vary over time and allows different weights for the lagged absolute unexpected returns. Since the work of French, Schwert and Stambaugh (1987), Schwert (1989), Schwert (1990a), Schwert (1990b) and Schwert and Seguin (1990), the use of the sum of the squared daily returns over a month, called the RV method, has emerged as the most popular method for the estimation of monthly stock volatility. On the other hand, monthly volatility can also be estimated using GARCH models estimated with monthly data, or by aggregating daily conditional variances over a month estimated from GARCH models with daily data. In this section we provide a brief review of the estimation of monthly volatility using RV and GARCH methods.

2.2.1 RV Method

Let r_{ti} denote the return on day i in month t, $\bar{r}_t$ denote the average daily return in month t and N_t denote the number of trading days in month t. The basic RV estimate of monthly variance, denoted by V_R, is defined as

$$V_{RM} = \sum_{i=1}^{N_t} (r_{ti} - \bar{r}_t)^2. \tag{2.2.1}$$

It is well known that non-synchronous trading of securities causes daily portfolio returns to be autocorrelated, particularly at lag one (see Fisher (1966) and Scholes and Williams (1977)). Akgiray (1989) showed that there exists linear dependence in daily return series of market indices, and the presence of linear dependence can be attributed to various market phenomena and anomalies. The presence of a common market factor, the problem of thin trading in some stocks, the speed of information processing by market participants, and the day-of-week effects may all contribute to the observed first-order autocorrelation. Because of this autocorrelation, French, Schwert and Stambaugh (1987) proposed to estimate the variance of the monthly return as the sum of the squared daily returns plus twice the sum of the products of adjacent returns, thus resulting in the following estimate of monthly volatility

$$V_R^* = \sum_{i=1}^{N_t} r_{ti}^2 + 2 \sum_{i=1}^{N_t - 1} r_{ti} r_{t,i+1}. \qquad (2.2.2)$$

Note that in the above equation the sample mean of the return is not subtracted from the daily return, as the effect of this adjustment is usually very small. However, this return correlation adjustment is found not helpful in improving the monthly volatility estimation.

The application of V_{RM} and V_R^* using daily closing prices has been widely adopted in the literature. Apart from the simplicity of the calculation, the problem of overnight price jumps is not an issue when daily data are used. However, as monthly volatility estimates using daily data makes use of only about 21 observations for each estimate, the accuracy of the estimates may be a concern. As high-frequency data have become more easily available, we extend the use of monthly RV estimation to transaction data. For the purpose of using as much data as possible, shorter sampling intervals are preferred. However, returns over short sampling intervals may be contaminated by market microstructure noise. To balance between these two conflicting goals, we use 5-minute price data to calculate the RV. This is in contrast to Jiang and Tian (2005) and Becker, Clements and White (2007), who used 30-minute data to compute the RV.

To extend the use of equation (2.2.1) to intraday returns, we have to consider the treatment of overnight price jumps. Let x be the closing price of the stock on a trading day, y be the opening price of the stock on the next trading day, and z be the price of the stock at the 5th minute of the next trading day. We modify the calculation of RV in equation (2.2.1), without the mean correction, using two methods. In the first method, we add, for each overnight transition, the term $(\log y - \log x)^2 + (\log z - \log y)^2$ to take account of the overnight price change. In the second method, we treat trading as continuous and add the term $(\log z - \log x)^2$ only. We denote the RV estimates of the monthly volatility using these two methods by V_{R1} and V_{R2}, respectively.

2.2.2 GARCH Method

Another popular method for constructing monthly volatility is to use the GARCH model. The ARCH model proposed by Engle (1982) is well-known to capture the clustering of volatility of many economic and financial time series. Bollerslev (1986) extended the ARCH model to the GARCH model, which provides a more flexible framework to capture the dynamic structure of conditional variance.

The original GARCH specification assumes that the response of the conditional variance of a stock to a shock is symmetric with respect to the sign of the shock. Several extensions of the GARCH model, however, have been proposed to accommodate the asymmetry in the response. These include the GJR-GARCH model of Glosten, Jagannathan, Runkle (1993), the asymmetric GARCH models of Engle and Ng (1993a), the quadratic GARCH model of Sentana (1995) and the Exponential GARCH (EGARCH) model of Nelson (1991). Pagan and Schwert (1990) fitted a number of different models to monthly US stock-return data and found that the EGARCH model is the best in overall performance. Engle and Ng (1993b) also concluded that the EGARCH model does a good job in capturing the asymmetry of conditional volatilities. In this chapter we adopt the EGARCH model to estimate monthly stock volatility.

In the EGARCH model, the conditional variance, σ_t^2, is an asymmetric function of the lagged disturbances ε_{t-i}. Specifically, we have

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i g(z_{t-i}) + \sum_{j=1}^{p} \beta_j \log \sigma_{t-j}^2, \tag{2.2.3}$$

where

$$g(z_t) = \theta z_t + \gamma[|z_t| - \mathrm{E}|z_t|], \tag{2.2.4}$$

with $z_t = \varepsilon_t / \sigma_t$. As argued in Nelson (1991), the generalized error distribution (GED) is a more flexible assumption for the distribution of ε_t, because it encompasses the normality assumption as a special case, as well as many other distributions. The density function of the GED distribution with parameter $\nu > 0$ is defined as

$$f(z) = \frac{\nu \ \exp[-(\frac{1}{2})|z/\lambda|^{\nu}]}{\lambda \ 2^{1+1/\nu}\Gamma(1/\nu)}, \quad -\infty < z < \infty, 0 < \nu \leq \infty, \tag{2.2.5}$$

where $\Gamma(\cdot)$ denotes the gamma function and $\lambda \equiv [2^{(-2/\nu)}\Gamma(1/\nu)/\Gamma(3/\nu)]^{1/2}$.

In this chapter, we adopt the EGARCH(1, 1) model defined by

$$\log \sigma_t^2 = \omega + \alpha \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \log \sigma_{t-1}^2. \tag{2.2.6}$$

There are two approaches of using the EGARCH(1, 1) method to estimate monthly volatility. First, we use monthly data and calculate the estimated conditional variance $\hat{\sigma}_t^2$ and denote it by V_G^M. Second, we estimate the EGARCH(1, 1) model using daily data and compute the daily conditional variance $\hat{\sigma}_{ti}^2$, for $t = 1, \cdots, N$, over N days of the month t. We then aggregate the estimated daily conditional variances and denote it by V_G^D, so that

$$V_G^D = \sum_{t=1}^{N} \hat{\sigma}_{ti}^2. \tag{2.2.7}$$

In our study we find that V_G^D performs dramatically better than V_G^M; hence, we only report V_G^D for the GARCH measures.

In sum, we consider RV estimates V_{RM}, V_{R1} and V_{R2}, as well as EGARCH estimates V_G^D. In addition, V_{R1} and V_{R2} use 5-minute data, whereas V_{RM} and V_G^D use daily data.

2.3 Monthly Volatility Estimation using ACD Models

The ACD model was first proposed by Engle and Russell (1998) to analyze the durations of transaction data. A recent review of the literature on the ACD models and their applications to finance can be found in Pacurar (2008). Analogous to the GARCH model, which captures the clustering of volatility, the ACD model analyzes the clustering of transaction duration. The latter phenomenon describes the stylized fact that short (long) transaction durations tends to be followed by short (long) transaction durations.

Adopting the augmented ACD (AACD) model proposed by Fernandes and Grammig (2006), Tse and Yang (2012) proposed to estimate intraday volatility by integrating the instantaneous conditional variance per unit time estimated from the AACD model, resulting in the ACD-ICV method. In this section, we first review the ACD-ICV method proposed by Tse and Yang (2012), followed by an outline of the modification of this method for the estimation of monthly volatility.

2.3.1 ACD-ICV Method

Let $t_0, t_1, \cdots, t_N$ denote a sequence of times for which t_i is the time of the ith price event, to be defined below.[1] Thus, $x_i = t_i - t_{i-1}$, for $i = 1, 2, \cdots, N$, are the intervals between consecutive price events, called price duration. In Tse and Yang (2012), a price event occurs if the cumulative change in the logarithm transaction price since the last price event is at least of a preset amount δ, called the price range. Thus, from time t_{i-1} to t_i, the price changes by at least an amount δ, whether upwards or downwards. Let Φ_i be the information set upon the transaction at time t_i, and denote $\psi_i = \mathrm{E}(x_i \mid \Phi_{i-1})$, which is the conditional expectation of the transaction duration. The standardized durations, $\varepsilon_i = x_i / \psi_i$, are assumed to be independently and

[1] In this section t denotes the intraday time, not month notation t.

identically distributed positive random variables. If $\sigma^2(t\,|\,\Phi_i)$ is the instantaneous variance of the return per unit time at time t, for $t_i < t < t_{i+1}$, conditional upon the information Φ_i, the integrated conditional variance over the interval (t_i, t_{i+1}), denoted by ICV$_i$, is

$$\begin{aligned} \mathrm{ICV}_i &= \int_{t_i}^{t_{i+1}} \sigma^2(t\,|\,\Phi_i)\,dt, \\ &= \frac{\delta^2}{\psi_{i+1}} \int_{t_i}^{t_{i+1}} \lambda\left(\frac{t-t_i}{\psi_{i+1}}\right) dt, \end{aligned} \tag{2.3.1}$$

where $\lambda(\cdot)$ is the hazard function of the standardized durations ε. See the details of the above derivation in Tse and Yang (2012).

Tse and Yang (2012) proposed estimating the hazard function using a semiparametric (SP) method, which does not specify the distribution of ε. However, if ε are assumed to follow the standard exponential distribution the hazard function is constant and the integrated conditional variance in the interval (t_i, t_{i+1}) is reduced from equation (2.3.1) to the simple result

$$\mathrm{ICV}_i = \delta^2 \left[\frac{t_{i+1} - t_i}{\psi_{i+1}}\right]. \tag{2.3.2}$$

For estimating the volatility over one trading day, t_0 and t_N are the opening and closing times of the day, respectively, while $t_1, \cdots, t_{N-1}$ are the time of occurrence of the price events within the day. Thus, the integrated conditional variance of the day, denoted by ICV, is

$$\mathrm{ICV} = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\psi_{i+1}}. \tag{2.3.3}$$

The implementation of the ACD-ICV method to estimate monthly volatility depends on the price data available. If only daily data are available, the price changes over each day may be too big for the transaction durations to be precisely approximated. One remedy to overcome this difficulty is to linearly intrapolate the daily closing prices to obtain a continuous-time price function, from which the transaction durations are computed. This approach, however, is found to incur large errors in the estimation and has to be abandoned. An alternative is to resort to using higher frequency data. In this chapter we use transaction data. We ignore the overnight close of the market and treat the first trade of each day as continuously away from the last trade of the previous trading day. Hence, given the return range δ we compile $t_1, \cdots, t_{N-1}$ based on the continues transaction data as the price-event times over a month. We then apply equation (2.3.3) to estimate the integrated conditional variance of the month.

2.3.2 ACD and AACD Models

The use of equation (2.3.3) requires estimates of the conditional expected duration ψ_{i+1}. Tse and Yang (2012) employ the AACD model for this purpose.

Engle and Russell (1998) proposed the ACD(p,q) model for the analysis of transaction duration, which is defined by

$$\psi_i = \omega + \sum_{j=1}^{p} \alpha_j x_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}. \tag{2.3.4}$$

Setting $p = q = 1$, we obtain the ACD$(1,1)$ model as

$$\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}, \tag{2.3.5}$$

where α, β and $\omega \geq 0$, with $\alpha + \beta \leq 1$.

Recently, Fernandes and Grammig (2006) proposed some extensions of the ACD$(1,1)$ model, including incorporating a Box-Cox type transformation with possible asymmetry in the duration shocks. In Tse and Yang (2012), they adopt the AACD model of Fernandes and Grammig (2006), which is defined by

$$\psi_i^{\lambda} = \omega + \alpha \psi_{i-1}^{\lambda}[|\varepsilon_{i-1} - b| + c(\varepsilon_{i-1} - b)]^{\upsilon} + \beta \psi_{i-1}^{\lambda}. \tag{2.3.6}$$

The AACD model nests the ACD$(1,1)$ model as a special case and provides a more flexible model for the conditional expected duration. The parameter $\lambda > 0$ determines the shape of the transformation, with $\lambda \geq 1$ representing a convex transformation and $\lambda \leq 1$ representing a concave transformation. Asymmetric responses in duration shocks are permitted through the shift parameter b and the rotation parameter c. The shape parameter υ assumes a similar role as λ. As in the case of the ACD$(1,1)$ model, the parameters α, β and ω are assumed to be nonnegative. However, Tse and Yang (2012) find that the ICV estimates are not sensitive to the choice of ACD$(1,1)$ or AACD model, so considering estimation of monthly volatility using hundred of thousands tick data, AACD model is computationally expensive and beneficial little over ACD$(1,1)$ model, we shall use ACD$(1,1)$ model for this purpose.

Given an assumed density function $f(\cdot)$ for ε, the maximum likelihood estimates (MLE) of the parameters of the ACD equation can be computed straightforwardly. A particularly simple model is the case when ε are assumed to be standard exponential, which results in the quasi MLE (QMLE) method. As shown by Drost and Werker (2004) the QMLE method is consistent provided the conditional expected duration equation is correctly specified, regardless of the true distribution of the standardized duration. However, if the exponential distribution assumption is incorrect it may induce error in the computation of the conditional intensity function and hence the integrated conditional variance. While this issue can be resolved by employing the semiparametric (SP) method, Tse and Yang (2012) showed that the QMLE and SP methods produce very similar results for the ACD-ICV estimates. As the SP method is computationally very intensive, we shall adopt the QMLE method in this chapter.

Under the exponential assumption the ACD-ICV estimate, denoted by V_A, is given by

$$V_A = \delta^2 \sum_{i=0}^{N-1} \frac{t_{i+1} - t_i}{\hat{\psi}_{i+1}}, \tag{2.3.7}$$

where $\hat{\psi}_{i+1}$ is the QMLE of ψ_{i+1}. The choice of δ affects the fit of the ACD$(1,1)$ model for price duration, and hence the performance of V_A as an estimate of the monthly ICV. We shall vary δ from 0.15% through 0.35% in steps of 0.05%, and denote the resulting estimates by V_{Aj} for $j = 1, \cdots, 5$, respectively, to correspond to δ being 0.15%, 0.20%, 0.25%, 0.30% and 0.35%.

We compare the performance of the different estimates of monthly volatility using a MC study and empirical data from the NYSE. The MC study will throw light on the optimal choice of δ for the ACD-ICV method. In the next section we describe the data used in the empirical study.

2.4 NYSE Data

We apply the RV, GARCH and ACD-ICV methods to estimate monthly volatility using empirical data from the Trade and Quotation (TAQ) database provided through the Wharton Research Data Services (WRDS). The TAQ data files contain continuously recorded information on the trades and quotations for the securities listed on the NYSE, the American Stock Exchange (AMEX), and the National Association of Security Dealers Automated Quotation system (NASDAQ). We select ten actively traded stocks listed on the NYSE without company merger and acquisition from 2003 through 2007, with 60 months of data. The price changes due to stock splits are adjusted according to the capitalization of the company. The selected stocks and their codes are summarized in Table 2.1.

We extract stock transaction prices from 9:30 to 16:00 on each day. When no trade occurred exactly at the required end of interval, the price of the last transaction was recorded. In addition, we also record the overnight price jumps of the stocks.

2.5 Monte Carlo Study

We conduct MC experiments to compare the performance of the RV, GARCH and ACD-ICV estimates of monthly volatility. As the assumption underlying the theoretical derivation of RV in the literature is that the logarithmic stock price follows a Brownian semimartingale (BSM), we create artificial data generation processes along this line. We denote the observed price at time t by $p(t)$ and denote

$\tilde{p}(t) = \log p(t)$, which is assumed to follow the BSM

$$\tilde{p}(t) = \int_0^t \mu(t)\,dt + \int_0^t \sigma(t)\,dW(t), \tag{2.5.1}$$

where $\mu(t)$ is the instantaneous drift rate, $\sigma^2(t)$ is the instantaneous variance and $W(t)$ is a standard Brownian process. For each MC sample, we generate data over 5 years, with a total of 60 months of observations.

For the drift term $\mu(t)$ we consider two different artificial processes, which are plotted in Figure 2.1. For the variance process $\sigma^2(t)$ we consider two methods: deterministic volatility and stochastic volatility models, which will be described in the next two subsections. Given the drift term $\mu(t)$ and the variance term $\sigma^2(t)$, we generate the logarithmic price series $\tilde{p}(t)$ by the equation

$$\tilde{p}(t+\Delta t) = \tilde{p}(t) + \mu(t)\Delta t + \sigma(t)\sqrt{\Delta t}\,\varepsilon, \tag{2.5.2}$$

where $\varepsilon \sim N(0,1)$. We take Δt to be one second so that we generate price series at one-second intervals and the starting price is \$100. We further add to the series $s_B(u)$ a jump component $s_J(u)$, which is assumed to follow a Poisson process with a mean of 0.4 per five minutes. When a jump occurs, it takes value of –\$0.05, –\$0.03, \$0.03 and \$0.05 with probabilities of 0.25 each. Finally, we consider a price process consisting of a BSM and a white noise; following the definition of noise-to-signal (NSR) ratio NSR $= [\mathrm{Var}\{\varepsilon(t)\}/\mathrm{Var}\{\sigma(t)\}]^{\frac{1}{2}}$ in Tse and Yang (2012), for similarly, we set NSR = 0.6.

From the generated price series p_t we round the price to the cent and sample the rounded price by 1 cent to get the transaction price for ACD-ICV estimation; after which we also sample the transaction price at 5-minute and one-day intervals, depending on the estimation method for the monthly volatility.

2.5.1 Deterministic Volatility Model

For the deterministic instantaneous variance term $\sigma^2(t)$, we assume two artificial processes: a sinusoidal function and an empirical function. The sinusoidal function has cycles of volatility taking the form

$$\sigma(t) = 0.05\sin\left(\frac{2t\pi}{30} + \frac{\pi}{2}\right) + 0.15, \qquad t = 1, \cdots, 60, \tag{2.5.3}$$

called DV Model 1. To construct an empirical volatility process we estimate an EGARCH(1, 1) model described by equation (2.2.6) using 5-year daily prices of the stock GE employing data in the period 2003-2007. The estimated parameters are: $\hat{\omega} = -0.1873$, $\hat{\alpha} = -0.0264$, $\hat{\beta} = 0.1191$ and $\hat{\gamma} = 0.9899$ with $\hat{r} = 1.3909$. The deterministic empirical volatility function is then smoothed by applying a spline function to the estimated conditional variance function over 60 months. This model is called DV Model 2. Figure 2.2 shows the two deterministic volatility models in our MC experiments.

2.5.2 Stochastic Volatility Model

For the stochastic volatility model we consider the set-up due to Heston (1993). Thus, we assume the following generation process for the logarithmic price

$$d\tilde{p}(t) = \left(\mu(t) - \frac{\sigma^2(t)}{2}\right) dt + \sigma(t)\, dW_1(t), \tag{2.5.4}$$

and

$$d\sigma^2(t) = \kappa(\alpha - \sigma^2(t))\, dt + \gamma\sigma(t)\, dW_2(t), \tag{2.5.5}$$

where $W_1(t)$ and $W_2(t)$ are standard Brownian processes with a correlation coefficient of ρ. Two different sets of parameters are adopted for the Heston model.

First, we use the model defined in Aït-Sahalia and Mancini (2008), which was also adopted by Tse and Yang (2012). Specifically, we set $\kappa = 5$, $\alpha = 0.04$, $\gamma = 0.5$ and $\rho = -0.5$. Furthermore, we apply the same drift terms $\mu(t)$ defined in Figure

1 to the stochastic volatility models. We call this parametric set-up SV Model 2. Second, we vary SV Model 2 by setting the reversion-rate parameter κ to 4 and the volatility-rate parameter γ to 0.4. We call this parametric set-up SV Model 1; we also set the reversion-rate parameter κ to 6 and the volatility-rate parameter γ to 0.6 and this set-up is called SV model 3. As before, price data are generated second by second, and then sampled to get transaction data, 5-minute and one-day intervals depending on the estimation method required.

2.5.3 Overnight Price Jump

While BSM may approximate price movements when the market is open, the process is disrupted when the market is closed. To this effect, it is important to examine how overnight price jumps affect the performance of the monthly volatility estimates. While the estimation of intraday volatility can be studied without taking account of overnight price jumps, this issue cannot be overlooked when the objective is to estimate monthly volatility. Table 2.2 summarizes some statistics for the distribution of the overnight returns of the ten NYSE stocks in our sample. The results show that the absolute values of the minimum and maximum of many of the stocks are larger than 10%.

We consider two models for fitting the overnight return: the generalized normal-distribution model and the t-distribution model. We also consider the empirical price jumps randomly drawn from the true jumps computed from the 10 NYSE stocks. Thus, if y denotes the overnight return, the generalized normal-distribution model states that $y \sim GN(\mu, \alpha, \beta)$, with

$$f(y) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|y-u|/\alpha)}. \tag{2.5.6}$$

On the other hand, the t-distribution model states that the density function of y is

$$f(y) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+\left(\frac{y-\mu}{\sigma}\right)^2}{\nu}\right]^{-\frac{\nu+1}{2}}, \tag{2.5.7}$$

where μ and σ are the location and scale parameters, respectively, and ν is the degrees of freedom.

We estimate the generalized normal model and t model for overnight returns of the sample of ten stocks using the maximum likelihood method. The results are summarized in Table 2.3. Figure 2.3 presents the QQ plots of five stocks in the sample. It can be seen that the generalized normal model and t model both appear to fit the overnight returns well. In our MC experiments, however, we consider both models. Based on the results in Table 2-3, we set $\mu = 0$, $\alpha = 0.0026$, $\beta = 0.69$ for the generalized normal distribution model in our MC study. For the parameters of the t-distribution model, we set $\sigma = 0.004$, $\mu = 0$ and $\nu = 2.5$. We also consider less fatter tails of the overnight returns by setting $\mu = 0$, $\alpha = 0.0032$, $\beta = 0.74$ for the generalized normal distribution model and $\sigma = 0.0046$, $\mu = 0$ and $\nu = 2.75$, which is called robust check 1. Further, for the fatter tailed model, we set $\mu = 0$, $\alpha = 0.0021$, $\beta = 0.64$ for the generalized normal distribution model and $\sigma = 0.003$, $\mu = 0$ and $\nu = 2.25$, which is called robust check 2.

In sum, we consider the following four processes of generating stock prices: BSM without empirical price jumps, BSM with overnight returns following generalized normal distribution, BSM with overnight returns following the t distribution and BSM with randomly drawn empirical price jumps.

2.5.4 Monte Carlo Results

Tables 2.4 through 2.9 summarize the MC results comparing the performance of different monthly volatility estimates based on their mean error (ME) and root mean-squared error (RMSE). All results are estimated using 1,000 MC replications. As the results for the two drift-term models are qualitatively similar, for the robustness MC check in Table 2.8 and Table 2.9, we present the results for Drift Model 1 only.

Tables 2.4 shows the results for the case when there are no overnight price jumps. These results throw light on markets with continuous trading for which the price process can be approximated by pure BSM. In this case, V_{R1} and V_{R2} perform better than the ACD-ICV method, and the ACD-ICV methods perform better than V_{RM} and V_G^D. Since there are no overnight price jumps, V_{R1} is numerically the same as V_{R2}, so we only report V_{R1} in this case. It should be noted, however, that the V_A estimates use tick data, while the V_R and V_G^D estimates use daily data.Surprisingly, V_G^D outperforms V_{RM} in all reported cases, although the latter is far more widely used in the literature.

The results for the cases when there are overnight price jumps are summarized in Tables 2.5 through 2.9. The V_A estimates perform the best for the deterministic volatility and stochastic volatility models, providing lower RMSE versus V_{R1} and V_{R2}, which use tick data. Of the two RV estimates using 5-minute data, V_{R1} performs slightly better than V_{R2}, although the difference is not large. Rather surprisingly, for the deterministic volatility models, V_G^D based on daily data performs well against the V_{Rj} estimates based on 5-minute data. The results are, however, slightly different for the stochastic volatility models, for which V_{Rj} clearly outperforms better than V_G^D. As in the case with no overnight price jumps, if only daily data are available for estimation, GARCH estimates outperform RV estimates.

Table 2.5 and 2.6 present the results of overnight price jumps with parameters estimated empirically for generalized normal distribution and t distribution respectively. Table 2.7 shows the results of overnight jumps randomly drawn from the empirical price jumps. The optimal price range is different for different volatility models and different distribution models. For the generalized normal distribution, lowest RMSE is achieved at $\delta = 0.35\%$, 0.25%, 0.3%, 0.25% and 0.2% for MV1, MV2, SV1, SV2 and SV3 respectively; for t distribution, lowest RMSE is achieved at $\delta = 0.15\%$ for all volatility models. However, the ACD-ICV measure which achieved lowest RMSE is slightly lower than the other ACD-ICV measures for all the volatility models, and the difference is quite small. Not surprisingly, the case for overnight prices jumps randomly drawn from empirical jumps is similar to the case of generalized normal distribution.

Table 2.8 and 2.9 present the results of Robustness MC check for the generalized normal distribution and *t* distribution. Panel A in Table 2.8 and 2.9 show less fatter tails for the overnight returns, however, panel B present much fatter tails. The results are similar to Table 2.5 and 2.6.

Figure 2.4 through 2.9 present examples of monthly volatility plots over a sample of 60 months. Note that all estimates trace the true volatility quite closely, although the RV estimates appear to be more volatile, especially for the stochastic volatility model.

In sum, the ACD-ICV method compares very well against the RV method for estimating monthly volatility when existence with overnight price jumps. The best results for the ACD-ICV method are obtained for δ in the range of 0.15% to 0.35%. When high-frequency data are available, the ACD-ICV method gives lower RMSE than the RV method. The GARCH method based on daily data, which is less popular in the literature for estimating monthly volatility, performs better than the widely used RV method.

2.6 Empirical Results for NYSE Data

We estimate the monthly volatility for the ten NYSE stocks in our sample. Figure 2.10 plots the monthly volatility estimates of the ten stocks for the 60 months over the period 2003 through 2007. To avoid jamming the figures, only the estimates V_{A3} (for $\delta = 0.25\%$), V_{R1} and V_G^D are presented. It can be seen that all estimates track each other quite closely, and there does not appear to be any systematic bias among the different methods. The RV method V_{R1}, however, exhibits a few extreme values of high volatility estimates and generally have the largest fluctuations among the three methods. It is interesting to observe that the volatility paths of the different stocks show significant co-movements.

Table 2.10 summarizes the average correlations across selected volatility estimation methods over the ten stocks in the sample period. It is noted that the pairwise correlations of V_{A1}, V_{A3}, V_{A5}, V_{R1}, V_{RM} and V_G^D are all above 0.6. The pairwise correlation coefficients of the volatility estimates V_{A3}, V_{R1} and V_G^D of the ten stocks are summarized in Table 2.11. It can be seen that the correlations are highest for V_{A3}, followed by V_G^D and then V_{R1}. Specifically, of the 45 pairs of volatility correlations 93.3% are maximized when V_{A3} is used as the volatility estimate, the remaining 6.7% are maximized when V_G^D is used as the volatility estimate, while none for V_{R1}. Many studies in the literature examine the effects of macroeconomic variables on stock volatility, and generally points to the co-movements of volatility across stocks. Thus, the ACD-ICV estimates support a higher volatility co-movement versus estimates based on the RV method. It will be interesting to further investigate volatility co-movements using the ACD-ICV estimates, in particular in relation to macroeconomic variables such as inflation, exchange rate, GDP growth and interest rate movements.

2.7 Volatility of S&P500

Implied volatility computed from option prices has often been used as a predictor for future historical volatility. The S&P500 Index volatility has been a case of particular research interest in the literature due to the popular reference to the CBOE volatility index VIX. Whaley (2009) provided a description of VIX and discussed some of its properties. In this section we examine the use of VIX as a predictor for future historical volatility when RV, GARCH and ACD-ICV estimates are used as proxies for historical volatility.

VIX is calculated and disseminated in real time by CBOE. It is a forward-looking index of the expected return volatility of the S&P500 Index over the next 30 calendar days and is implied from the prices of S&P500 Index options. The VIX index is quoted in percentage points as the annualized standard deviation of the return of the S&P500 Index over the next 30 days. It is based upon a model-free formula using a wide range of selected near- and near-term put and call options. Studies in the literature on the forecasting performance of implied volatility often use RV as the proxy for historical volatility. Jiang and Tian (2005) and Becker, Clements and White (2007) used RV computed over 30-day intervals as proxy for 30-day historical volatility in their studies on the information content of VIX on the volatility of the S&P500. Recently, Chung, Tsai, Wang and Weng (2011) considered both VIX and VIX options as predictors for the RV of the S&P500, although they did not specify the RV method used. We shall investigate the forecasting performance of VIX for the volatility of S&P500 when historical volatility is estimated by GARCH and ACD-ICV, and compare the results against using RV.

We downloaded daily closing values of VIX from the website of the CBOE. S&P500 tick data were obtained from The Institute for Financial Markets (IFM) Data Center. The sample period is from 1998 through 2007, with 2516 daily observations. 5-minute S&P500 data were extracted from 8:30 to 15:00 (Chicago time) each day.

We select N time points in the sample period that are at least 30 calendar days apart and denote them by t_i, for $i = 1, \cdots, N$. Altogether there are 117 nonoverlapping 30-day intervals in our sample ($N = 117$). Let VIX_i be the closing value of VIX at date t_i. We denote Y_i generically as an estimate of the historical volatility over the 30-day period starting from time t_i. Let Y_i be V_{R1}, V_G^D or V_{Aj}, for $j = 1, \cdots, 5$, so that historical estimates using the RV, GARCH and ACD-ICV methods are considered. We also denote R_i as the 30-day return of the S&P500 starting from time t_i. Table 2.12 summarizes the correlations between VIX and different estimates of 30-day volatility over the forecast intervals of VIX. It can be seen that V_{R1} has the lowest correlation with VIX, whereas V_{A1} and V_{A2} have the highest correlations.

To examine the relationship between return and volatility we consider the regression of return on VIX and contemporaneous volatility. Thus, we estimate the following regression equations

$$R_i = \alpha + \beta X_i + \xi_i, \tag{2.7.1}$$

and

$$R_i = \alpha + \beta X_i^2 + \xi_i, \tag{2.7.2}$$

for $i = 1, \cdots, N$, where X_i is VIX at time t_i or historical volatility estimate Y_i. In the former case, return is regressed on volatility forecast, while in the latter case we consider a contemporaneous return-volatility relationship. The results are summarized in Table 2.13. We can see that VIX, V_{A1}, V_{R1}, V_{RM} and V_G^D are statistically significant, while all other regressors are insignificant at the 5% level. Overall the return-volatility relationship is quite weak, which is in line with the results in Chung, Tsai, Wang and Weng (2011).

More importantly, we consider the regressions of historical volatility estimates on volatility forecasts using VIX. Thus, we estimate the following regression equations

$$Y_i = \alpha + \beta \, \mathrm{VIX}_i + \xi_i, \tag{2.7.3}$$

and

$$Y_i^2 = \alpha + \beta \, \mathrm{VIX}_i^2 + \xi_i, \tag{2.7.4}$$

for $i = 1, \cdots, N$. The results are summarized in Table 2.14. It can be seen that R^2 is the highest for the regressions with the ACD-ICV measures as the dependent variables. Rather remarkably, the regressions with V_{R1} as the dependent variable produce the lowest R^2. These results show that VIX is a more successful predictor of future volatility if volatility is estimated by the ACD-ICV method but not the RV method. Figure 11 plots VIX and some historical volatility estimates. VIX appears to be more volatile than the historical volatility it predicts. There are some periods for which VIX over-predicts volatility as estimated by V_{R1}. This over-prediction, however, is not evident if historical volatility is estimated by the ACD-ICV method. Overall, our results show that VIX has higher prediction value if its performance is measured against historical estimates using the ACD-ICV method.

2.8 Conclusion

In this chapter we have extended the ACD-ICV method proposed by Tse and Yang (2012) to estimate stock volatility over longer intervals such as a month. Estimation of low-frequency volatility is important for studies involving macroeconomic data that are available only monthly or quarterly. In addition, returns over longer intervals are less susceptible to the contamination of noise over short intervals and may be preferred in studies on asset pricing. Our MC study suggests that price events defined by return of about 0.15% to 0.35% are appropriate for the ACD-ICV method. Based on the transaction data, the ACD-ICV method outperforms the RV method in our MC experiments. On the other hand, if daily data are used, the GARCH method based on aggregating the daily estimates of conditional variance is superior to the RV method, which is widely used in the literature.

Our empirical results using ten NYSE stocks show that the ACD-ICV, RV and GARCH estimates track each other quite closely. The RV estimates, however, have larger fluctuations and exhibit occasionally extreme volatility estimates. Co-movements of volatility across different stocks are highest according to the ACD-ICV estimates. Our empirical study on VIX and the S&P500 index shows that VIX is a more successful predictor of future volatility if volatility is estimated by the ACD-ICV method than the RV method.

Overall we have shown that using the ACD-ICV method on high-frequency data (tick transaction data) provides superior estimates of low-frequency volatility (over monthly intervals) to the RV method. If only daily data are available, however, monthly volatility computed by aggregating the daily conditional variance estimates of the GARCH model provides a better estimate than the RV method. As better volatility estimates may help improve the pricing of derivatives and enhance the robustness of inference in asset pricing, the ACD-ICV method should provide a useful tool in empirical research.

Table 2.1: Stocks

Stock	Code
Bank of America Corp	BAC
General Electric	GE
Merck & Co Inc	MRK
Johnson & Johnson	JNJ
JP Morgan	JPM
Wal Mart	WMT
IBM	IBM
Pfizer	PFE
AT&T Inc.	T
Chevron Corporation	CVX

Table 2.2: Summary statistics of overnight returns

	Quantile						
Code	Min	5%	25%	50%	75%	95%	Max
BAC	-0.2559	-0.0124	-0.0037	0.0000	0.0032	0.0119	0.1061
GE	-0.1030	-0.0108	-0.0032	0.0000	0.0033	0.0112	0.0777
MRK	-0.2997	-0.0127	-0.0035	0.0000	0.0036	0.0112	0.0633
JNJ	-0.1809	-0.0111	-0.0030	0.0000	0.0034	0.0107	0.0489
JPM	-0.1245	-0.0153	-0.0037	0.0000	0.0044	0.0157	0.0808
WMT	-0.0931	-0.0130	-0.0039	0.0000	0.0047	0.0136	0.0849
IBM	-0.1553	-0.0118	-0.0030	0.0000	0.0034	0.0126	0.1377
PFE	-0.1638	-0.0134	-0.0033	0.0000	0.0044	0.0136	0.1078
T	-0.1313	-0.0121	-0.0031	0.0000	0.0035	0.0111	0.0603
CVX	-0.0518	-0.0097	-0.0025	0.0000	0.0034	0.0094	0.0382

Table 2.3: Estimation of overnight-return models

	Generalized Normal			Local t		
Code	μ	α	β	μ	σ	ν
BAC	-0.0003	0.0025	0.6757	-0.0001	0.0043	2.1674
GE	0.0001	0.0027	0.7258	0.0000	0.0042	2.4911
MRK	-0.0002	0.0029	0.7224	0.0001	0.0045	2.4651
JNJ	0.0001	0.0022	0.6731	0.0002	0.0043	2.6833
JPM	0.0003	0.0023	0.6107	0.0004	0.0051	2.0563
WMT	0.0004	0.0039	0.7847	0.0004	0.0051	2.5951
IBM	0.0002	0.0023	0.6572	0.0002	0.0039	1.9239
PFE	0.0004	0.0034	0.7432	0.0005	0.0048	2.2945
T	0.0000	0.0004	0.4173	0.0002	0.0042	2.3526
CVX	0.0004	0.0034	0.8735	0.0005	0.0039	2.9249

Table 2.4: Monte Carlo results without overnight jumps.

	Volatility Model									
	MV1		MV2		SV1		SV2		SV3	
method	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
Panel A: Drift Model 1										
V_{A1}	0.538	0.654	0.578	0.681	0.894	1.115	0.839	1.061	0.816	1.041
V_{A2}	0.422	0.604	0.458	0.614	0.724	1.014	0.686	0.984	0.685	0.975
V_{A3}	0.350	0.603	0.377	0.596	0.616	0.994	0.599	0.977	0.624	0.990
V_{A4}	0.302	0.630	0.322	0.607	0.544	1.012	0.558	1.021	0.600	1.054
V_{A5}	0.272	0.669	0.298	0.639	0.507	1.060	0.540	1.089	0.587	1.131
V_{R1}	0.043	0.272	0.034	0.325	0.025	0.519	0.028	0.486	0.031	0.466
V_{RM}	-0.514	2.395	-0.624	2.892	-1.037	4.675	-0.957	4.358	-0.906	4.165
V_G^D	0.133	1.745	0.191	1.916	0.446	3.769	0.535	3.797	0.654	3.825
Panel B: Drift Model 2										
V_{A1}	0.554	0.666	0.596	0.699	0.949	1.172	0.882	1.106	0.856	1.080
V_{A2}	0.435	0.612	0.472	0.627	0.771	1.059	0.725	1.022	0.715	1.003
V_{A3}	0.356	0.607	0.387	0.604	0.654	1.030	0.631	1.007	0.655	1.018
V_{A4}	0.311	0.631	0.331	0.613	0.576	1.042	0.587	1.045	0.624	1.072
V_{A5}	0.286	0.673	0.306	0.643	0.534	1.091	0.562	1.109	0.606	1.149
V_{R1}	0.044	0.272	0.035	0.325	0.026	0.520	0.029	0.486	0.032	0.466
V_{RM}	-0.514	2.395	-0.624	2.892	-1.037	4.675	-0.957	4.358	-0.906	4.165
V_G^D	0.133	1.747	0.203	1.915	0.446	3.768	0.536	3.797	0.656	3.825

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1000 MC replications of 5-year monthly volatility. All figures are in percentage. V_{A1}, V_{A2},V_{A3}, V_{A4}, V_{A5} is ACD-ICV volatility measures for δ =0.15%, 0.2%, 0.25%, 0.3%, 0.35% respectively. V_{R1} is realized volatility defined in Section 2.1. V_{RM} is realized volatility defined in equation (1). V_G^D is GARCH estimates based on daily data.

Table 2.5: Monte Carlo results with overnight jumps following Generalized Normal Distribution

	Volatility Model									
	MV1		MV2		SV1		SV2		SV3	
method	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
Panel A: Drift Model 1										
V_{A1}	0.249	1.890	0.502	1.321	0.827	1.659	0.606	1.850	0.424	2.061
V_{A2}	0.094	1.890	0.328	1.287	0.566	1.547	0.359	1.780	0.193	2.022
V_{A3}	0.006	1.892	0.227	1.280	0.406	1.513	0.218	1.762	0.077	2.046
V_{A4}	-0.048	1.873	0.160	1.288	0.301	1.501	0.142	1.783	0.030	2.072
V_{A5}	-0.077	1.849	0.111	1.288	0.237	1.526	0.106	1.820	0.024	2.085
V_{R1}	-0.182	2.860	-0.124	2.563	-0.066	2.026	-0.087	2.164	-0.104	2.273
V_{R2}	-0.184	2.870	-0.127	2.576	-0.068	2.046	-0.089	2.182	-0.106	2.289
V_{RM}	-0.885	4.042	-0.919	4.211	-1.224	5.403	-1.170	5.170	-1.138	5.039
V_G^D	0.436	2.328	0.390	2.301	0.330	3.808	0.349	3.864	0.385	3.927
Panel B: Drift Model 2										
V_{A1}	0.274	1.895	0.530	1.327	0.904	1.735	0.672	1.901	0.485	2.103
V_{A2}	0.111	1.890	0.347	1.287	0.632	1.603	0.411	1.823	0.240	2.056
V_{A3}	0.020	1.889	0.243	1.279	0.459	1.562	0.263	1.795	0.116	2.070
V_{A4}	-0.036	1.868	0.173	1.284	0.347	1.541	0.179	1.815	0.065	2.093
V_{A5}	-0.069	1.847	0.121	1.285	0.275	1.558	0.139	1.845	0.050	2.104
V_{R1}	-0.181	2.860	-0.123	2.563	-0.064	2.026	-0.086	2.164	-0.103	2.272
V_{R2}	-0.183	2.870	-0.126	2.575	-0.066	2.046	-0.088	2.182	-0.105	2.289
V_{RM}	-0.884	4.042	-0.919	4.211	-1.224	5.403	-1.169	5.170	-1.138	5.039
V_G^D	0.439	2.327	0.413	2.329	0.327	3.804	0.350	3.863	0.385	3.926

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1000 MC replications of 5-year monthly volatility. All figures are in percentage. V_{A1}, V_{A2},V_{A3}, V_{A4}, V_{A5} is ACD-ICV volatility measures for δ =0.15%,0.2%, 0.25%, 0.3%, 0.35% respectively. V_{R1} and V_{R2} are realized volatility defined in Section 2.1. V_{RM} is realized volatility defined in equation (1). V_G^D is GARCH estimates based on daily data. Overnight returns are generalized normally distributed with $\mu = 0, \alpha = 0.0026, \beta = 0.69$.

Table 2.6: Monte Carlo results with overnight jumps following *t* Distribution

	Volatility Model									
	MV1		MV2		SV1		SV2		SV3	
method	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
Panel A: Drift Model 1										
V_{A1}	-1.165	2.106	-0.767	1.402	-0.065	1.496	-0.341	1.779	-0.562	2.061
V_{A2}	-1.320	2.202	-0.941	1.510	-0.321	1.511	-0.587	1.826	-0.796	2.121
V_{A3}	-1.411	2.256	-1.041	1.581	-0.484	1.558	-0.732	1.872	-0.913	2.179
V_{A4}	-1.463	2.272	-1.108	1.636	-0.589	1.595	-0.810	1.921	-0.959	2.216
V_{A5}	-1.493	2.274	-1.154	1.670	-0.659	1.648	-0.846	1.962	-0.974	2.230
V_{R1}	-1.705	3.537	-1.456	3.145	-1.038	2.462	-1.143	2.647	-1.225	2.791
V_{R2}	-1.707	3.543	-1.458	3.152	-1.037	2.473	-1.142	2.656	-1.225	2.799
V_{RM}	-2.369	4.685	-2.215	4.717	-2.167	5.716	-2.197	5.534	-2.230	5.443
V_G^D	-1.092	2.460	-0.905	2.405	-0.595	3.759	-0.654	3.801	-0.680	3.847
Panel B: Drift Model 2										
V_{A1}	-1.143	2.094	-0.741	1.380	0.011	1.552	-0.276	1.818	-0.502	2.082
V_{A2}	-1.304	2.192	-0.923	1.494	-0.257	1.548	-0.534	1.848	-0.746	2.139
V_{A3}	-1.398	2.246	-1.027	1.570	-0.432	1.591	-0.687	1.891	-0.874	2.188
V_{A4}	-1.454	2.269	-1.097	1.624	-0.545	1.623	-0.773	1.935	-0.927	2.224
V_{A5}	-1.484	2.269	-1.146	1.661	-0.617	1.669	-0.813	1.976	-0.945	2.238
V_{R1}	-1.704	3.537	-1.455	3.144	-1.037	2.462	-1.142	2.646	-1.224	2.790
V_{R2}	-1.706	3.542	-1.457	3.152	-1.036	2.472	-1.141	2.655	-1.223	2.798
V_{RM}	-2.369	4.685	-2.215	4.717	-2.167	5.716	-2.196	5.534	-2.231	5.443
V_G^D	-1.116	2.469	-0.922	2.398	-0.595	3.758	-0.657	3.799	-0.680	3.845

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1000 MC replications of 5-year monthly volatility. All figures are in percentage. V_{A1}, V_{A2},V_{A3}, V_{A4}, V_{A5} is ACD-ICV volatility measures for δ =0.15%,0.2%, 0.25%, 0.3%, 0.35% respectively. V_{R1} and V_{R2} are realized volatility defined in Section 2.1. V_{RM} is realized volatility defined in equation (1). V_G^D is GARCH estimates based on daily data. Overnight returns are *t* distributed with $\mu = 0, \sigma = 0.004, \nu = 2.5$.

Table 2.7: Monte Carlo results with overnight jumps randomly drawn from the empirical jumps

	Volatility Model									
	MV1		MV2		SV1		SV2		SV3	
method	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
Panel A: Drift model 1										
V_{A1}	0.240	1.939	0.487	1.377	0.729	1.608	0.524	1.828	0.356	2.060
V_{A2}	0.093	1.935	0.321	1.338	0.498	1.522	0.303	1.783	0.145	2.036
V_{A3}	0.012	1.930	0.226	1.330	0.349	1.506	0.176	1.773	0.039	2.054
V_{A4}	-0.040	1.905	0.161	1.334	0.256	1.511	0.103	1.798	0.004	2.091
V_{A5}	-0.065	1.880	0.120	1.333	0.195	1.534	0.081	1.838	0.001	2.097
V_{R1}	-0.286	3.459	-0.200	3.142	-0.099	2.509	-0.127	2.667	-0.150	2.788
V_{R2}	-0.287	3.465	-0.202	3.151	-0.103	2.527	-0.131	2.683	-0.154	2.802
V_{RM}	-0.986	4.525	-0.994	4.633	-1.256	5.663	-1.209	5.464	-1.184	5.358
V_G^D	0.369	2.359	0.386	2.341	0.342	3.862	0.357	3.923	0.387	3.993
Panel B: Drift model 2										
V_{A1}	0.263	1.944	0.513	1.375	0.791	1.665	0.577	1.870	0.403	2.092
V_{A2}	0.110	1.937	0.340	1.341	0.546	1.568	0.345	1.814	0.181	2.064
V_{A3}	0.022	1.932	0.238	1.331	0.393	1.546	0.208	1.801	0.070	2.076
V_{A4}	-0.030	1.907	0.173	1.336	0.292	1.540	0.134	1.819	0.029	2.103
V_{A5}	-0.055	1.878	0.129	1.332	0.229	1.558	0.106	1.855	0.021	2.112
V_{R1}	-0.285	3.459	-0.200	3.142	-0.098	2.509	-0.126	2.667	-0.149	2.788
V_{R2}	-0.286	3.465	-0.201	3.151	-0.102	2.527	-0.130	2.682	-0.153	2.802
V_{RM}	-0.986	4.525	-0.994	4.632	-1.256	5.663	-1.209	5.465	-1.185	5.358
V_G^D	0.327	2.332	0.407	2.369	0.341	3.861	0.357	3.922	0.387	3.992

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1000 MC replications of 5-year monthly volatility. All figures are in percentage. V_{A1}, V_{A2},V_{A3}, V_{A4}, V_{A5} is ACD-ICV volatility measures for δ =0.15%,0.2%, 0.25%, 0.3%, 0.35% respectively. V_{R1} and V_{R2} are realized volatility defined in Section 2.1. V_{RM} is realized volatility defined in equation (1). V_G^D is GARCH estimates based on daily data. Overnight returns are randomly drawn from the empirical jumps.

Table 2.8: Robustness Monte Carlo results with overnight jumps following GN Distribution

	Volatility Model									
	MV1		MV2		SV1		SV2		SV3	
method	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
Panel A: Robustness check 1										
V_{A1}	0.356	1.926	0.599	1.369	0.890	1.691	0.674	1.879	0.497	2.085
V_{A2}	0.202	1.913	0.425	1.324	0.632	1.572	0.428	1.801	0.266	2.043
V_{A3}	0.115	1.909	0.324	1.307	0.472	1.532	0.287	1.778	0.149	2.058
V_{A4}	0.061	1.882	0.258	1.309	0.366	1.518	0.212	1.798	0.104	2.086
V_{A5}	0.035	1.851	0.211	1.301	0.302	1.536	0.180	1.832	0.098	2.091
V_{R1}	-0.051	2.745	-0.012	2.463	0.019	1.948	0.006	2.080	-0.005	2.183
V_{R2}	-0.053	2.754	-0.014	2.476	0.017	1.968	0.004	2.098	-0.007	2.199
V_{RM}	-0.755	3.949	-0.808	4.138	-1.140	5.354	-1.077	5.113	-1.039	4.975
V_G^D	0.529	2.313	0.482	2.291	0.411	3.816	0.436	3.874	0.474	3.940
Panel B: Robustness check 2										
V_{A1}	2.552	2.992	2.484	2.680	0.931	1.728	0.717	1.910	0.541	2.120
V_{A2}	2.467	2.924	2.383	2.594	0.675	1.604	0.470	1.835	0.311	2.074
V_{A3}	2.415	2.870	2.316	2.543	0.514	1.565	0.330	1.810	0.192	2.090
V_{A4}	2.381	2.844	2.270	2.507	0.407	1.554	0.253	1.832	0.150	2.118
V_{A5}	2.375	2.829	2.251	2.494	0.343	1.565	0.219	1.868	0.138	2.129
V_{R1}	2.250	3.866	2.006	3.466	0.038	2.232	0.023	2.381	0.010	2.496
V_{R2}	2.245	3.879	2.003	3.473	0.035	2.251	0.021	2.397	0.008	2.511
V_{RM}	1.543	4.439	1.204	4.455	-1.111	5.457	-1.051	5.237	-1.016	5.116
V_G^D	2.897	3.790	2.567	3.520	0.453	3.845	0.480	3.905	0.524	3.974

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1000 MC replications of 5-year monthly volatility. All figures are in percentage. V_{A1}, V_{A2},V_{A3}, V_{A4}, V_{A5} is ACD-ICV volatility measures for δ =0.15%,0.2%, 0.25%, 0.3%, 0.35% respectively. V_{R1} and V_{R2} are realized volatility defined in Section 2.1. V_{RM} is realized volatility defined in equation (1). V_G^D is GARCH estimates based on daily data. Overnight returns in panel A are generalized normally distributed with $\mu = 0, \alpha = 0.0032, \beta = 0.74$; overnight returns in panel B are generalized normally distributed with $\mu = 0, \alpha = 0.0021, \beta = 0.64$.

Table 2.9: Robustness Monte Carlo results with overnight jumps following *t* Distribution

	Volatility Model									
	MV1		MV2		SV1		SV2		SV3	
method	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
Panel A: Robustness check 1										
V_{A1}	-0.599	1.918	-0.264	1.245	0.291	1.488	0.038	1.750	-0.165	2.010
V_{A2}	-0.752	1.978	-0.435	1.301	0.038	1.449	-0.205	1.747	-0.393	2.031
V_{A3}	-0.839	2.008	-0.535	1.347	-0.125	1.470	-0.346	1.765	-0.510	2.067
V_{A4}	-0.891	2.007	-0.600	1.386	-0.230	1.488	-0.420	1.799	-0.556	2.099
V_{A5}	-0.921	1.996	-0.646	1.404	-0.295	1.523	-0.456	1.839	-0.567	2.108
V_{R1}	-1.095	3.225	-0.923	2.887	-0.639	2.294	-0.709	2.455	-0.764	2.581
V_{R2}	-1.096	3.233	-0.925	2.897	-0.639	2.309	-0.709	2.469	-0.765	2.594
V_{RM}	-1.766	4.399	-1.687	4.497	-1.777	5.597	-1.772	5.393	-1.780	5.284
V_G^D	-0.500	2.289	-0.398	2.262	-0.212	3.750	-0.237	3.796	-0.240	3.849
Panel B: Robustness check 2										
V_{A1}	-2.444	2.876	-1.929	2.182	-0.715	1.671	-1.031	1.970	-1.277	2.255
V_{A2}	-2.534	2.955	-2.032	2.281	-0.974	1.759	-1.281	2.080	-1.513	2.371
V_{A3}	-2.590	2.994	-2.102	2.354	-1.139	1.844	-1.426	2.162	-1.635	2.458
V_{A4}	-2.627	3.033	-2.149	2.401	-1.248	1.907	-1.509	2.223	-1.688	2.504
V_{A5}	-2.639	3.038	-2.171	2.425	-1.318	1.967	-1.548	2.271	-1.706	2.526
V_{R1}	-2.866	4.040	-2.462	3.531	-1.755	2.712	-1.923	2.941	-2.056	3.123
V_{R2}	-2.869	4.053	-2.463	3.538	-1.755	2.722	-1.923	2.950	-2.056	3.130
V_{RM}	-3.506	5.174	-3.199	5.075	-2.859	5.885	-2.951	5.744	-3.035	5.690
V_G^D	-2.343	3.101	-1.950	2.885	-1.306	3.834	-1.418	3.900	-1.483	3.956

Notes: ME = mean error, RMSE = root mean-squared error. The results are based on 1000 MC replications of 5-year monthly volatility. All figures are in percentage. V_{A1}, V_{A2},V_{A3}, V_{A4}, V_{A5} is ACD-ICV volatility measures for δ =0.15%,0.2%, 0.25%, 0.3%, 0.35% respectively. V_{R1} and V_{R2} are realized volatility defined in Section 2.1. V_{RM} is realized volatility defined in equation (1). V_G^D is GARCH estimates based on daily data. Overnight returns in panel B are *t* distributed with $\mu = 0, \sigma = 0.0046, \nu = 2.75$; overnight returns in panel B are *t* distributed with $\mu = 0, \sigma = 0.0030, \nu = 2.25$.

Table 2.10: Correlations of different volatility estimates

	V_{A1}	V_{A3}	V_{A5}	V_{R1}	V_{RM}	V_G^D
V_{A1}	1.000	0.972	0.958	0.745	0.628	0.697
V_{A3}		1.000	0.978	0.763	0.653	0.727
V_{A5}			1.000	0.786	0.680	0.755
V_{R2}				1.000	0.892	0.700
V_{RM}					1.000	0.701
V_G^D						1.000

Table 2.11: Correlations of volatility estimates of different stocks

	Volatility estimate: V_{A3}									
	BAC	GE	MRK	JNJ	JPM	WMT	IBM	PFE	T	CVX
BAC	1.000	**0.714**	**0.734**	**0.441**	**0.898**	**0.819**	**0.867**	**0.550**	**0.642**	**0.658**
GE		1.000	**0.603**	**0.843**	**0.903**	0.713	**0.828**	**0.772**	**0.806**	**0.316**
MRK			1.000	**0.517**	**0.720**	**0.648**	**0.704**	**0.691**	**0.632**	**0.744**
JNJ				1.000	**0.687**	0.455	**0.649**	**0.802**	**0.734**	**0.218**
JPM					1.000	**0.843**	**0.925**	**0.690**	**0.759**	**0.541**
WMT						1.000	**0.870**	**0.512**	0.547	**0.599**
IBM							1.000	**0.678**	**0.701**	**0.616**
PFE								1.000	**0.704**	**0.375**
T									1.000	**0.385**
CVX										1.000

	Volatility estimate: V_{R1}									
	BAC	GE	MRK	JNJ	JPM	WMT	IBM	PFE	T	CVX
BAC	1.000	0.600	0.106	0.253	0.848	0.672	0.667	0.036	0.251	0.564
GE		1.000	0.042	0.776	0.859	0.685	0.716	0.109	0.558	0.207
MRK			1.000	0.062	0.058	0.005	-0.043	0.171	0.065	0.023
JNJ				1.000	0.577	0.384	0.530	0.187	0.535	-0.007
JPM					1.000	0.794	0.808	0.057	0.471	0.480
WMT						1.000	0.669	-0.048	0.327	0.514
IBM							1.000	0.031	0.435	0.451
PFE								1.000	0.226	0.037
T									1.000	-0.052
CVX										1.000

	Volatility estimate: V_G^D									
	BAC	GE	MRK	JNJ	JPM	WMT	IBM	PFE	T	CVX
BAC	1.000	0.704	0.057	0.247	0.817	0.559	0.655	0.147	0.418	0.120
GE		1.000	0.175	0.686	0.859	**0.812**	0.697	0.349	0.708	-0.049
MRK			1.000	0.332	0.125	0.005	-0.008	0.371	0.163	-0.082
JNJ				1.000	0.517	**0.536**	0.445	0.556	0.641	-0.158
JPM					1.000	0.788	0.781	0.268	0.679	0.174
WMT						1.000	0.600	0.179	**0.604**	-0.069
IBM							1.000	0.182	0.554	0.299
PFE								1.000	0.231	0.055
T									1.000	-0.114
CVX										1.000

Table 2.12: Correlations of VIX and S&P500 30-day volatility estimates

	VIX	V_{A1}	V_{A2}	V_{A3}	V_{A4}	V_{A5}	V_{R1}	V_{RM}	V_G^D
VIX	1.000	0.989	0.986	0.979	0.978	0.834	0.845	0.921	0.910
V_{A1}		1.000	0.989	0.983	0.985	0.853	0.864	0.921	0.901
V_{A2}			1.000	0.988	0.986	0.862	0.873	0.910	0.900
V_{A3}				1.000	0.981	0.862	0.873	0.900	0.897
V_{A4}					1.000	0.878	0.888	0.893	0.882
V_{A5}						1.000	0.999	0.754	0.681
V_{R1}							1.000	0.770	0.701
V_{RM}								1.000	0.911
V_G^D									1.000

Table 2.13: Regression results of return on VIX and volatility estimates

	VIX	V_{A1}	V_{A2}	V_{A3}	V_{A4}	V_{A5}	V_{R1}	V_{RM}	V_G^D
Panel A: $R_i = \alpha + \beta X_i + \xi_i$									
α	-0.0251	0.0208	0.0160	0.0156	0.0123	0.0123	0.0132	0.0303	0.0315
	(-1.9263)	(1.4835)	(1.2225)	(1.2160)	(0.9847)	(1.0258)	(1.6967)	(2.7931)	(2.4280)
β	0.1442	-0.0755	-0.0531	-0.0511	-0.0350	-0.0356	-0.0459	-0.1577	-0.1618
	(2.4668)	(-1.1714)	(-0.8768)	(-0.8632)	(-0.6065)	(-0.6373)	(-1.2470)	(-2.5215)	(-2.1512)
R^2	0.0503	0.0118	0.0066	0.0064	0.0032	0.0035	0.0133	0.0524	0.0387
Panel B: $R_i = \alpha + \beta X_i^2 + \xi_i$									
α	-0.0125	0.0134	0.0109	0.0100	0.0088	0.0086	0.0077	0.0181	0.0178
	(-1.7400)	(1.7116)	(1.4754)	(1.3974)	(1.2476)	(1.2727)	(1.5814)	(2.8498)	(2.4304)
β	0.3574	-0.1731	-0.1215	-0.1028	-0.0767	-0.0737	-0.0559	-0.4274	-0.4232
	(3.0684)	(-1.2657)	(-0.9618)	(-0.8530)	(-0.6537)	(-0.6655)	(-1.2025)	(-2.7581)	(-2.1370)
R^2	0.0757	0.0137	0.0080	0.0063	0.0037	0.0038	0.0124	0.0620	0.0382

Notes: R_i is the return of the ith 30-day interval. X_i is the forecast of the volatility of the ith interval by VIX or a historical estimate of the volatility of the interval, Y_i (e.g., V_{A2}). Numbers in parentheses are t-statistics.

Table 2.14: Regression results of 30-day volatility estimates on VIX

	V_{A1}	V_{A2}	V_{A3}	V_{A4}	V_{A5}	V_{R1}	V_{RM}	V_G^D
Panel A: $Y_i = \alpha + \beta\,\mathrm{VIX}_i + \xi_i$								
α	0.0407	0.0267	0.0223	0.0134	0.0066	-0.0580	0.0145	0.0299
	(4.0632)	(2.5170)	(2.0564)	(1.2418)	(0.5818)	(-2.3715)	(1.1098)	(3.1344)
β	0.7896	0.8449	0.8627	0.8988	0.9213	1.1078	0.6896	0.6330
	(17.5326)	(17.6902)	(17.6637)	(18.5012)	(18.0094)	(10.0801)	(11.7403)	(14.7747)
R^2	0.7254	0.7289	0.7283	0.7463	0.7360	0.4645	0.5412	0.6520
Panel B: $Y_i^2 = \alpha + \beta\,\mathrm{VIX}_i^2 + \xi_i$								
α	0.0125	0.0094	0.0075	0.0058	0.0037	-0.0184	0.0059	0.0070
	(4.1117)	(2.8253)	(2.1425)	(1.6487)	(0.9559)	(-1.4203)	(1.7788)	(3.1855)
β	0.7032	0.7584	0.7972	0.8259	0.8600	1.2749	0.4891	0.4602
	(14.2688)	(14.0031)	(14.0789)	(14.3752)	(13.6903)	(6.0709)	(9.0759)	(12.8213)
R^2	0.6359	0.6271	0.6296	0.6394	0.6164	0.2361	0.4123	0.5848

Notes: Y_i is a historical estimate of the volatility of the ith 30-day interval (e.g., V_{A2}). Numbers in parentheses are t-statistics.

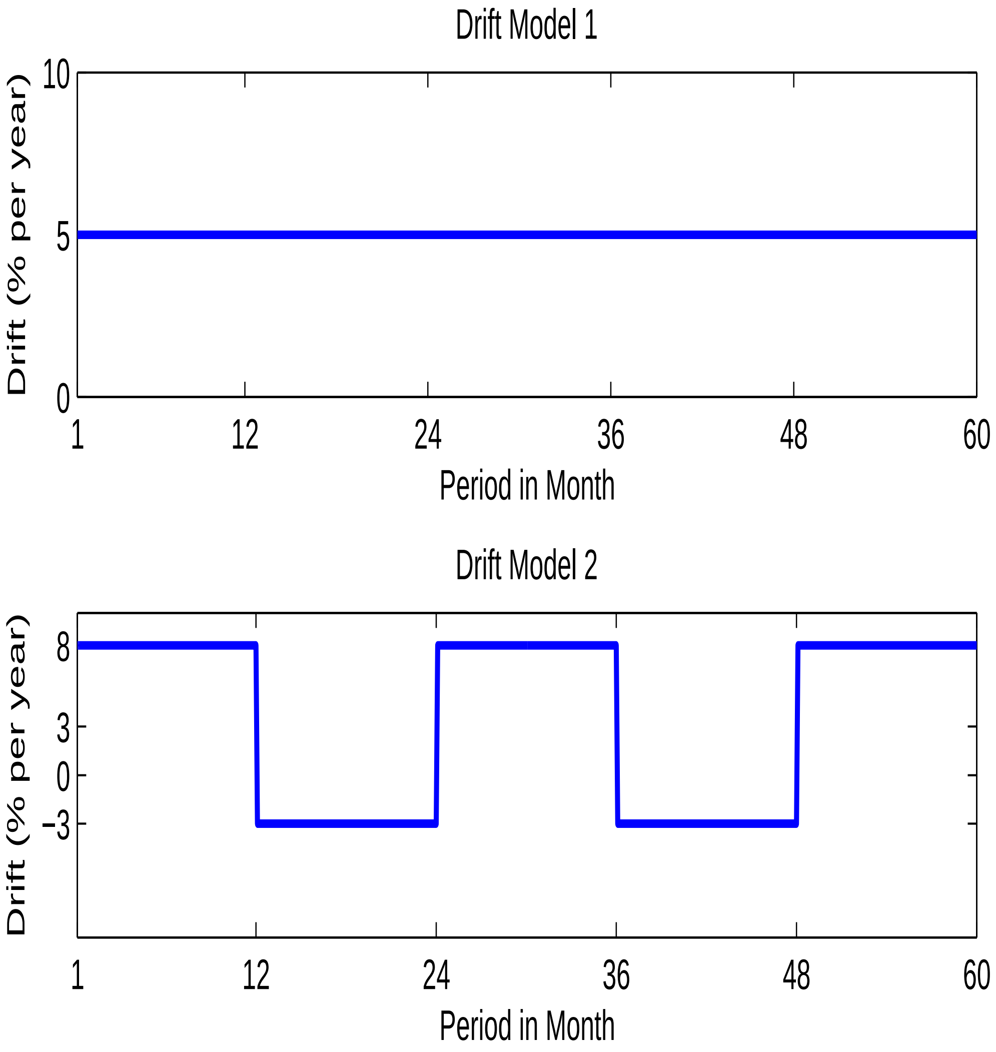


Figure 2.1: The drift term

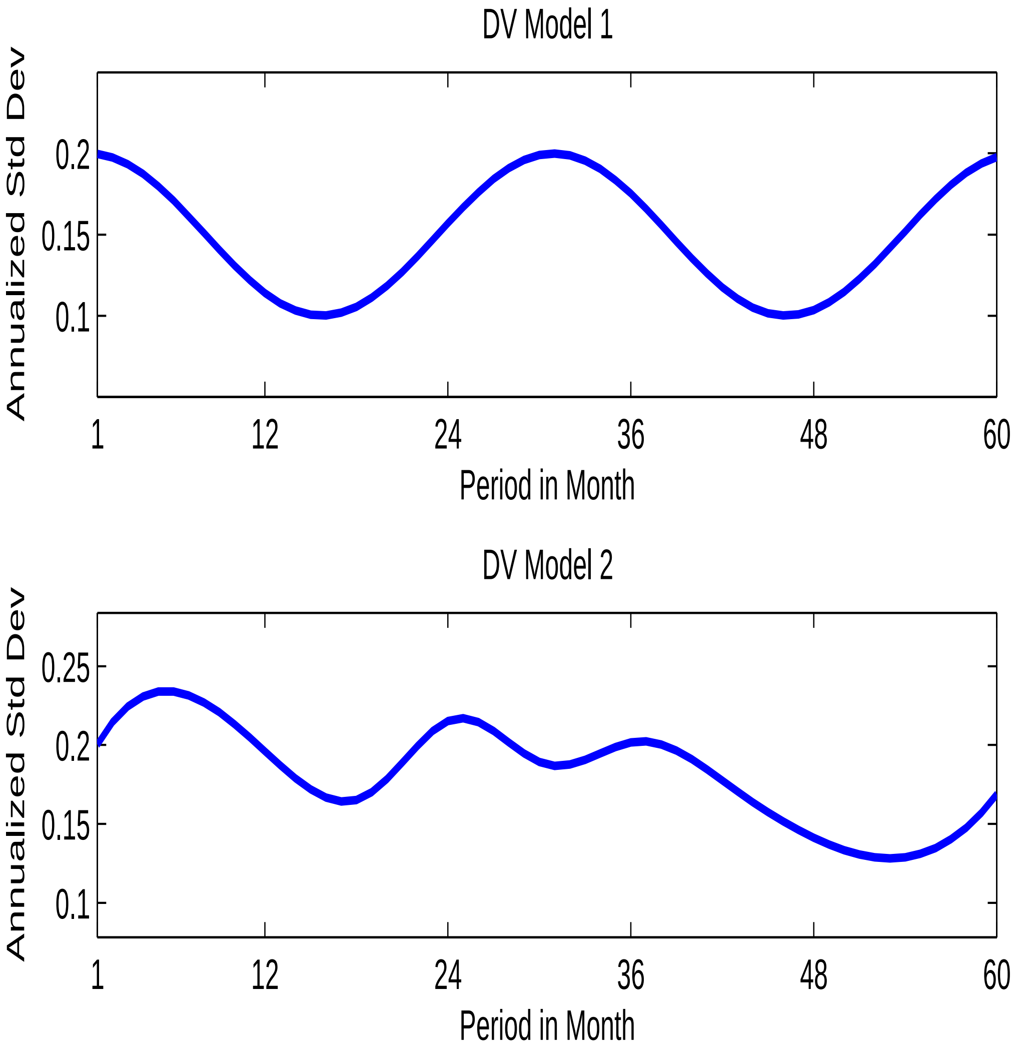


Figure 2.2: Deterministic volatility models

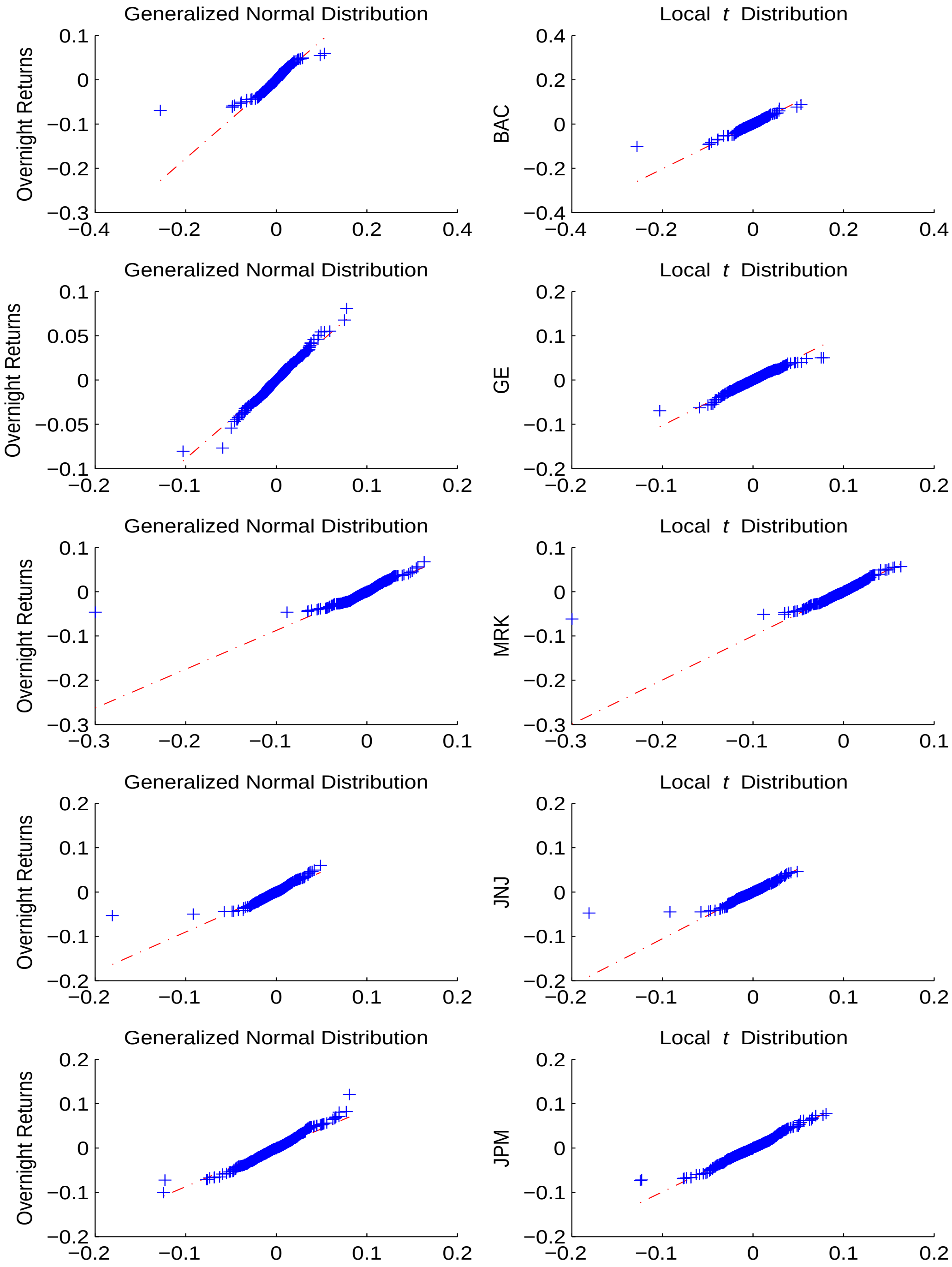


Figure 2.3: QQ plot of overnight returns

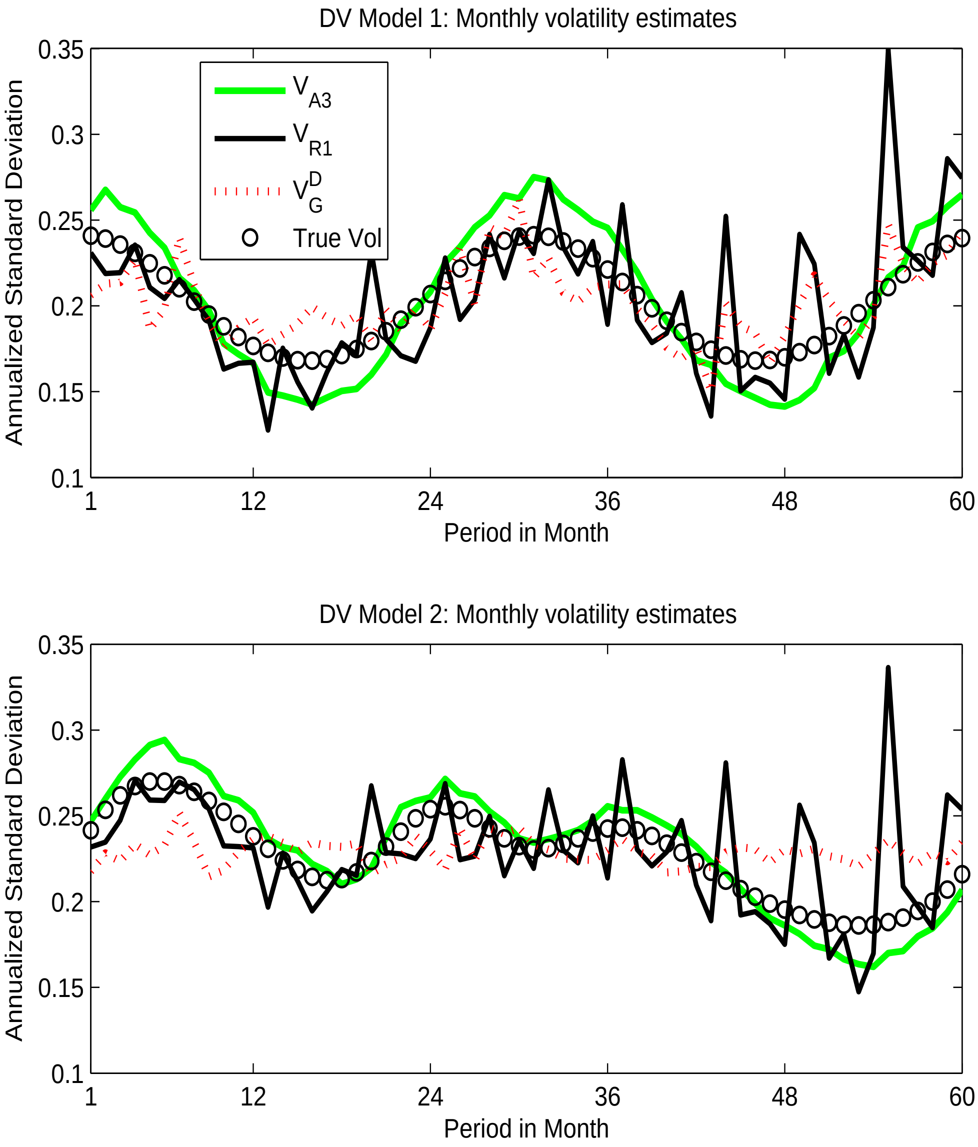


Figure 2.4: Estimation of deterministic volatility with generalized normal overnight price jumps

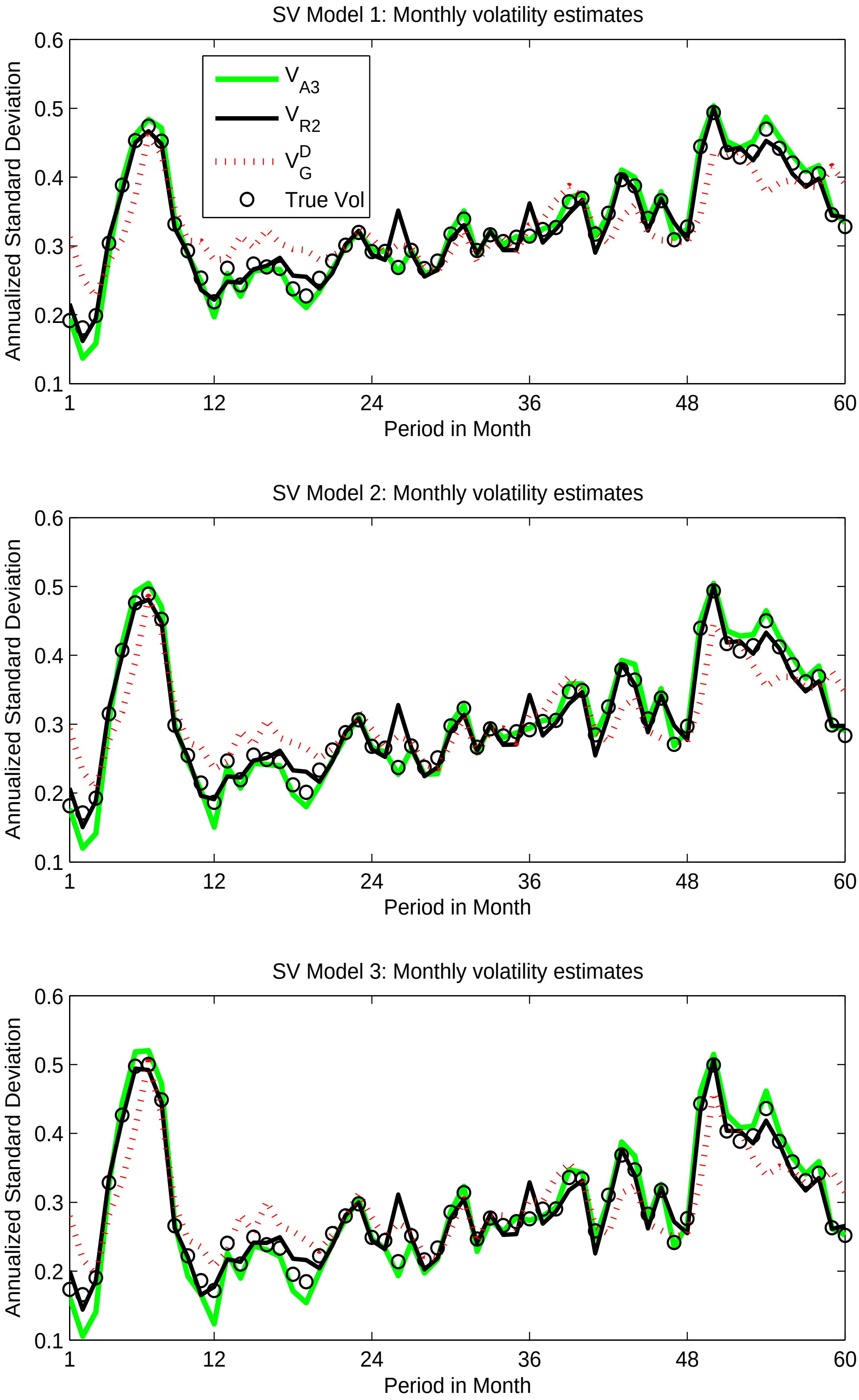


Figure 2.5: Estimation of stochastic volatility with generalized normal overnight price jumps

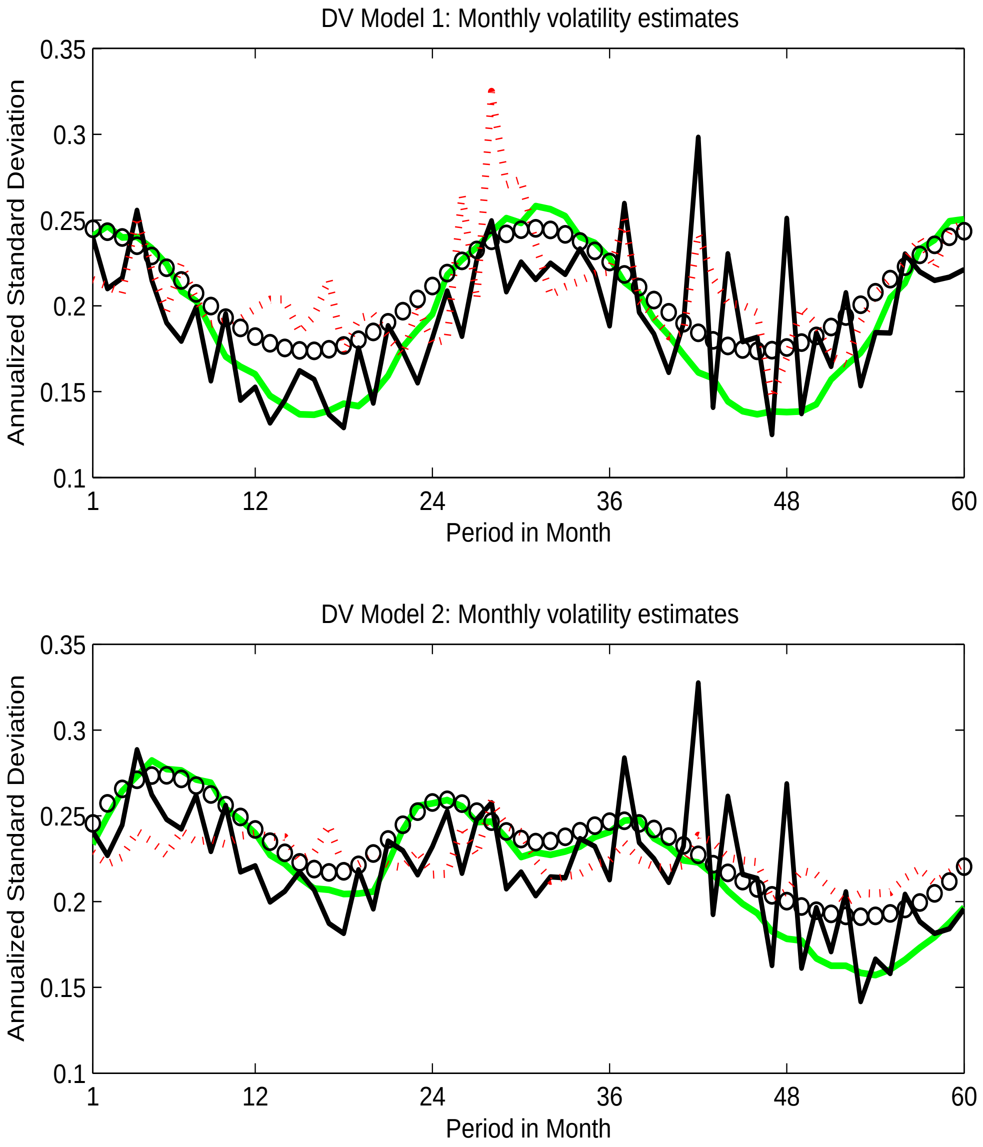


Figure 2.6: Estimation of deterministic volatility with *t* overnight price jumps

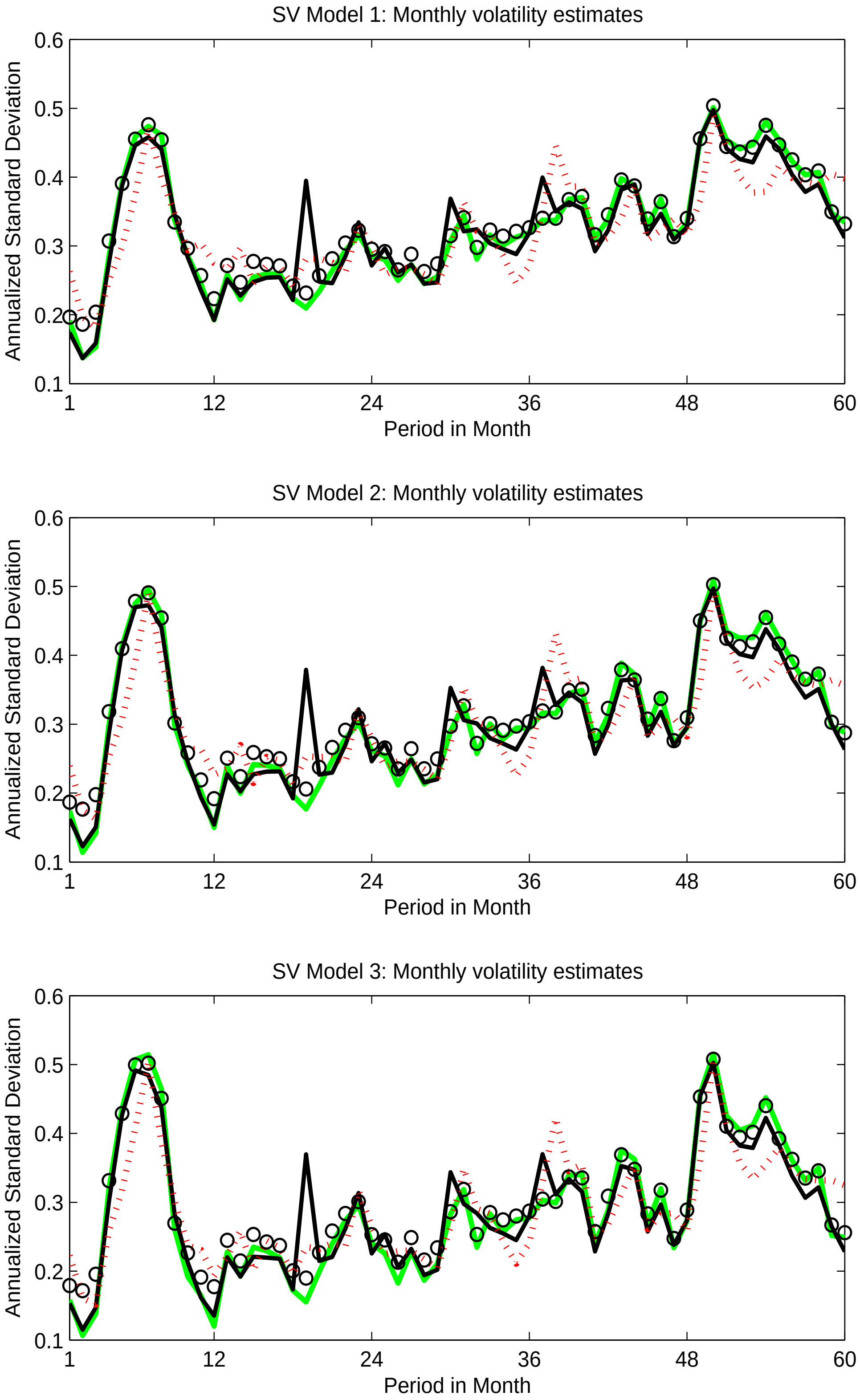
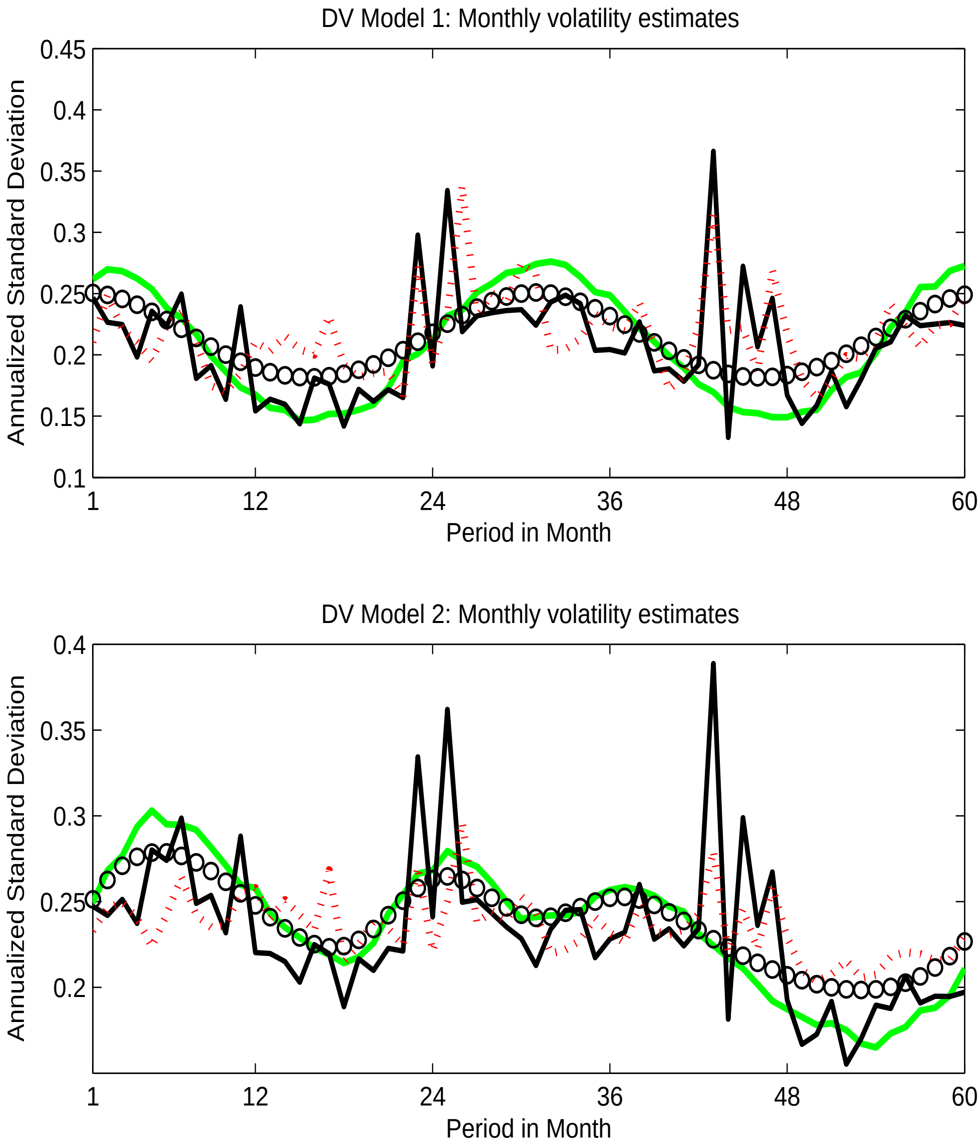


Figure 2.7: Estimation of stochastic volatility with with *t* overnight price jumps


DV Model 1: Monthly volatility estimates

DV Model 2: Monthly volatility estimates

Figure 2.8: Estimation of deterministic volatility with empirical overnight price jumps

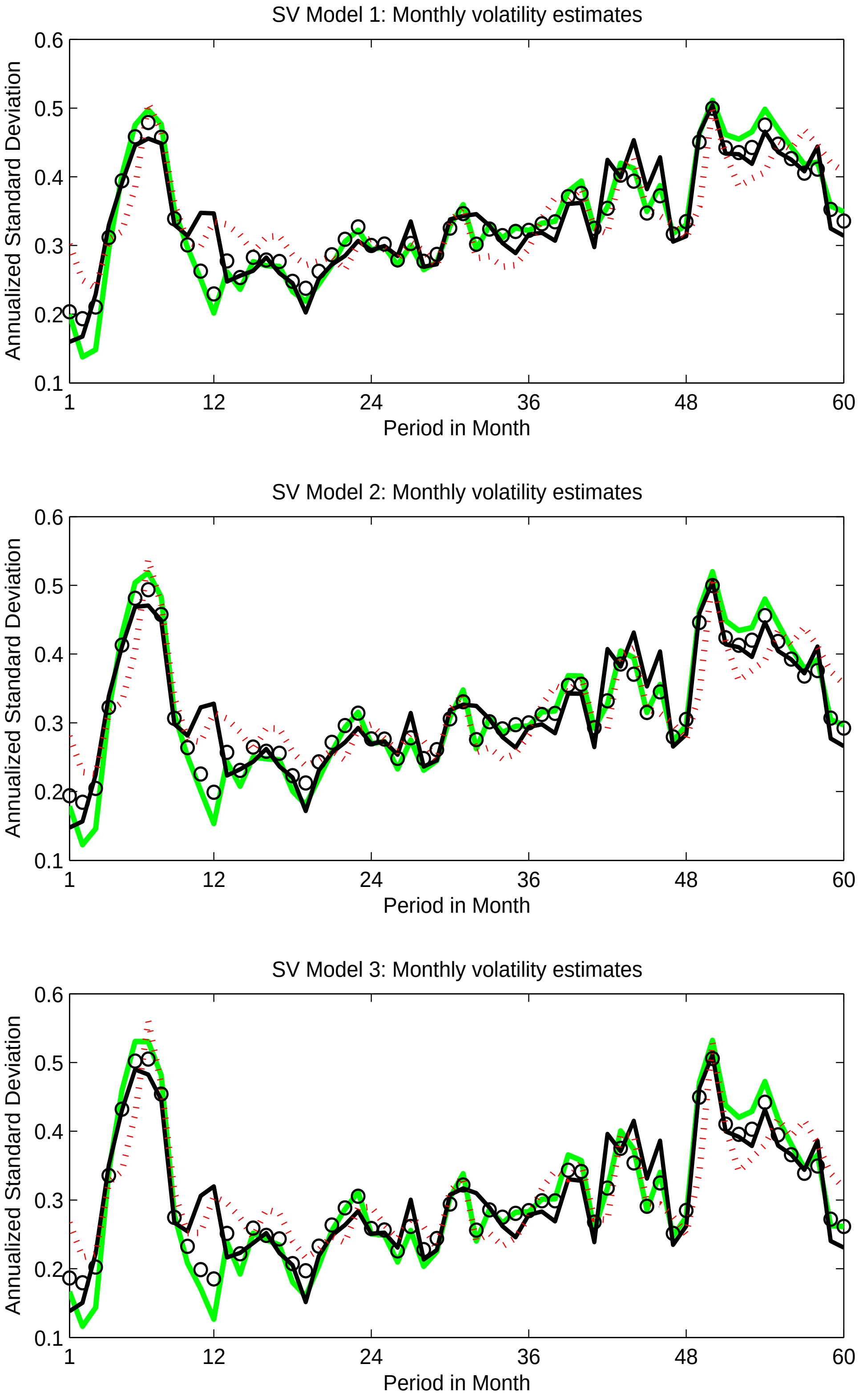


Figure 2.9: Estimation of stochastic volatility with empirical overnight price jumps

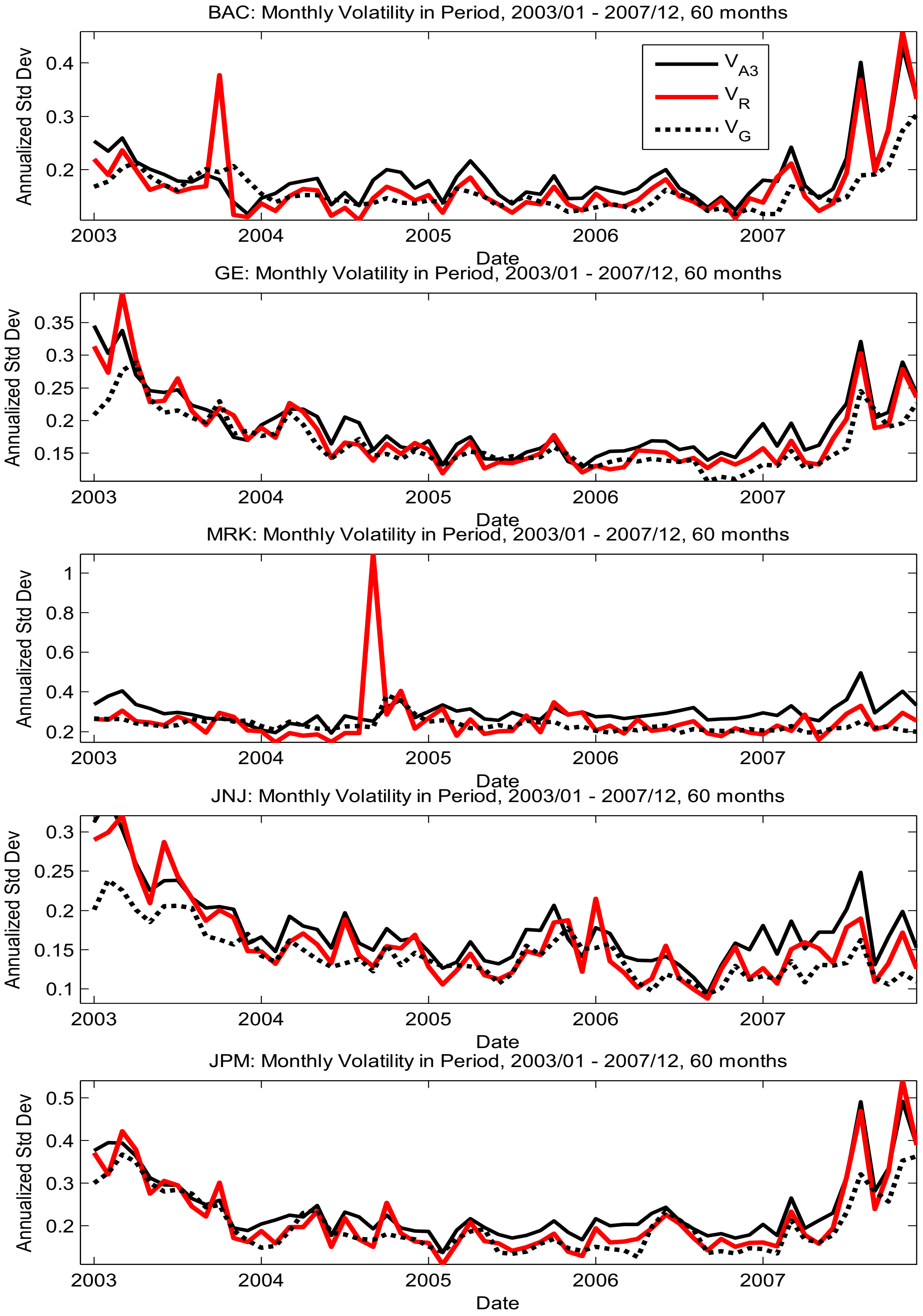
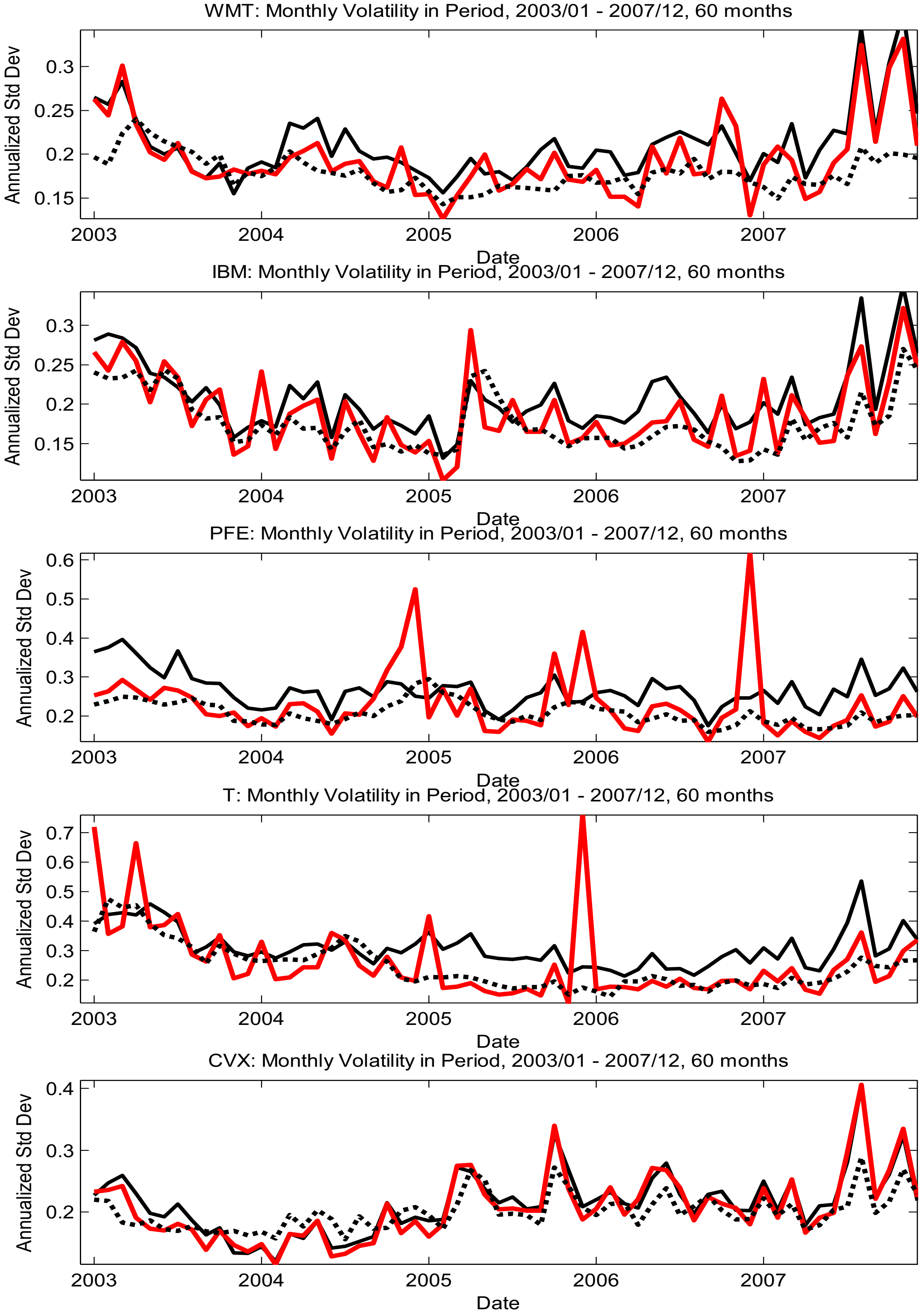


Fig 2.10A: Empirical estimates of monthly volatility


WMT: Monthly Volatility in Period, 2003/01 - 2007/12, 60 months

IBM: Monthly Volatility in Period, 2003/01 - 2007/12, 60 months

PFE: Monthly Volatility in Period, 2003/01 - 2007/12, 60 months

T: Monthly Volatility in Period, 2003/01 - 2007/12, 60 months

CVX: Monthly Volatility in Period, 2003/01 - 2007/12, 60 months

Fig 2.10B: Empirical estimates of monthly volatility

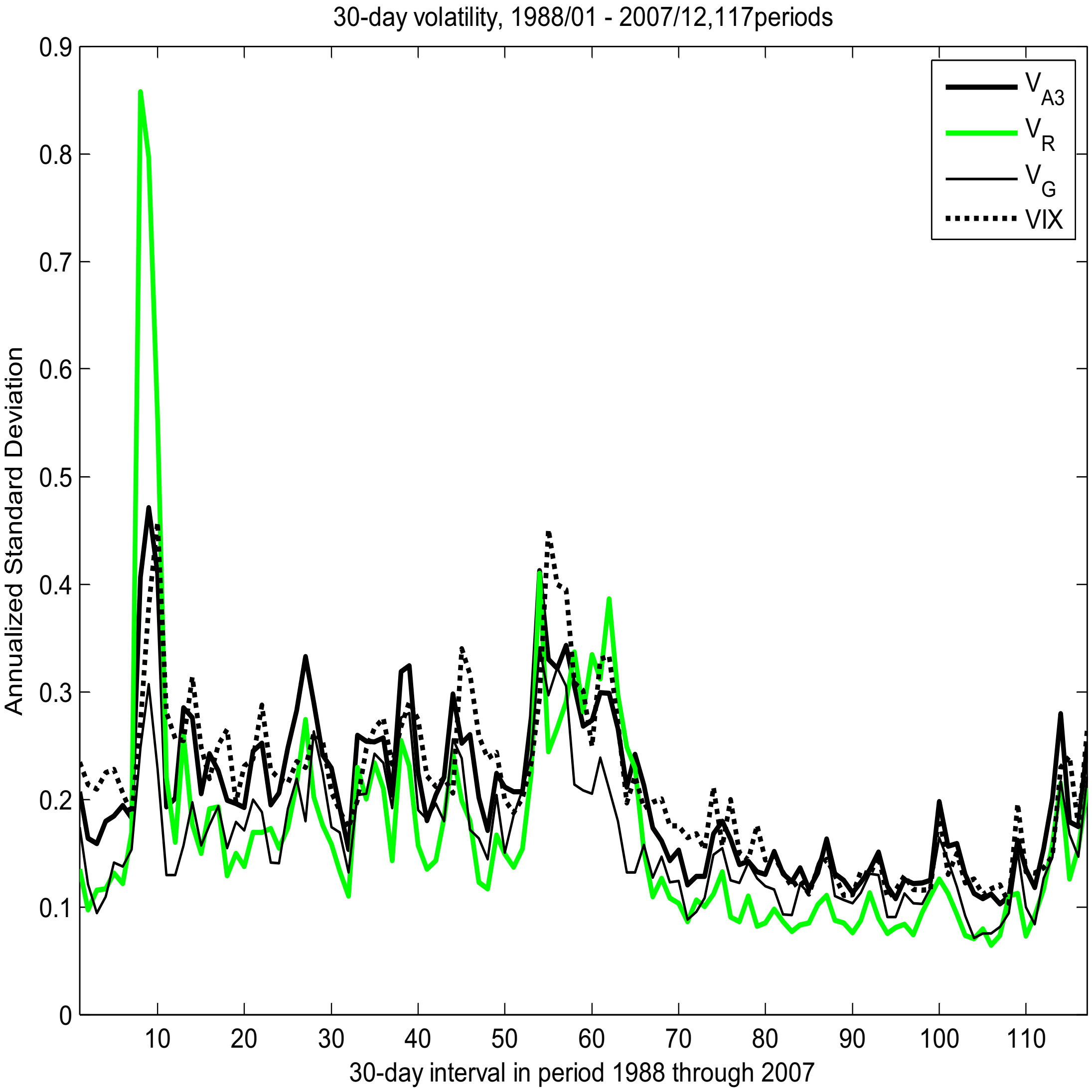


Figure 2.11: VIX and S&P500 30-day volatility estimates

Chapter 3 Trade Volume, Trade Frequency and Trade Size: Their Implications for Informed Trading and Volatility

3.1 Introduction

Trade volume consists of two components: number of trades (trade frequency) and trade size. As pointed out by Jones, Kaul and Lipson (1994), theoretical models of stock trading, such as Wang (1993, 1994) and Harris and Raviv (1993), standardize trades to be of unit size. Thus, these models have little to say in segregating the effects of trade frequency and trade size on return volatility, and they can only provide guidance on the interaction between aggregate trade volume and volatility. Furthermore, most empirical work in the literature treats volume as *exogenous*. This has been criticized by Jones, Kaul and Lipson (1994), who stated that their results "suggest that theoretical models need to endogenize both the frequency and size of trade". This call for action, however, has not yet been answered.

The purpose of this chapter is to propose a statistical model for the aggregate volume of trade. We model trade volume as a compound Poisson distribution, with trade frequency as the *primary* distribution (i.e., the random variable determining the number of summation terms) and trade size as the *secondary* distribution (i.e., the random variable determining the distribution of each term in the summation). Our focus is on the quote-driven (specialist) market, which was the dominant market in the New York Stock Exchange (NYSE) prior to 2006.[1] We consider two types of traders in this market: informed and uninformed traders, who trade against the specialists. Our notion of informed traders differs from that of Easley, Hvidkjaer and O'Hara (2002), who assume that information can only be either good news or bad news, and that trade direction (buy or sell) is completely determined by such

[1] Section 2 describes briefly the introduction of the hybrid market in 2006 and the subsequent dominance of the limit-order market in the NYSE.

news. In contrast, "informed traders" in our model are determined by their motivation, for which divergence of opinion is possible. We assume that the aggregate volume of trades originating from each group of traders follows a separate compound Poisson distribution. Thus, the aggregate volume of all trades is the sum of two independent compound Poisson distributions. In addition, the parameters of the compound Poisson distribution for the informed traders are postulated to depend on some information variables.

Using the estimated aggregate trade volume model, we propose two measures of intensity of informed trading: relative frequency of informed trading (RFIT) and relative volume of informed trading (RVIT), which estimate the relative intensity of informed traders according to their trade frequency and trade volume, respectively. Thus, our model provides a volume-based method to calibrate the intensity of informed trading, which may be used as a measure of asymmetric information. To the best of our knowledge, this chapter is the first attempt to model trade volume, trade frequency and trade size *endogenously*. Furthermore, our measures of the relative intensity of informed trading do not make use of return data. We do not assume *a priori* that volume and volatility are related, as in the case of the MDH approach of Andersen (1996) and Li and Wu (2006). Our approach contrasts that of Li and Wu (2006), who use return data to calibrate their model for volume of trade and estimate the "informed component of volume".

Our empirical analysis involves 50 stocks traded on the NYSE. We use transaction data of trade frequency and trade size over 30-minute intervals to estimate the aggregate-volume model. We use two approaches to estimate the model. First, we use covariates as proxies for information intensity, conditional on the covariates the likelihood function of trade frequency and trade size can be obtained, from which the parameters of the model can be estimated using the MLE method. Alternatively, we may calculate the unconditional moments and cross moments of trade volume and trade frequency, treating the moments of the information intensity as unknown parameters. The parameters of the primary and secondary distributions of the compound Poisson distribution, together with the moments of the information intensity, can then be estimated using the GMM method. For each stock we calculate the daily RFIT and RVIT measures. We then study the effects of informed trade frequency, informed trade volume, uninformed trade frequency and uninformed trade volume on volatility. Our results show that the empirical relation between volume and volatility under the MLE approach is similar to the empirical relation under the GMM approach. Both the MLE and GMM approaches show that trade frequency dominates trade volume in explaining volatility, although trade volume still has incremental information for volatility in the presence of trade frequency. Rather interestingly, while informed trading volume has positive effects on volatility, uninformed trading volume has negative effects on volatility. This result is consistent with Li and Wu (2006), who find negative correlation between volatility and trade volume due to liquidity traders. However, both the informed and uninformed trading frequency has positive effect on volatility.

In sum, this chapter contributes to the literature in several aspects. First, we propose a statistical model for the aggregate trade volume of a stock with informed and uninformed traders. Our model can be estimated using high-frequency transaction data and provides estimated measures of the relative intensity of informed (and uninformed) traders over intraday intervals. While we use these measures in this chapter to study the effects of informed and uninformed trades on volatility, they can also be used to examine return-volume relationships. Second, we contribute to the frequency-size debate using improved econometric methodology. Unlike many studies in the literature that uses daily absolute return as proxy for volatility, we use the integrated conditional variance (ICV) estimate proposed recently by Tse and Yang (2012).[2] Our results confirm the dominance of trade frequency over trade volume in influencing return volatility, although trade volume still has incremental information for volatility beyond trade frequency. Third, in segregating trading frequency/volume into their informed and uninformed components, we find that informed trading volume increase volatility and uninformed trading volume reduce volatility. Surprisingly, both the informed and uninformed components of trade frequency have positive effects on volatility.

The balance of this chapter is as follows. In Section 2 we describe our data, which are high-frequency transaction data extracted from the Trade and Quote (TAQ) database of the NYSE. We present some summary statistics of the data and motivate the construction of the statistical model. The compound Poisson model for the aggregate trade volume originating from both informed and uninformed traders are outlined in Section 3, in which we also discuss the estimation of the model parameters. Section 4 reports the empirical results of the compound Poisson model and the resulting RFIT and RVIT measures of the NYSE stocks. We examine the effects of informed and uninformed trade frequency and volume on volatility in Section 5. Some concluding remarks are summarized in Section 6.

3.2 The Data

The data used in this chapter were extracted and compiled from the TAQ database provided through the Wharton Research Data Services. We downloaded the following variables from the Consolidated Trade (CT) file: date, time, price, and trade size, and the following variables from Consolidated Quote (CQ) file: date, time, bid price, ask price, bid size, and ask size.

From the CT file we compiled the data for the trade frequency over every 30-minute interval and recorded the size of every trade in the interval. We also computed the tick imbalance, which is the (absolute) difference between the number of up-tick and down-tick trades. Due to the unusual large volumes traded at the opening and the closing of trades, we only used trade data between 9:45 and 15:45.

[2] Some authors (see, e.g., Chan and Fong (2006)) use realized volatility as a measure of daily volatility. As shown by Tse and Yang (2012), the ICV estimates perform very well against the realized volatility method, and can provide intraday volatility estimates over short intervals such as half hour.

We selected 50 stocks from the S&P500 index stocks for our study. All selected stocks were without any stock split or stock dividend in the sample period. To economize on space we report detailed graphical results only on a short-list of 10 stocks from the top size-quintile, although the overall conclusions are drawn from the 50 stocks analyzed.[3]

Trading in NYSE went through a transition in 2006. OpenBook was first introduced in 2002 in the NYSE, providing limit-order-book information to traders off the floor. Electronic trading grew from 2002 onwards and hybrid activation was rolled out gradually between October 2006 and January 2007, when the limitation of orders of up to 1,099 shares was removed for immediate automatic execution (Auto-X).[4] Under the hybrid trading system the TAQ database identifies the mode of trading in the field COND. Of the 50 stocks in our sample, regular trading (i.e., a trade without any stated condition in the field COND) under the quote-driven (specialist) system represented 75.53% of the trade frequency in 2005, but dropped to 8.04% in 2007. In contrast, Auto-X trade (a trade with code E for COND) increased from 24.45% in 2005 to 80.58% in 2007. As the model we propose below identifies informed versus uninformed traders based on their trading motivation when they trade against the market makers, we used only data from the specialist market. Specifically, we extracted and used data from regular trading in 2005, before the transition to the hybrid market. Of the 50 stocks in our sample, the proportion of trade frequency under regular trade varied from 67.49% to 85.60%. After filtering out data with abnormality, our sample consists of 245 to 251 days of trades for the 50 stocks. Figure 1 plots the relative frequency distribution of trade size. It can be observed that there are spikes at certain trade sizes, notably 1000 and multiples of thousand. Overall, the relative-frequency diagrams exhibit a profile commensurate with an exponential distribution for the trade size.

There are two possible approaches to model information and calibrate the model: one approach uses some suitable proxies for information and the other approach treats information intensity as random and unobservable. In the first approach, possible candidates for the information proxy are order imbalance, tick imbalance and quote revision. Given the data on the transaction volume and the information proxy, the parameters of the model can be computed using the MLE method. In the second approach, the likelihood function is not available. However, if the parameters of the model are simple linear functions of the intensity of information, unconditional moments of trade volume, trade frequency and trade size can be derived, from which the model parameters can be estimated using the generalized method of moments (GMM). For the first method, the performance of the model may be compromised if the information proxy used is not appropriate. For the second method, although no errors due to a wrong proxy will incur, the model may only adopt simple parametric forms. In this chapter, we adopt both the MLE approach specifying the information

[3] These 10 stocks are Exxon (XOM), General Electric (GE), Procter & Gamble (PG), Johnson & Johnson (JNJ), AT & T (T), Chevron (CVX), JP Morgan Chase (JPM), Wal Mart (WMT), IBM (IBM) and Pfizer (PFE).

[4] See Hendershott and Moulton (2011) for an account of the hybrid transition in the NYSE.

proxy and the GMM approach treating the information proxy as random and parameters to estimate. It is crucial to select the appropriate proxy in the first approach. More recently, many researchers question the order imbalance still being a valid signal of informed trading. Quote revision, however, only stand for the information represented by a small proportion of market participants (market markers). In this chapter, we consider the use of tick imbalance as possible candidate for information proxy in the MLE approach.

3.3 The Model

We assume that there are two types of traders in the quote-driven market: informed traders and uninformed traders. Our definitions for the two groups of traders are based on their trading motivation. Thus, informed traders trade when there is information, private or public, that is made available to them, although their interpretation of the information may differ. On the other hand, uninformed traders do not trade in response to specific information, but for other reasons such as liquidity needs or portfolio rebalancing. Our notion of information is different from that of Easley, Kiefer, O'Hara and Paperman (1996) and Easley, Hvidkjaer and O'Hara (2002), who assume that information can only be either good news or bad news. They assume that good news induces traders to buy and bad news induces traders to sell. In our model we allow informed traders to have divergence in opinions, so that different traders may trade differently in response to the same news. This is also in line with Duarte and Young (2009), who allow the arrival of public news to be interpreted differently by uninformed traders, causing increases in both buy- and sell-orders. Duarte and Young (2009) call the unconditional probability of a trade coming from uninformed traders due to their response to news the probability of symmetric order-flow shock (PSOS).

We consider the trade volume of a stock over a given time interval, set to be 30 minutes in this chapter. Let N be the number of trades (trade frequency) and Y be the trade volume (aggregate number of lots traded) in the time interval. We denote the numbers of trades initiated by the informed and uninformed traders in the interval by N_I and N_U, respectively. Likewise, we denote the volumes of trades initiated by the informed and uninformed traders by Y_I and Y_U, respectively. Thus, the aggregate number of trades N in the interval is given by $N = N_I + N_U$ and the aggregate volume of trades Y is given by $Y = Y_I + Y_U$. If we denote the size of each trade initiated by the informed traders by X_{Ii}, for $i = 1, \cdots, N_I$, and the size of each trade initiated by the uninformed traders by X_{Ui}, for $i = 1, \cdots, N_U$, we have

$$Y_I = \sum_{i=1}^{N_I} X_{Ii}, \tag{3.3.1}$$

and

$$Y_U = \sum_{i=1}^{N_U} X_{Ui}. \tag{3.3.2}$$

We assume that the trade frequencies N_I and N_U are distributed as Poisson variables with means λ_I and λ_U, respectively, so that $N_I \sim \mathscr{P}(\lambda_I)$ and $N_U \sim \mathscr{P}(\lambda_U)$. If we further assume that X_{Ii} are independently and identically distributed (iid) and are independent of N_I, then Y_I follows a compound Poisson distribution. Similar assumptions for X_{Ui} imply that Y_U also follows a compound Poisson distribution. N_I and N_U are called the *primary distributions*, while X_{Ii} and X_{Ui} are called the *secondary distributions*.[5]

As we are modeling stock trades in a quote-driven market in which the transactions are executed against the market makers, we can, in principle, classify a trade as initiated by an informed or uninformed trader according to his trading motivation, i.e., whether the trade is driven by relevant market information or for liquidity needs.[6] To distinguish between the behavior of informed and uninformed traders, we assume that λ_I depends on an information-intensity variable K, so that

$$\lambda_I = \lambda_I(K) = \beta K, \tag{3.3.3}$$

where $\beta > 0$ and K is nonnegative. Thus, the intensity of the trade frequency of informed traders λ_I varies positively with the information intensity. On the other hand, we assume the intensity of the trade frequency of uninformed traders λ_U to be constant. We further assume that informed and uninformed traders trade in different quanta. For informed traders, we assume that their trade-size variable X_{Ii} are distributed exponentially with mean μ_I, i.e., $X_{Ii} \sim$ iid $\mathscr{E}(\mu_I)$. Similarly, for uninformed traders we assume $X_{Ui} \sim$ iid $\mathscr{E}(\mu_U)$.

We denote the size of the ith trade in the interval by X_i, so that

$$X_i = X_{Ii}\mathbf{1}_{\{i\in I\}} + X_{Ui}\mathbf{1}_{\{i\in U\}}, \qquad i = 1, \cdots, N, \tag{3.3.4}$$

where $\mathbf{1}_{\{i\in I\}}$ and $\mathbf{1}_{\{i\in U\}}$ are indicator variables taking value 1 when the ith trade is initiated by informed and uninformed traders, respectively, and zero otherwise. Note that the identity of informed versus uninformed traders is not available from

[5] Compound distributions are sums of iid random variables, where the number of summation terms is random. If the primary distribution (number of summation terms) is Poisson, we have a compound Poisson distribution. Compound distributions have been used extensively in the actuarial science literature to model aggregate insurance losses. See Tse (2009) for further properties of compound Poisson distributions.

[6] In contrast, for an order-driven market in which orders are executed through a computerized order book without the intermediation of dealers, traders may choose to enter a market or limit order. While many authors assume that informed traders always submit market orders, limit orders may also be information-motivated (see Pascual and Veredas (2010)). The interpretation of informed traders in the limit-order literature goes beyond characterization by motivation, as in Goettler, Parlour and Rajan (2009) and Dumitrescu (2010).

the data. Hence, while X_i are observable, $\mathbf{1}_{\{i\in I\}}$ and $\mathbf{1}_{\{i\in U\}}$ are not. The observable total trade volume is (Y_I and Y_U are not observable)

$$Y = Y_I + Y_U = \sum_{i=1}^{N} X_i. \tag{3.3.5}$$

Thus, we have completed the definition of our model of trade frequency N, trade size X_i and trade volume Y. Note that only N, X_i and Y are *observable*, while N_U, N_I, Y_U, Y_I, X_{Ui} and X_{Ii} are not.

To estimate the parameters of the model, two approaches may be considered. First, we may use covariates as proxies for information intensity. Conditional on the covariates, the likelihood function of trade frequency and trade size can be obtained, from which the parameters of the model can be estimated using the MLE method. The MLE approach is facilitated by the independence assumption of the primary and secondary distributions of the compound Poisson model, and that the sum of two compound Poisson distributions has also a compound Poisson distribution (See Tse (2009)). An advantage of this approach is that trade frequency can be modeled as a nonlinear function of the possible candidates of information proxy.[7]

Alternatively, we may calculate the unconditional moments and cross moments of trade volume and trade frequency based on the Poisson assumption for the primary distribution and the exponential assumption for the secondary distribution, treating the moments of the information intensity as unknown parameters. The parameters of the primary and secondary distributions of the compound Poisson distributions, together with the moments of the information intensity, can then be estimated using the GMM. This approach parallels that of Andersen (1996) and Li and Wu (2006) in estimating the enhanced MDH. An advantage of this approach is that data for the information proxy is not required, which will circumvent the problem due to an inappropriate proxy. On the other hand, the moments and cross moments of the trade data are only tractable when the parameters are linear in the information intensity, which limits the applicability of the GMM approach.

3.3.1 The Case of Proxied Information

We now outline the MLE procedure for the estimation of the compound Poisson distribution. Let there be n (30-minute) intervals in the sample, with trade frequency n_i in the ith interval, for $i = 1, \cdots, n$. Denote the (observable) covariate for the information proxy in the ith interval by K_i and the trade sizes by x_{ij}, for $i = 1, \cdots, n$ and $j = 1, \cdots, n_i$. As $N = N_I + N_U$ is the sum of two independent Poisson variates, we have $N \sim \mathscr{P}(\lambda_U + \beta K)$. From equation (3.3.4), $X_i \sim$ iid $w_U \mathscr{E}(\mu_U) + (1 - w_U)\mathscr{E}(\mu_I)$,

[7] We also estimated the model treating $\lambda_I = \beta K^{\alpha}$. however, the parameter estimation of α turns out to be quite small with average value around 0.1 for all the stocks. Empirical results also inferior to the cases when $\lambda_I = \beta K$. In order to compare with the GMM cases we will not present the estimation results with nonlinear dependence of the information proxies.

where

$$w_U = \frac{\lambda_U}{\lambda_U + \beta K}. \tag{3.3.6}$$

If we denote $\mathbf{n} = \{n_1, \cdots, n_n\}$ and $\mathbf{K} = \{K_1, \cdots, K_n\}$, the conditional log-likelihood of $\mathbf{n}$ is

$$\log f_{\mathbf{N}|\mathbf{K}}(\mathbf{n}) = \sum_{i=1}^{n} f_i, \tag{3.3.7}$$

where, dropping the irrelevant constant,

$$f_i = n_i \log(\lambda_U + \beta K_i) - (\lambda_U + \beta K_i). \tag{3.3.8}$$

Furthermore, the conditional log-likelihood of $\mathbf{x} = \{x_{ij}\}$ is

$$\log f_{\mathbf{X}|\mathbf{K}}(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} g_{ij}, \tag{3.3.9}$$

where

$$g_{ij} = \log\left[\frac{\lambda_U}{\lambda_U + \beta K_i}\left[\frac{1}{\mu_U}\right] \exp\left[-\frac{x_{ij}}{\mu_U}\right] + \frac{\beta K_i}{\lambda_U + \beta K_i}\left[\frac{1}{\mu_I}\right] \exp\left[-\frac{x_{ij}}{\mu_I}\right]\right]. \tag{3.3.10}$$

Denoting $\theta = (\lambda_U, \beta, \mu_U, \mu_I)$ as the parameter vector of the model, the joint log-likelihood of $\mathbf{n}$ and $\mathbf{x}$ is

$$L(\theta) = \log f_{\mathbf{N}|\mathbf{K}}(\mathbf{n}) + \log f_{\mathbf{X}|\mathbf{K}}(\mathbf{x}) = \sum_{i=1}^{n} f_i + \sum_{i=1}^{n} \sum_{j=1}^{n_i} g_{ij}. \tag{3.3.11}$$

We denote the MLE of θ by $\hat{\theta}$, with robust variance-covariance matrix $\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$, where

$$\hat{B} = \sum_{i=1}^{n} \frac{\partial f_i}{\partial \theta} \frac{\partial f_i}{\partial \theta'} + \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\partial g_{ij}}{\partial \theta} \frac{\partial g_{ij}}{\partial \theta'} \tag{3.3.12}$$

and

$$\hat{A} = -\frac{\partial^2 L(\hat{\theta})}{\partial \theta \partial \theta'}. \tag{3.3.13}$$

3.3.2 The Case of Unobservable Information

To apply the GMM, we first derive the moments of N and Y up to the third order.[8] The derivation turns out to be more complex when then information variable K

[8] In order to adjust the 30-minute intraday periodicity of trade frequency N and trade volume Y, we simply divide them by their corresponding 30-minute sample average.

is not observable.[9] In this approach, we treat K as a random variable and derive the moments of the trade data as functions of the moments of K, which are some parameters to be estimated, among others.

We define $\mu_K = \mathrm{E}(K)$, so $\mathrm{E}(\lambda_I) = \beta\mu_K$, the unconditional moments involving N and Y are given by

$$\mathrm{E}(N) = \lambda_U + \beta\mu_K = \bar{N} \tag{3.3.14}$$

$$\mathrm{E}(N-\bar{N})^2 = (\lambda_U + \beta\mu_K) + \beta^2\mathrm{E}(K-\mu_K)^2 \tag{3.3.15}$$

$$\mathrm{E}(N-\bar{N})^3 = (\lambda_U + \beta\mu_K) + 3\beta^2\mathrm{E}(K-\mu_K)^2 + \beta^3\mathrm{E}(K-\mu_K)^3 \tag{3.3.16}$$

$$\mathrm{E}(Y) = \lambda_U\mu_U + \beta\mu_K\mu_I = \bar{Y} \tag{3.3.17}$$

$$\mathrm{E}(Y-\bar{Y})^2 = 2(\lambda_U\mu_U^2 + \beta\mu_K\mu_I^2) + \mu_I^2\beta^2\mathrm{E}(K-\mu_K)^2 \tag{3.3.18}$$

$$\mathrm{E}(Y-\bar{Y})^3 = 6(\lambda_U\mu_U^3 + \beta\mu_K\mu_I^3) + 6\mu_I^3\beta^2\mathrm{E}(K-\mu_K)^2 + \mu_I^3\beta^3\mathrm{E}(K-\mu_K)^3 \tag{3.3.19}$$

$$\mathrm{E}(N-\bar{N})(Y-\bar{Y}) = (\lambda_U\mu_U + \beta\mu_K\mu_I) + \mu_I\beta^2\mathrm{E}(K-\mu_K)^2 \tag{3.3.20}$$

$$\mathrm{E}(N-\bar{N})^2(Y-\bar{Y}) = (\lambda_U\mu_U + \beta\mu_K\mu_I) + 3\mu_I\beta^2\mathrm{E}(K-\mu_K)^2 + \mu_I\beta^3\mathrm{E}(K-\mu_K)^3 \tag{3.3.21}$$

$$\mathrm{E}(N-\bar{N})(Y-\bar{Y})^2 = 2(\lambda_U\mu_U^2 + \beta\mu_K\mu_I^2) + 4\mu_I^2\beta^2\mathrm{E}(K-\mu_K)^2 + \mu_I^2\beta^3\mathrm{E}(K-\mu_K)^3 \tag{3.3.22}$$

The parameters θ to be estimated including λ_U, $\beta\mu_K$, $\beta^2\mathrm{E}(K-\mu_K)^2$, $\beta^3\mathrm{E}(K-\mu_K)^3$, μ_U and μ_I.[10] There are 6 parameters to be estimated and 9 unconditional moments, resulting 3 over-identifying restrictions.

[9] The derivation is presented in the Appendix.

[10] Another way to estimate the model is to isolate β from parameters $\beta\mu_K$, $\beta^2\mathrm{E}(K-\mu_K)^2$ and $\beta^3\mathrm{E}(K-\mu_K)^3$, treating β as a separated parameter to be estimated. However, this makes the estimation more complicated and since to compute the RFIT and RVIT only need $\beta\mu_K$, so we treat $\beta\mu_K$, $\beta^2\mathrm{E}(K-\mu_K)^2$ and $\beta^3\mathrm{E}(K-\mu_K)^3$ as three individual parameters.

We denote $g(\theta)$ as the vector of unconditional moments derived from equation (3.3.14) through (3.3.22). To estimate θ using the GMM approach, we calculate the sample moments to match $g(\theta)$. For $i = 1, \cdots, M$, we define f_i as

$$f_i = \Big(N_i, (N_i - \bar{N})^2, (N_i - \bar{N})^3, Y_i, (Y_i - \bar{Y})^2,\ (Y_i - \bar{Y})^3, (N_i - \bar{N})(Y_i - \bar{Y}), (N_i - \bar{N})^2(Y_i - \bar{Y}), (N_i - \bar{N})(Y_i - \bar{Y})^2\Big)', \tag{3.3.23}$$

and let

$$f = \frac{1}{M}\sum_{i=1}^{M} f_i \tag{3.3.24}$$

The GMM estimate of θ is computed as the value that minimizes the objective function

$$Q = (f - g(\theta))' S(\theta)^{-1} (f - g(\theta)) \tag{3.3.25}$$

where

$$S(\theta) = \frac{1}{M}\sum_{i=1}^{M} (f - g(\theta))(f - g(\theta))' \tag{3.3.26}$$

3.3.3 Relative Intensity of Informed Trading

Conditional on the information proxy K_i in interval i, the relative intensity of informed traders measured by trade frequency can be computed as

$$\frac{\beta K_i}{\lambda_U + \beta K_i} \tag{3.3.27}$$

and the relative intensity of informed traders measured by trade volume can be computed as[11]

$$\frac{\mu_I \beta K_i}{\lambda_U \mu_U + \mu_I \beta K_i}. \tag{3.3.28}$$

With the model parameters estimated by MLE, we compute the Relative Frequency of Informed Trading (RFIT) over the n intervals as

$$\text{RFIT} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\beta} K_i}{\hat{\lambda}_U + \hat{\beta} K_i}. \tag{3.3.29}$$

[11] Note that $\mu_I \beta K_i$ and $\lambda_U \mu_U$ are the means of the trade volume in the interval due to informed and uninformed traders, respectively.

Likewise, the Relative Volume of Informed Trading (RVIT) is computed as[12]

$$\mathrm{RVIT} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mu}_I\hat{\beta}K_i}{\hat{\mu}_U\hat{\lambda}_U+\hat{\mu}_I\hat{\beta}K_i}. \tag{3.3.30}$$

However, when the information variable is random and not observable, it has to be estimated. For this purpose, we adopt the approach suggested by Li and Wu (2006) to recover the values of K_i. First, we define v_i as

$$v_i = \begin{bmatrix} N_i - \mathrm{E}(N_i|K_i) \\ Y_i - \mathrm{E}(Y_i|K_i) \end{bmatrix} = \begin{bmatrix} N_i - \mu_N(\theta|K_i) \\ Y_i - \mu_Y(\theta|K_i) \end{bmatrix} \tag{3.3.31}$$

Thus, we have

$$\begin{aligned}
\mathrm{E}(v_i v_i^{'}|\theta,K_i) &= \begin{bmatrix} \mathrm{E}(N_i^2|\theta,K_i)-[\mathrm{E}(N_i|\theta,K_i)]^2 & \mathrm{E}(N_iY_i|\theta,K_i)-\mathrm{E}(N_i|\theta,K_i)\mathrm{E}(Y_i|K_i) \\ \mathrm{E}(N_iY_i|\theta,K_i)-\mathrm{E}(N_i|\theta,K_i)\mathrm{E}(Y_i|K_i) & \mathrm{E}(Y_i^2|\theta,K_i)-[\mathrm{E}(Y_i|\theta,K_i)]^2 \end{bmatrix} \\
&= \begin{bmatrix} \mu_{N2}(\theta|K_i)-\mu_N^2(\theta|K_i) & \mu_{NY}(\theta|K_i)-\mu_N(\theta|K_i)\mu_Y(\theta|K_i) \\ \mu_{NY}(\theta|K_i)-\mu_N(\theta|K_i)\mu_Y(\theta|K_i) & \mu_{Y2}(\theta|K_i)-\mu_Y^2(\theta|K_i) \end{bmatrix} \\
&= \Sigma(\theta,K_i) \\
&= \Sigma_i
\end{aligned} \tag{3.3.32}$$

We compute $\hat{K}_i$ to minimize $L_i = v_i^{'}\Sigma_i^{-1}v_i$, from which we calculate estimates of RFIT and RVIT over the given interval as[13]

$$\hat{\mathrm{RFIT}} = \frac{1}{n}\sum_{i=1}^{n}\frac{[\hat{\beta K}]_i}{\hat{\lambda_U}+[\hat{\beta K}]_i} \tag{3.3.33}$$

and

$$\hat{\mathrm{RVIT}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mu}_I[\hat{\beta K}]_i}{\hat{\mu_U}\hat{\lambda_U}+\hat{\mu}_I[\hat{\beta K}]_i} \tag{3.3.34}$$

[12] RFIT and RVIT can also be calculated over any subinterval in the sample.

[13] We actually compute $[\hat{\beta K}]_i$, since they are always combine together in the model estimation.

3.4 Empirical Results on Modeling Trade Volume

We estimate the compound Poisson model for the 50 selected stocks from NYSE. We use tick imbalance as the information proxy for the MLE approach. In order to compare the model output to the unobservable information case, we standardize the information proxy to have unit mean over the estimation sample period.

We measure the information proxy as the absolute difference between the number of up- and down-tick movements within each 30-minute intervals. The trade size variables X_i are the number of lots traded. The MLE of the 50 stocks are presented in Table 3.1, with standard errors given in parentheses. It can be seen that the results across the 50 stocks are quite similar. Estimates of μ_I are larger than those of μ_U for all stocks except stock DVN, showing that the average trade size of informed traders are larger than that of uninformed traders. Table 3.2 presents the average RFIT and RVIT of the stocks over the sample period. RFIT varies between 9.31% (for WB) and 23.68% (for GE) with a mean of 15.45%, whereas RVIT varies between 9.93% (for DVN) and 58.53% (for GE) with a mean of 32.91%. While the informed traders are responsible for about 15% of the number of trades, their trade volume takes up about 30% of the market. Ratio of the estimates of μ_I to μ_U varies from 0.82 (for DVN) and 6.14 (for GE) with a mean of 3.24.

The GMM of the 50 stocks are presented in Table 3.3, with standard errors given in parentheses. It can be seen that the results across the 50 stocks are quite different, which is contrary to the MLE approach. One reasonable reason is that the ‘true’ information proxy might be quite different for different stocks under our model assumption. For example, for stock A the ‘true’ information proxy is tick imbalance, for stock B is order imbalance and for stock C it is other information proxy under the compound Poisson assumption. Estimates of μ_I are larger than those of μ_U for all stocks except stock DIS, showing that the average trade size of informed traders are larger than that of uninformed traders. Table 3.4 presents the average RFIT and RVIT of the stocks over the sample period. RFIT varies between 0.58% (for RIG) and 86.03% (for MCD) with a mean of 28.19%, whereas RVIT varies between 0.72% (for RIG) and 90.06% (for VZ) with a mean of 47.13%. While the informed traders are responsible for about 30% of the number of trades, their trade volume takes up about 50% of the market. Ratio of the estimates of μ_I to μ_U varies from 0.77 (for DIS) and 6.40 (for GE) with a mean of 3.08.

We compute the means of RFIT and RVIT over each of the 30-minute intervals from 9:45 through 15:45 for the ten short-listed stocks over the sample. In order to economize the space, we only present the result under unobservable information cases. The results are plotted in Figure 3.2. It can be seen that both RFIT and RVIT exhibit intraday periodicity. In particular, there is an intraday “information-intensity smile”, with intensity being the lowest in the (11:45, 12:45) interval for most stocks. The information-intensity smiles based on frequency and volume are quite similar, except that the RVIT curves are much smoother than the RFIT curves.

Lastly we plotted the daily RFIT and RVIT for all the stocks under the unobservable information assumption in our sample period. The results for the ten short-listed stocks are given in Figure 3. As excepted the paths of RFIT and RVIT are quite similar, and both exhibit variations over time in the sample.

3.5 Empirical Results on the Volume-Volatility Relationship

We now re-visit the issue of the effect of trade volume on volatility as studied by Jones, Kaul and Lipson (1994). Using regression models, we examine the effects of trade frequency, trade volume and trade size in the time series context. Furthermore, the regression models incorporates trade frequency and trade volume that are due to informed and uninformed traders separately.

The notations used in our regressions are defined as follows (all on daily frequency): F = trade frequency, V = trade volume, S = average trade size = V/F, RF = RFIT, RV = RVIT, IF = informed trade frequency = F*RF, IV = informed trade volume = V*RV, UF = uninformed trade frequency = F*(1 – RFIT) and UV = uninformed trade volume = V*(1 – RVIT).

The dependent variable VL of the regressions is the daily volatility computed using the ACD-ICV method of Tse and Yang (2012).[14] We regressed VL on one or more of the regressors defined above for each of the 50 stocks. The results of 13 models are summarized in Table 3.5. For each model we present the number of stocks with the computed t-ratio larger than 2 (indicated by a positive number) or less than -2 (indicated by a negative number).[15] The last three columns of the table summarizes the mean, minimum and maximum of $\bar{R}^2$ over the 50 regressions. We assess the comparative importance of the regressors based on the number of cases of extreme t-ratios, as well as the size of $\bar{R}^2$.

The first 6 regressions are regression of volatility on trade frequency, trade volume and trade size, which are presented in Panel A. Models 1 and 2 show that trade frequency explains volatility better than trade volume. However, Models 3 and 4 suggest that both trade volume and average trade size have incremental information on volatility beyond what is contained in trade frequency. Model 5 is misspecified, as both frequency and volume are left out in the model, resulting in low $\bar{R}^2$. Finally, average trade size is found to have negative coefficients in Model 6, as trade frequency, which is negatively correlated with average trade size, is left out of the model.

Panel B and C present regressions of volatility on the information variables constructed by RFIT and RVIT, which are calibrated from the MLE and GMM

[14] The ACD-ICV method estimates the integrated conditional variance (ICV) over an intraday interval using tick data. It is computed as the weighted sum of instantaneous conditional variances estimated from an Autoregressive Conditional Duration (ACD) model.

[15] For example, in Model 6, 49 of the stocks have the t-ratio of V larger than 2, and 45 stocks have the t-ratio of S less than -2.

approaches, respectively. The Regression results in Panel B are similar to the regression results in Panel C, only with numerical difference. Model 7 and Model 8 show that both the relative frequency of informed trading (RFIT and RVIT) calibrated from the MLE(and GMM) approach are informative in explaining volatility. Model 9 show that the trade frequency of informed traders have positive effect in explaining volatility, however, Model 10 show that trade frequency of uninformed traders still have incremental information on volatility. Model 11 and Model 12 show that while trade volume of informed trades increases volatility, trade volume of uninformed trades reduces volatility. Model 13 reinforce the result of Model 3 that informed trade volume have incremental information for volatility beyond what is contained in informed trade frequency. The converse volatility effects of uninformed trade frequency and uninformed trade volume show that liquidity traders' behavior have different effects on volatility. Liquidity traders with small trade sizes (when trade size is small, trade frequency is equivalent to trade volume) have positive effect on volatility, while liquidity traders with moderate or large trade sizes has negative effects on volatility.

In sum, while our results confirm the dominance of trade frequency over trade volume and average trade size in explaining volatility, we find that volume and average trade size have incremental information beyond trade frequency. Surprisingly, the trade frequency and trade volume of uninformed traders have converse effects on volatility. The trade frequencies of both informed and uninformed traders increase volatility. However, the trade volume of informed traders increases volatility, while the trade volume of uninformed trader reduces volatility. From the Volume-Volatility relationship, the relative frequency (volume) of informed trading calibrated from tick imbalance have similar properties with the relative frequency (volume) of informed trading calibrated from the unobservable information assumption.

3.6 Conclusion

We have proposed to model the aggregate trade volume of stocks in a quote-driven market using a compound Poisson distribution. In our model trades may be initiated by informed or uninformed traders, differentiated by their motivation of trade. We assume that the aggregate volume of each group of traders follow a compound Poisson distribution, with the parameters for the distribution of trades due to informed traders dependent on some information variables. We use two approaches to estimate the model. First, we use tick imbalance as proxy for information variable. Conditional on the tick imbalance, MLE method is used to estimate the model; Second, we treat the information variable random and unobservable, GMM method is used to estimate the model. We then calibrate the model and propose measures of relative intensity of informed trading based on trade frequency and trade volume. Our model treats volume endogenously and does not assume *a priori* that volume and volatility are related.

Our empirical analysis of the daily volatility estimates of 50 NYSE stocks confirm that trade frequency dominates trade volume and trade size in affecting volatility. Yet trade volume and trade size have incremental information for volatility beyond that contained in trade frequency. Tick imbalance is an appropriate information proxy under our compound Poisson distribution assumption. Our results also show that informed trading volume increase volatility, while uninformed trading volume reduce volatility. However, for both informed and uninformed traders, the disaggregated effect of trade frequency is to increase volatility. The converse effects of liquidity traders on volatility remains the future research.

Since the transition year of 2007 in which the NYSE started to operate a hybrid market, the limit-order market has gained importance over the specialist market. Electronic trading provides the platform for the development of High Frequency Trading (HFT). The SEC (2010) report showed that trade frequency and trade volume increased substantially from 2005 though 2009, while the average trade size had dropped. While the model proposed in this chapter applies only to the specialist market, in which informed traders are defined against the market maker, the compound Poisson approach proposed in this chapter remains a strong candidate as a statistical tool for modeling trade volume in the limit-order market. The distinction between the information role of market order and limit order, however, has to be carefully studied in developing such a model.

Table 3.1: Estimated models of 50 stocks of the MLE approach

	Parameters			
Stocks	λ_U	β	μ_U	μ_I
XOM	320.3 (2.755)	81.36 (2.259)	1043 (11.78)	3497 (9.358)
C	264.8 (1.694)	76.58 (1.252)	775.4 (4.862)	3573 (14.94)
CVX	319.1 (2.468)	47.68 (1.73)	611.5 (3.419)	1449 (6.338)
GE	231.8 (1.409)	84.64 (1.113)	800.4 (4.396)	4910 (19.32)
GS	194 (1.688)	26.18 (1.307)	602.2 (3.206)	1181 (7.335)
JPM	193.5 (1.353)	61.52 (1.005)	782.3 (5.996)	3655 (14.09)
RIG	240.5 (2.67)	43.23 (1.986)	538 (2.398)	909.9 (6.76)
BAC	207.6 (1.237)	54.46 (0.8097)	694.6 (3.556)	3335 (14.7)
PFE	261.8 (1.838)	90.22 (1.501)	1135 (8.044)	6512 (30.5)
WMT	240.9 (1.66)	78.71 (1.213)	714.6 (4.603)	3585 (15)
FCX	162.6 (1.632)	19.92 (1.249)	479.4 (1.726)	647.5 (14.69)
TXN	264.6 (2.47)	65.82 (1.998)	941 (11.68)	3063 (8.747)
AIG	196.5 (2.355)	57.88 (1.971)	918.6 (17.32)	3287 (10.44)
IBM	235 (1.91)	42.98 (1.447)	558.6 (4.702)	1651 (6.869)
T	115.9 (1.151)	31.47 (0.7313)	776 (6.853)	2962 (15.44)
PG	192.6 (1.291)	54.34 (0.8734)	599.4 (3.552)	2719 (11.85)
HD	205.5 (1.738)	45.78 (1.311)	771.5 (7.86)	2492 (8.498)
HPQ	190.1 (1.312)	57.73 (1.017)	872.9 (6.538)	4433 (20.16)
JNJ	209.8 (1.298)	52.49 (0.8697)	571.9 (3.267)	2446 (10.53)
NEM	220.6 (2.214)	37.54 (1.639)	612 (3.282)	1173 (7.173)
BHI	206.2 (1.951)	27.87 (1.522)	462.9 (1.437)	524.4 (10.55)
MRK	192.3 (1.648)	51.87 (1.24)	804.9 (7.529)	2987 (12.29)
CFC	161.7 (1.42)	23.73 (1.12)	593.7 (4.435)	1376 (7.291)
AA	183.2 (1.512)	23.71 (1.039)	775.8 (5.406)	1890 (8.659)
WB	179.4 (1.533)	19.59 (1.111)	665 (4.134)	1415 (7.392)
GLW	171.8 (1.499)	50.93 (1.032)	872.4 (6.771)	3413 (15.66)
TGT	208.2 (1.827)	31.7 (1.443)	526.4 (2.827)	1046 (6.453)
VZ	192.2 (1.472)	50 (0.9657)	715.4 (4.873)	2970 (13.79)
AMD	171.9 (1.891)	44.98 (1.384)	956.6 (12.87)	3050 (11.9)
BA	182.2 (1.676)	28.62 (1.268)	562.3 (4.171)	1292 (6.279)
X	182.6 (1.695)	28.35 (1.228)	605.7 (4.51)	1418 (7.185)
KO	214.1 (1.662)	35.03 (1.23)	675 (5.659)	1926 (7.25)
DVN	201.2 (1.788)	29.19 (1.291)	540.5 (1.819)	440.5 (10.97)
AXP	147.3 (1.018)	36.51 (0.6097)	521.5 (2.818)	2421 (12.37)
EMC	166.4 (1.215)	45.44 (0.7889)	901.5 (6.728)	4955 (23.57)
TWX	219.9 (1.489)	67.97 (1.036)	983.9 (5.966)	5604 (25.39)
MCD	191.8 (1.259)	43.77 (0.7316)	562.3 (2.558)	2380 (12.06)
DIS	199.5 (1.492)	35.93 (1.105)	773.6 (6.786)	2520 (9.796)
APA	174.4 (1.711)	22.1 (1.379)	470.9 (1.313)	628.8 (11.32)
WYE	192.1 (1.736)	27.31 (1.216)	544.2 (4.036)	1376 (7.136)
WM	144.9 (1.202)	27.49 (0.8943)	583.1 (4.68)	1762 (9.482)
ABT	178.1 (1.338)	21.23 (0.9198)	600.9 (3.91)	1540 (7.77)
DD	193.5 (1.415)	31 (0.936)	463.3 (2.839)	1333 (6.628)
DOW	174.4 (1.427)	29.93 (1.068)	554.2 (4.739)	1568 (6.574)
MDT	176.5 (1.423)	21.12 (1.081)	602.7 (4.04)	1358 (6.65)
MMM	173.6 (1.388)	22.79 (0.9798)	421.6 (2.467)	1034 (6.942)
HON	180 (1.533)	24.36 (1.102)	574.7 (3.593)	1306 (6.977)
KSS	137.1 (1.132)	26.86 (0.7692)	454.2 (3.444)	1463 (7.992)
NSM	139 (1.382)	33.62 (0.9984)	651.1 (6.514)	2032 (9.468)
BMY	156.2 (1.261)	21.2 (0.9831)	934.3 (8.376)	2505 (11.04)

Notes: Information proxy is tick imbalance. Numbers in parentheses are standard errors. The data used are regular trades in 2005. Since the sample average of tick imbalance is 1, so the value of β is also the value of $\beta\mu_K$. We keep 4 significant digital figures in the table.

Table 3.2: Summary statistics of the MLE approach

Stocks	RFIT	RVIT	μ_I/μ_U	Stocks	RFIT	RVIT	μ_I/μ_U
XOM	0.1835	0.3930	3.3527	C	0.2006	0.4803	4.6073
CVX	0.1207	0.2310	2.3691	GE	0.2368	0.5853	6.1351
GS	0.1112	0.1888	1.9613	JPM	0.2144	0.5014	4.6716
RIG	0.1403	0.2084	1.6913	BAC	0.1878	0.4711	4.8016
PFE	0.2245	0.5506	5.7348	WMT	0.2189	0.5229	5.0171
FCX	0.1021	0.1308	1.3508	TXN	0.1789	0.3781	3.2550
AIG	0.2036	0.4333	3.5784	IBM	0.1431	0.3063	2.9552
T	0.1897	0.4214	3.8164	PG	0.1979	0.4752	4.5369
HD	0.1652	0.3560	3.2299	HPQ	0.2061	0.5044	5.0788
JNJ	0.1809	0.4368	4.2772	NEM	0.1342	0.2186	1.9171
BHI	0.1111	0.1231	1.1330	MRK	0.1900	0.4209	3.7106
CFC	0.1185	0.2235	2.3174	AA	0.1070	0.2124	2.4355
WB	0.0931	0.1717	2.1283	GLW	0.2002	0.4405	3.9123
TGT	0.1226	0.2074	1.9876	VZ	0.1858	0.4372	4.1519
AMD	0.1841	0.3789	3.1888	BA	0.1260	0.2350	2.2978
X	0.1248	0.2360	2.3416	KO	0.1302	0.2775	2.8527
DVN	0.1178	0.0993	0.8150	AXP	0.1805	0.4541	4.6422
EMC	0.1900	0.4927	5.4964	TWX	0.2090	0.5288	5.6960
MCD	0.1671	0.4099	4.2337	DIS	0.1401	0.3169	3.2574
APA	0.1057	0.1340	1.3353	WYE	0.1160	0.2340	2.5288
WM	0.1460	0.3132	3.0219	ABT	0.1000	0.2085	2.5634
DD	0.1275	0.2738	2.8762	DOW	0.1351	0.2842	2.8299
MDT	0.1005	0.1914	2.2532	MMM	0.1088	0.2176	2.4528
HON	0.1115	0.2103	2.2726	KSS	0.1504	0.3331	3.2204
NSM	0.1752	0.3635	3.1203	BMY	0.1112	0.2340	2.6810

Table 3.3: Estimated models of 50 stocks of the GMM approach

	Parameters			
Stocks	λ_U	$\beta\mu_K$	μ_U	μ_I
XOM	255.6 (10.02)	87.09 (4.181)	1730 (7.176)	5807 (265.2)
C	237.2 (17.46)	39.55 (8.953)	1709 (4.384)	6448 (62.05)
CVX	209.9 (18.14)	36.69 (8.838)	1358 (38.84)	2219 (121.8)
GE	196.3 (6.508)	74.19 (4.78)	2161 (6.868)	7463 (188.8)
GS	134.1 (33.4)	44.53 (24.07)	1064 (9.226)	1780 (151.3)
JPM	163.6 (5.527)	41.23 (4.127)	1584 (2.612)	5794 (39.98)
RIG	245.5 (53.02)	1.442 (55.59)	943.9 (53.03)	1159 (81.55)
BAC	177.6 (7.849)	58.39 (6.753)	1579 (11.6)	6624 (242.3)
PFE	209.1 (8.441)	93.34 (3.307)	2785 (7.627)	11280e (376.5)
WMT	175.9 (14.11)	78.61 (11.63)	1793 (19.69)	8168 (1007)
FCX	106.8 (22.51)	19.47 (13.35)	946.6 (25.98)	1314 (119.4)
TXN	271.8 (8.086)	2.643 (5.561)	2050 (3.291)	4935 (65.11)
AIG	219.1 (12.39)	15.98 (12.38)	3676 (15.34)	9627 (615.5)
IBM	153.5 (15.01)	63.09 (8.745)	804.9 (2.835)	2243 (77.54)
T	72.64 (5.687)	44.99 (3.683)	2103 (1.922)	4744 (179.5)
PG	158.7 (6.625)	57.7 (5.252)	1514 (15.18)	6279 (535.4)
HD	189.5 (9.678)	21.43 (6.86)	1696 (2)	3174 (80.33)
HPQ	153.1 (10.11)	57.27 (3.967)	2187 (10.21)	9225 (526.7)
JNJ	243.1 (12.96)	12.43 (14.18)	1850 (9.15)	4085 (367.6)
NEM	132.5 (19.85)	34.12 (9.217)	758.1 (32.13)	2380 (170.4)
BHI	139.7 (66.11)	10.97 (39.75)	1011 (91.29)	1323 (201)
MRK	93.1 (13.2)	103.4 (17.05)	1412 (0.2644)	8635 (0.326)
CFC	75.4 (16.93)	66.47 (6.564)	812.4 (2.993)	2866 (251.1)
AA	109.7 (5.883)	47.23 (3.704)	1127 (1.791)	2982 (66.8)
WB	145.5 (9.125)	3.107 (4.904)	1244 (7.139)	2231 (109.6)
GLW	104 (6.599)	6.015 (1.687)	730.8 (4.519)	6164 (208.3)
TGT	95.29 (19.37)	103.3 (9.343)	819.9 (3.47)	1891 (166.5)
VZ	39.49 (14.4)	127.7 (11.1)	1472 (3.271)	4122 (920)
AMD	113.1 (12.37)	46.77 (6.473)	1760 (7.917)	8326 (379.9)
BA	127.8 (12.1)	21.06 (7.529)	1180 (13.17)	1703 (85.78)
X	134.8 (44.28)	16.02 (18.54)	714 (6.936)	4175 (3.607)
KO	159.8 (18.41)	24.66 (11.43)	1438 (16.05)	2166 (111.6)
DVN	104.2 (36.84)	45.49 (10.26)	892.5 (44.98)	1563 (279.5)
AXP	58.72 (10.22)	92.5 (3.291)	908.7 (4.131)	4049 (476.8)
EMC	101 (10.5)	33.11 (8.242)	2849 (1.126)	7760 (26)
TWX	111 (20.52)	82.06 (9.976)	2715 (4.698)	9732 (358.3)
MCD	21.86 (8.573)	134.7 (8.263)	1293 (0.8475)	3229 (536.2)
DIS	111 (22.94)	94.67 (27.14)	2410 (10.37)	1863 (108.6)
APA	124.6 (14.72)	3.556 (9.487)	949.8 (19.11)	1187 (53.67)
WYE	178 (17.87)	10.07 (13.76)	711.3 (10.46)	4552 (27.64)
WM	86.13 (8.665)	45.92 (4.987)	1139 (1.342)	2494 (74.55)
ABT	75.27 (5.888)	74.97 (2.355)	703.2 (1.55)	2550 (101.4)
DD	122.9 (5.676)	73.14 (2.292)	663.6 (2.194)	2154 (98.7)
DOW	109.4 (12.39)	23.2 (5.571)	1020 (4.272)	2780 (116.6)
MDT	93.02 (7.398)	80.35 (4.256)	892.9 (1.449)	1796 (85.66)
MMM	97.01 (35.59)	74.79 (48.77)	642.3 (107)	1777 (655.8)
HON	67.06 (8.938)	94.53 (4.897)	851.8 (1.679)	2144 (162.5)
KSS	38 (7.674)	60.83 (3.166)	809 (3.825)	2123 (127.6)
NSM	104.8 (10.78)	36.32 (9.176)	1404 (10.44)	4991 (305)
BMY	122.9 (4.56)	1.61 (2.638)	1082 (2.373)	4511 (23.37)

Notes: Information proxy is unobservable. Numbers in parentheses are standard errors. The data used are regular trades in 2005.We keep 4 significant digital figures in the table.

Table 3.4: Summary statistics of the GMM approach

Stocks	RFIT	RVIT	μ_I/μ_U	Stocks	RFIT	RVIT	μ_I/μ_U
XOM	0.2541	0.5336	3.3574	C	0.1429	0.3862	3.7737
CVX	0.1488	0.2221	1.6335	GE	0.2743	0.5663	3.4542
GS	0.2493	0.3572	1.6729	JPM	0.2013	0.4797	3.6583
RIG	0.0058	0.0072	1.2282	BAC	0.2474	0.5797	4.1961
PFE	0.3086	0.6437	4.0486	WMT	0.3089	0.6706	4.5562
FCX	0.1542	0.2019	1.3883	TXN	0.0096	0.0229	2.4069
AIG	0.0680	0.1604	2.6188	IBM	0.2912	0.5338	2.7865
T	0.3824	0.5828	2.2557	PG	0.2666	0.6012	4.1473
HD	0.1016	0.1747	1.8714	HPQ	0.2723	0.6122	4.2183
JNJ	0.0487	0.1015	2.2082	NEM	0.2047	0.4469	3.1388
BHI	0.0728	0.0932	1.3086	MRK	0.5261	0.8716	6.1146
CFC	0.4685	0.7567	3.5283	AA	0.3010	0.5327	2.6470
WB	0.0209	0.0369	1.7936	GLW	0.0547	0.3279	8.4341
TGT	0.5201	0.7142	2.3061	VZ	0.7638	0.9006	2.8006
AMD	0.2926	0.6618	4.7315	BA	0.1415	0.1921	1.4423
X	0.1062	0.4100	5.8476	KO	0.1337	0.1886	1.5066
DVN	0.3039	0.4332	1.7510	AXP	0.6117	0.8753	4.4554
EMC	0.2468	0.4716	2.7239	TWX	0.4250	0.7260	3.5849
MCD	0.8603	0.9390	2.4982	DIS	0.4602	0.3972	0.7729
APA	0.0277	0.0344	1.2501	WYE	0.0535	0.2658	6.3999
WM	0.3477	0.5387	2.1903	ABT	0.4990	0.7832	3.6260
DD	0.3731	0.6589	3.2457	DOW	0.1749	0.3663	2.7270
MDT	0.4635	0.6346	2.0108	MMM	0.4353	0.6808	2.7669
HON	0.5850	0.7801	2.5166	KSS	0.6155	0.8077	2.6243
NSM	0.2573	0.5519	3.5539	BMY	0.0129	0.0518	4.1684

Table 3.5: Summary of regressions

Panel A: Regressions of trade frequency, trade volume and trade size

Model	F	V	S	RF	RV	IF	IV	UF	UV	$\bar{R}^2_{mean}$	$\bar{R}^2_{min}$	$\bar{R}^2_{max}$
1	+49									0.2613	0.0231	0.5316
2		+49								0.1385	-0.0037	0.3154
3	+46	+12								0.2765	0.0564	0.5304
4	+49		+9							0.2754	0.0547	0.5301
5			+22							0.0369	-0.0040	0.2146
6		+49	-45							0.2647	0.0566	0.5103

Panel B: Regressions using RFIT and RVIT calibrated from the MLE approach

Model	F	V	S	RF	RV	IF	IV	UF	UV	$\bar{R}^2_{mean}$	$\bar{R}^2_{min}$	$\bar{R}^2_{max}$
7				+50						0.1963	0.0360	0.3595
8					+50					0.1739	0.0312	0.3329
9						+50				0.2980	0.0687	0.5564
10						+50		+23		0.3175	0.0651	0.5803
11							+49			0.2012	0.0182	0.3747
12							+49		-23	0.2258	0.1044	0.4233
13						+46	+11			0.3089	0.1328	0.5549

Panel C: Regressions using RFIT and RVIT calibrated from the GMM approach

Model	F	V	S	RF	RV	IF	IV	UF	UV	$\bar{R}^2_{mean}$	$\bar{R}^2_{min}$	$\bar{R}^2_{max}$
7				+48						0.2397	0.0063	0.4565
8					+47					0.2201	0.0038	0.4555
9						+48				0.2551	0.0203	0.5264
10						+47		+16		0.2684	0.0256	0.5304
11							+49			0.1918	-0.0009	0.4156
12							+49		-35	0.2558	0.0474	0.4964
13						+39	+15			0.2701	0.0467	0.5271

Notes: The data used are regular trades in 2005. Volatility is estimated using the ACD-ICV method. $+n$ ($-n$) means n stocks have regression coefficient with a t ratio larger (less) than 2 (-2). $\bar{R}^2_{mean}$, $\bar{R}^2_{min}$ and $\bar{R}^2_{max}$ are, respectively, the mean, minimum and maximum of $\bar{R}^2$ of the 50 regressions. Panel A shows the regressions of volatility on trading frequency, trade volume and trade size. Panel B and Panel C show the regressions adopting RFIT and RVIT calibrated from the MLE and GMM approaches, respectively.

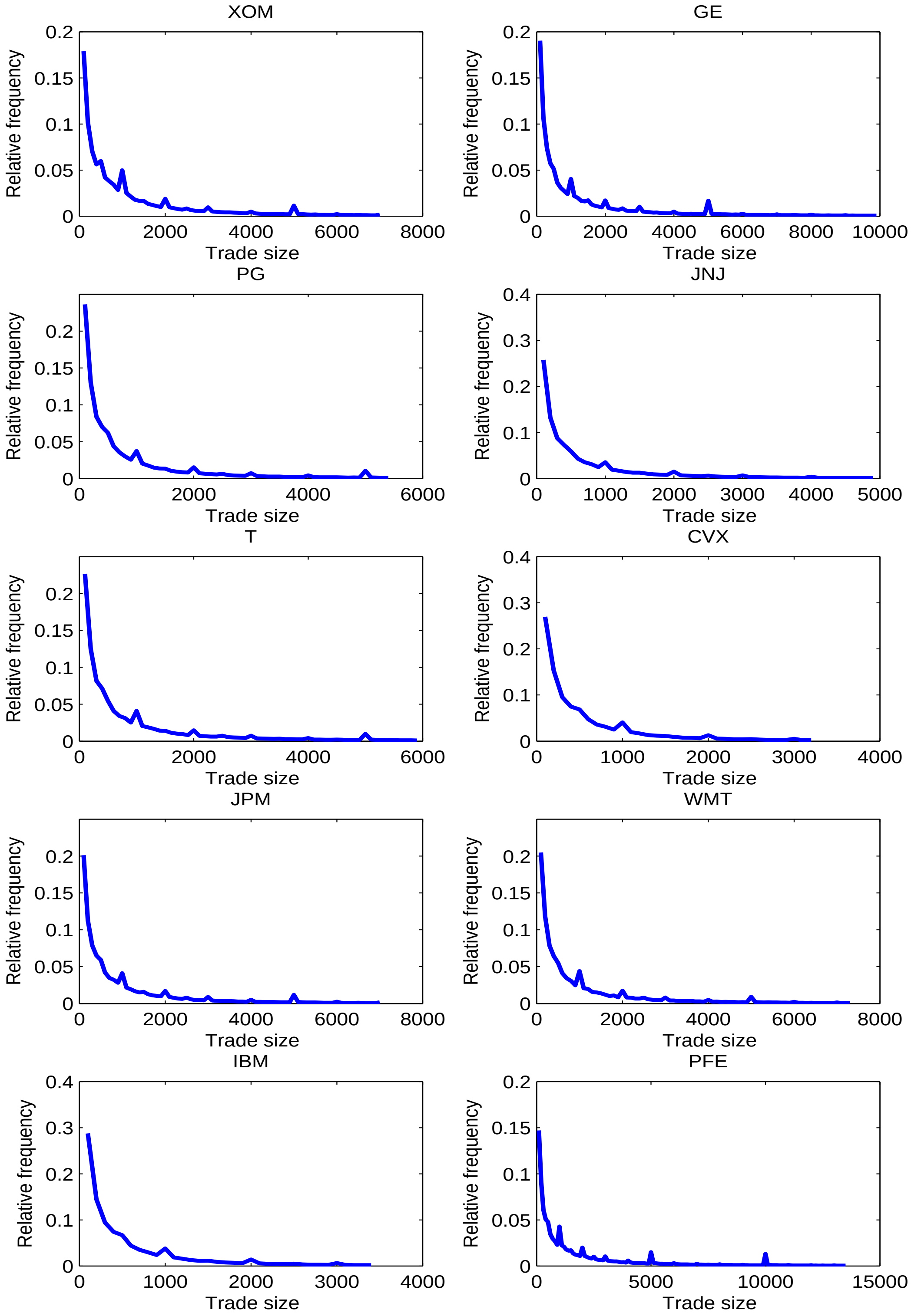


Figure 3.1: Relative frequency of trade size

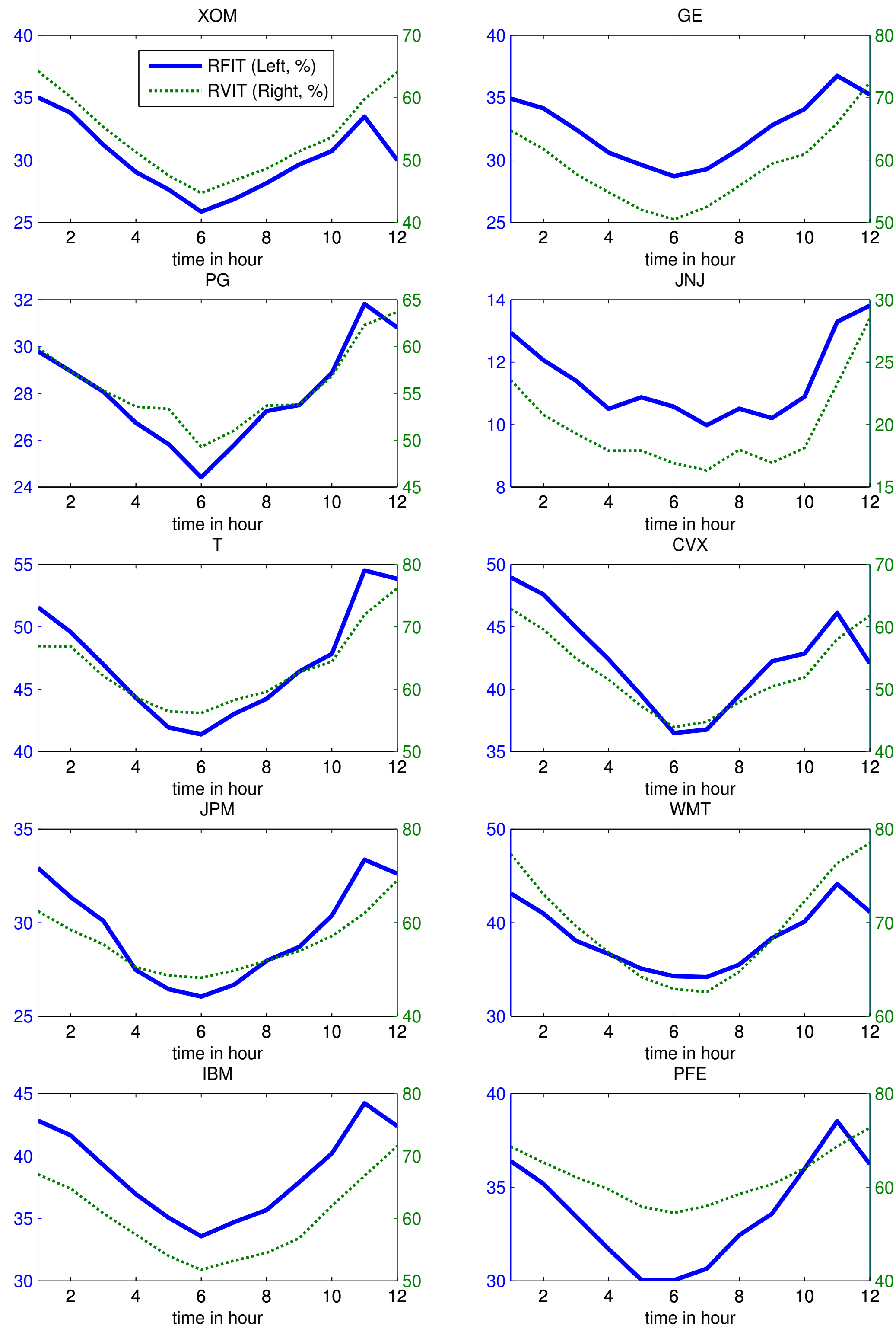


Figure 2: Intraday information profile

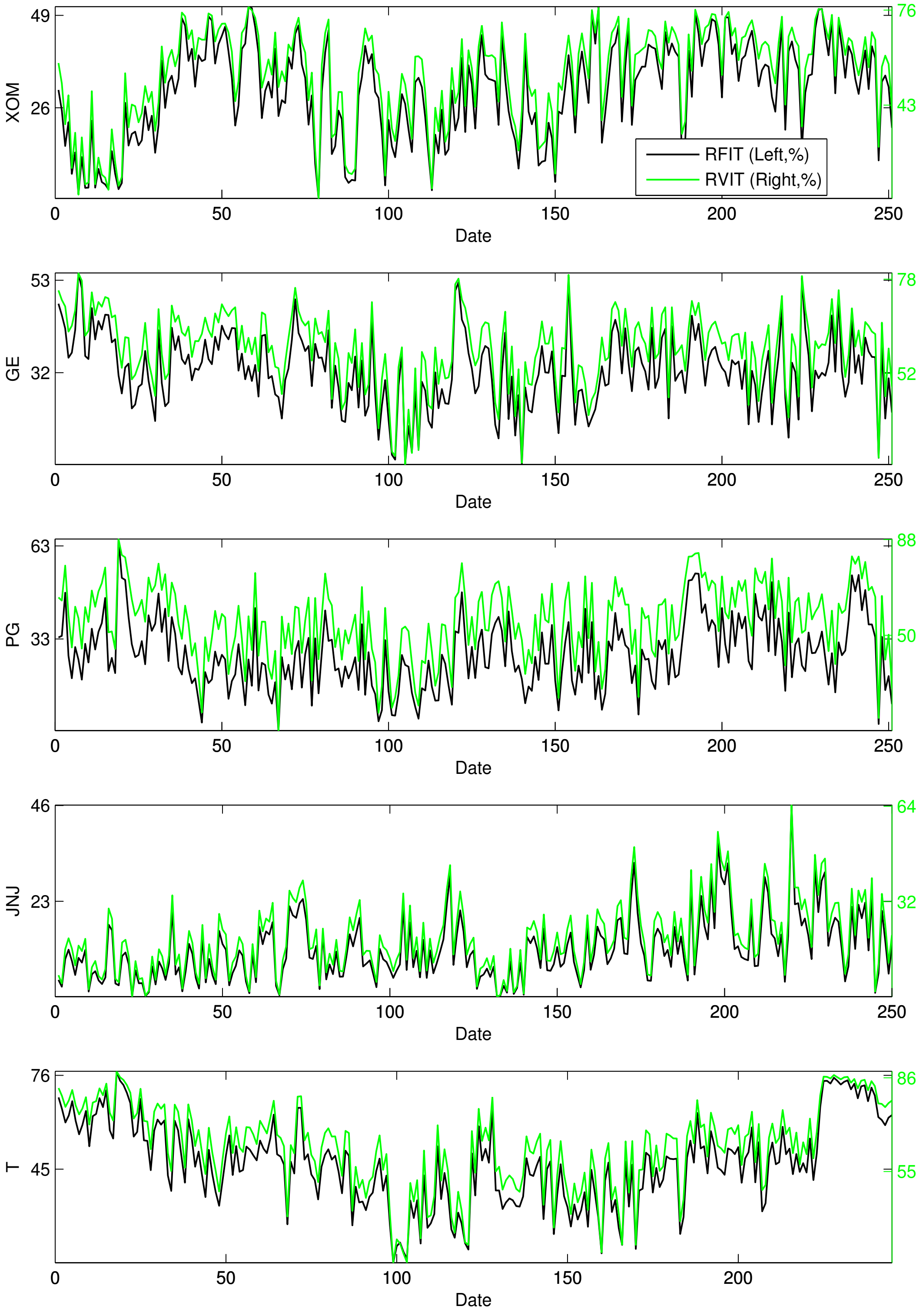


Figure 3A: Estimates of daily RFIT and RVIT

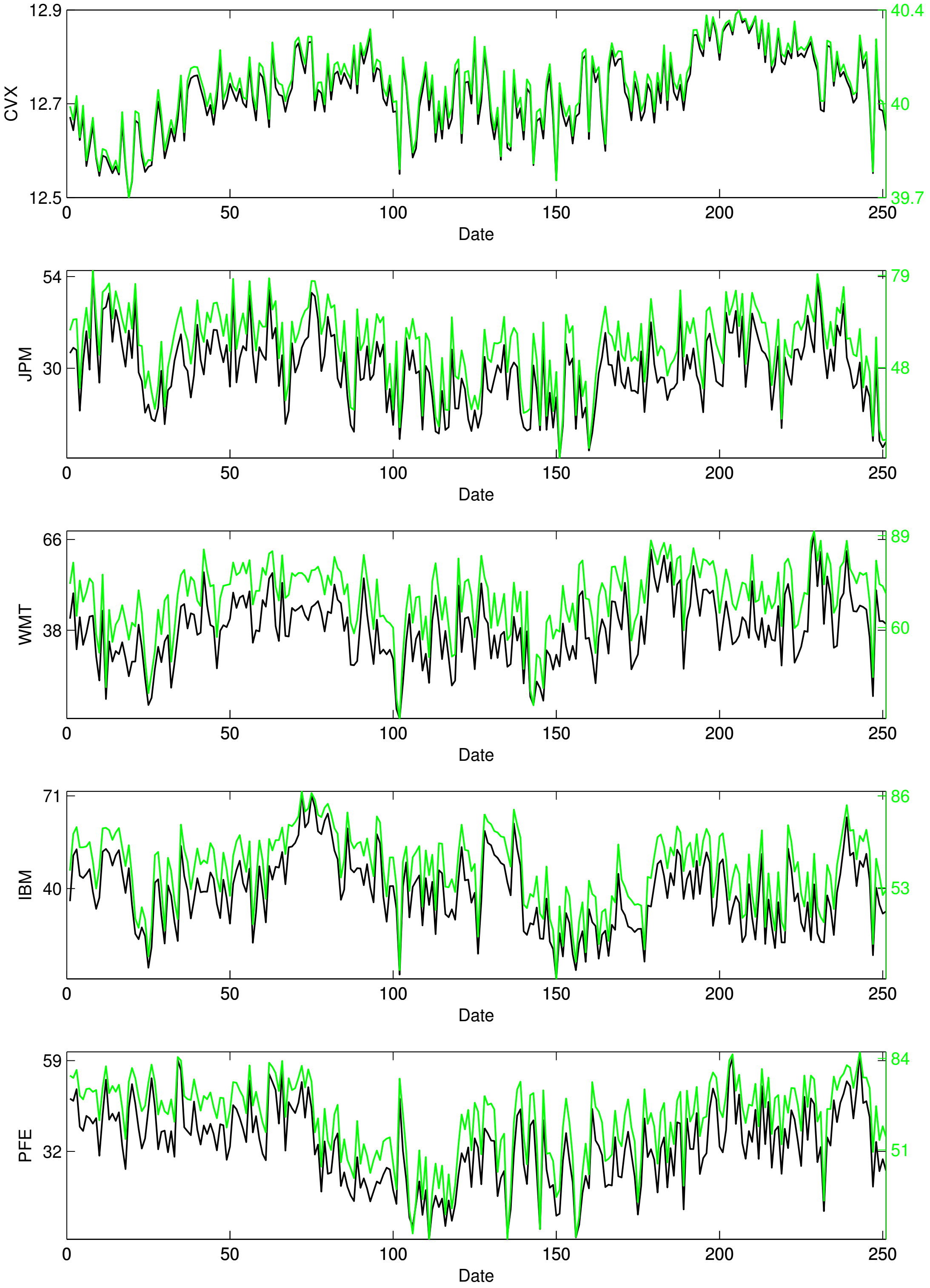


Figure 3B: Estimates of daily RFIT and RVIT

Chapter 4 Intraday Value at Risk: Asymmetric Autoregressive Conditional Duration Approach

4.1 Introduction

This chapter proposes a method to compute intraday Value at Risk (IVaR) using real-time high-frequency transaction data, which takes into account the instantaneous market events. With the rapid growth of high-frequency trading, transactions can be done within a second, millions of data points flow out of the stock markets driving investors' and traders' decision logic. In Table 4.1 the average daily transactions during the period of January 2008 to December 2010 for ten selected large-cap stocks are all above 10,000, with the average duration per trade ranging from 0.98 second for JPM to 2.17 seconds for IBM.[1] While the number of trades per day reduces dramatically after combining the trades in the same time stamp into one trade, it is still larger than 5,000 for all the stocks. Due to decimalization, the minimum tick size for the New York Stock Exchange (NYSE) is reduced to one cent. From Table 1, around half of the transactions of the selected 10 stocks are at tick zero (no price change), more than 5% of the transaction price changes are greater than or equal to 5 ticks, which leads to large price jumps in ticks and a larger range of observable discrete trade-by-trade price jumps. Since transaction price changes are quoted as multiples of the smallest divisor, the use of a continuous distribution to characterize price changes is not appropriate for stocks with high transaction intensities.

A problem in using high-frequency data is that the raw data contain excessive noise. The aim of applying econometric and other computational methods to high-frequency data is to filter noise and extract information. In the microstructure literature, due to the presence of transaction costs, a rational informed trader will only trade when the current price deviates from his/her estimate of the "true price" by more than the transaction cost incurred.

[1] We select 10 large-cap stocks traded on the NYSE for reporting convenience, information for other stocks is exhibited in Table 2.

We propose to sample the high-frequency data using price duration events to filter out the microstructure noise. Our method for estimating and evaluating the IVaR is based on an extension of the asymmetric autoregressive conditional duration (AACD) model of Bauwens and Giot (2003). For a pre-determined threshold δ, a price-event is triggered if the cumulative price change (either upwards or downwards) exceeds δ, and the time taken to observe the event is the duration. Upon modeling the price movements and the durations of price-events *jointly*, we forecast the next price movement and the price-event time. We then employ a simulation approach to estimate the IVaR within a pre-determined time horizon.

In providing intraday information, IVaR is a useful tool to define risk profiles, monitor risk and measure performance for traders. IVaR was first proposed by Giot (2005), who aggregates the irregularly-spaced high-frequency data to retrieve a regularly-sampled intraday returns and employs Gaussian GARCH, Student's t GARCH and RiskMetrics models to quantify IVaR. He also employs a duration (log-ACD) model to access the irregularly-spaced high-frequency data. Based on an empirical study on the NYSE stocks, he shows that the Student's t GARCH model performs the best. Dionne, Duchesne and Pacurar (2009) investigate the use of tick-by-tick data and estimate IVaR by intraday Monto Carlo simulations using irregularly-spaced high-frequency data. Coroneo and Veredas (2011) propose to estimate IVaR using quantile regression for regularly-spaced high-frequency financial data.

There are important advantages of our approach over the above methods. First, we employ price-duration point process to filter out the microstructure noise, which is crucial in modeling high-frequency transaction data. Second, we model the price movements and price durations *jointly* using AACD model, which avoids modeling the intraday return distribution. Third, we employ a time transaction method to adjust the intraday seasonality pattern of the price duration process, which allows us to switch from calendar time and diurnally-adjusted time conveniently for evaluating IVaR. Finally, we use all the information before the forecast IVaR intervals, which makes the IVaR forecast more accurate.

The autoregressive conditional duration (ACD) model is proposed by Engle and Russell (1998) to analyze the durations between transactions, irrespective of whether they correspond to a price increase or decrease. Bauwens and Giot (2003) extend the ACD model to study the mid-point of bid-ask quotes. They propose a two-state AACD model to analyze mid-price decrease and increase jointly with trade duration. In their model, the conditional expected duration of each state varies with conditional information. This conditional information may include lagged duration, lagged volume, and lagged spread. Recently, such models have been expanded rapidly with contribution by Gramming and Maurer (2000), Zhang, Russell and Tsay (2001), Engle and Lunde (2003), Gourieroux, Jasiak and Le Fol (2004), and Fernandes and Gramming (2005), among others. Paccurar (2008) provides a comprehensive suvey of ACD models.

We apply the AACD model to a two-state point process for price movements, where the two states represent an upward or a downward price movement of a pre-determined threshold δ. Following Bauwens and Giot (2003), we allow the expected duration to vary with the lagged durations, and the lagged conditional expected durations. Using an intraday Monte Carlo simulation approach, with information for the price movements before a stated time, we simulate the price movements starting from the stated time for any horizon during the trading hours. We study all index stocks of the S&P 500 traded on the NYSE for three different periods during and after the 2008 global financial crisis. Our empirical results of 30-min IVaR backtesting show that the AACD approach outperforms other IVaR evaluation methods. IVaR can be computed for any time horizon once the AACD model has been estimated without requiring new sampling and estimation when the time horizon changes, due to the the flexibility of irregularly-spaced information. 60-min IVaR backtesting results also indicate that the AACD approach performs well against other methods.

The structure of this chapter is as follows. Section 2 summarizes the IVaR methods considered for comparison in this chapter. Section 3 presents the IVaR evaluation method employing the AACD model. Section 4 describes the IVaR backtesting methods used in our paper. In Section 5 we describe our data, which are high-frequency transaction data extracted from the Trade and Quote (TAQ) database of the NYSE. Section 6 reports the empirical results of our study. Section 7 concludes the paper.

4.2 Review of Intraday Value at Risk

The Value at Risk (VaR) concept has emerged as one of the most prominent measures for downside market risk. VaR can be defined in terms of the conditional quantile of the asset portfolio return distribution for a given horizon and a given shortfall probability (typically chosen between 1% and 5%). Consider a time series of portfolio returns r_t and an associated series of *ex-ante* VaR forecasts with target probability p, denoted by $\mathrm{VaR}_t(p)$. The $\mathrm{VaR}_t(p)$ implied by a model M is defined by

$$\mathrm{Pr}^{\mathrm{M}}_{t-1}(r_t < -\mathrm{VaR}_t(p)) = p, \tag{4.2.1}$$

where $\mathrm{Pr}^{\mathrm{M}}_{t-1}$ denotes a probability derived from Model M using the information up to time $t-1$, and the negative sign in equation (4.2.1) is due to the convention of reporting VaR as a positive number.

Much effort has been spent on developing increasingly sophisticated risk models of VaR type for daily data and/or longer horizons. IVaR was initially discussed by Giot (2005), who applies VaR technique to intraday data. Giot (2005) aggregates irregularly-spaced high-frequency data to retrieve regularly-sampled intraday returns, which are in turn fitted into the normal GARCH, Student's t GARCH and RiskMetrics models in the framework of the VaR methodology. He also employs a duration (log-ACD) model for the irregularly-spaced high-frequency data. His study on the NYSE stocks shows that the Student's t GARCH model performs the best, while the high-frequency duration model behaves rather poorly in a regularly time-spaced framework. Similar to Giot (2005), Sakalauskas and Kriksciuniene (2006) suggest the modified RiskMetrics model of risk evaluation for short-term investments in the currency market, which is based on calculating VaR on hourly basis, using seasonal decomposition.

Giot and Grammig (2006) introduce the so-called liquidity adjusted IVaR by taking into account the potential price impact of liquidating an asset. In their model, liquidity risk is quantified by employing a new empirical technique which extends the classical frictionless IVaR methodology in an automated auction market. The liquidity adjusted IVaR measure is particularly relevant for impatient investors who submit market orders. Such models have been studied with recent contributions by Angelidis and Benos (2006), Qi and Ng (2010) and Groth and Muntermann (2011).

Coroneo and Veredas (2011) propose an IVaR model based on the conditional distribution of high-frequency financial returns by means of a two-component quantile regression model. In this model, the high-frequency returns are decomposed into two components, one to account for the intraday seasonality using a Fourier series and another for the return dynamics employing the lags of absolute returns. The application of this model on Spanish stocks outperforms GARCH-based IVaR methods.

The above-mentioned approaches are all based on regularly-spaced high-frequency data, which not only require finding the optimal aggregating scheme, but also inevitably lead to loss of important information contained in the time intervals between transactions as argued by Dionne, Duchesne and Pacurar (2009). These authors propose a method to compute the IVaR model based on irregularly-spaced high-frequency data and intraday Monte Carlo simulation. We call these kind of models irregularly-spaced IVaR models.

In this chapter, we will consider the Giot (2005) and Dionne, Duchesne and Pacurar (2009) methods for comparison.

4.2.1 Giot Method

One way to evaluate IVaR is to model the regularly-spaced intraday returns and their associated volatility. First, the feature of non-regularly time-spaced data requires a pre-determined sampling scheme. Define p_i as the sampled price, "raw return" can be computed as $r_i = \log(p_i) - \log(p_{i-1})$, where p_i is sampled at every s seconds so that $t_i - t_{i-1} = s$ and t_i is the calender time. Second, the GARCH-type volatility models can be used for volatility modeling provided that intraday seasonality is taken into account. Giot (2005) assumes deterministic seasonality in the intraday volatility, and the "deseasonalized return" R_i is computed from the raw returns r_i as

$$R_i = \frac{r_i}{\sqrt{\phi(t_i)}} \tag{4.2.2}$$

where $\phi(\cdot)$ is the deterministic intraday seasonal variance factor, defined as the expected variance conditional on time of day and usually computed by averaging the squared returns over s-second intervals for each day over the sample. Cubic splines are then used to smooth the time-of-day function. However, the choice for the time point in adjusting return R_i is quite arbitrary. For example, one can use the starting point of the interval t_{i-1} or the middle point of the interval $\frac{t_{i-1}+t_i}{2}$.

As intraday seasonality has been taken into account, the volatility models can be applied to the deseasonalized returns R_i. IVaR are then computed by re-including the seasonal component $\phi(t_i)$. In Giot (2005), the deseasonalized returns R_i are assumed to follow an AR(1)-GARCH(1,1) model, which can be written as

$$R_i = \mu + \delta R_{i-1} + e_i, \qquad e_i = \varepsilon_i \sqrt{h_i}, \tag{4.2.3}$$

with ε_i following an i.i.d. standard normal distribution and h_i given by

$$h_i = \omega + \alpha e_{i-1}^2 + \beta h_{i-1}. \tag{4.2.4}$$

Once all the parameters are determined, the one-step-ahead IVaR at time index $i+1$ is computed as

$$\mathrm{IVaR}_{i+1} = -\left(\hat{\mu}\sqrt{\phi(t_{i+1})} + \hat{\delta} R_i \sqrt{\phi(t_{i+1})} + z_p \sqrt{\hat{h}_{i+1}\phi(t_{i+1})}\right) \tag{4.2.5}$$

where z_p is the p-quantile of the standard normal distribution.[2]

4.2.2 Dionne-Duchesne-Pacurar Method

Dionne, Duchesne and Pacurar (2009) (DDP hereafter) adopt an intraday simulation method to evaluate IVaR, which considers the joint density of trade duration and tick-by-tick return, defined as the time and return between two consecutive transactions respectively. Let $f(x_i, r_i | x^{(i-1)}, r^{(i-1)}; \theta)$ represent the joint density of duration x_i and return r_i, and $x^{(i-1)}$ and $r^{(i-1)}$ denote the past observations of duration and return up to the $(i-1)$th transaction. The log-likelihood function for a sample of observations $\{x_i,\ r_i\}$, with $i = 1, \ldots, n$, can then be written as

$$\ell(\theta_1,\ \theta_2) = \sum_{i=1}^{n} \left[\log\{g(x_i | x^{(i-1)}, r^{(i-1)}; \theta_1)\} + \log\{q(r_i | x_i, x^{(i-1)}, r^{(i-1)}; \theta_2)\}\right] \tag{4.2.6}$$

where $g(x_i | x^{(i-1)}, r^{(i-1)}; \theta_1)$ is the marginal density of the duration x_i with parameter vector θ_1 conditional on past durations and returns, $q(r_i | x_i, x^{(i-1)}, r^{(i-1)}; \theta_2)$ is the conditional density of the return r_i with parameter vector θ_2 conditional on

[2] The negative sign in equation (5) is due to the convention of reporting IVaR as a positive number.

past durations and returns as well as the contemporaneous duration x_i. After the duration and tick-by-tick return models are determined, we can forecast the trade durations and tick-by-tick returns within a pre-determined time interval. From simulated replications, the empirical distribution of returns within the pre-determined interval is obtained.

In this chapter, we modify the DDP method by filtering microstructure noise employing volume duration, defined as the time until a given aggregate volume $\bar{v}$ is achieved.[3] Volume duration is introduced by Gourieroux, Jasiak, and Le Fol (1999) as a reasonable measure of liquidity that accounts simultaneously for the time and volume dimension of the trading process. There are few studies about volume durations and Paccurar (2008) provides an extensively survey.

Let $t_0, t_1, \cdots, t_N$ denote a sequence of time for which t_i is the time of occurrence of the ith volume event, which is said to have occurred if the cumulative trade volume since the last volume event is at least of a pre-set amount $\bar{v}$. Thus, $x_i = t_i - t_{i-1}$, for $i = 1, 2, \cdots, N$, are the intervals between consecutive volume events, called volume durations. As before, raw return $r_i = \log(p_i) - \log(p_{i-1})$, where p_i is the transaction price at time t_i.

The duration series $\{x_i\}_{i=1}^N$, generated by volume events in contrast to the trade durations in the original DDP method, is modeled by the ACD model. Let $\psi_i = \mathrm{E}(x_i|\Phi_{i-1})$ be the conditional duration, where Φ_i is the information set upon the volume event at time t_i. We assume the standardized duration

$$\varepsilon_i = \frac{x_i}{\psi_i} \tag{4.2.7}$$

to be a sequence of i.i.d. positive random variables with unit mean and finite variance. Following Bauwens, Giot, Grammig and Veredas (2004), we employ the log-ACD model with the flexible Weibull distribution to model the volume duration process, i.e.,

$$\log \psi_i = \omega + \alpha \log x_{i-1} + \beta \log \psi_{i-1}. \tag{4.2.8}$$

The standardized duration ε_i follows the Weibull distribution with density function

$$f(\varepsilon; \lambda, \phi) = \frac{\phi}{\lambda} \left(\frac{\varepsilon}{\lambda}\right)^{\phi-1} \exp\left[-\left(\frac{\varepsilon}{\lambda}\right)^{\phi}\right], \qquad \varepsilon > 0, \tag{4.2.9}$$

where we have imposed the restriction $\lambda = 1/\Gamma(1+1/\phi)$ to ensure unit mean.

The return series $\{r_i\}_{i=1}^N$, generated by the volume events, similar to the tick-by-tick returns in the original DDP method, can be modeled by the EGARCH method. Since both the volatility and duration exhibit intraday seasonality, the intraday periodicity of r_i has to be taken into account. Similar to $\phi(t_i)$ in Giot's method, we assume $\varphi(t_i)$ to be the deterministic intraday seasonal component of the intraday re-

[3] DDP method can also be modified by employing the frequency duration, which is, defined as the time until a given aggregate transaction amount is achieved.

turn variance. The difference between $\phi(t_i)$ and $\varphi(t_i)$ is that $\varphi(t_i)$ is the conditional variance of the irregularly-spaced intraday returns, while $\phi(t_i)$ is the conditional variance for the regularly-spaced intraday returns. Similarly, IVaR are computed by re-including the seasonal component $\varphi(t_i)$ for the original returns r_i. The deseasonalized return series $R_i = r_i/\sqrt{\varphi(t_i)}$ can be modeled by the EGARCH model, with

$$R_i = z_i\sqrt{h_i}, \tag{4.2.10}$$

where z_i denotes white nose, which is assumed to be i.i.d. random variables with standard normal distribution. The conditional variance h_i is given by

$$\log(h_i) = \gamma\log(x_i) + \tilde{\omega} + \tilde{\beta}(\log(h_{i-1}) - \gamma\log(x_{i-1})) + \xi|z_{i-1}| + \tilde{\alpha}z_{i-1}, \tag{4.2.11}$$

or in another form by simple transformation:

$$\log\left(\frac{h_i}{x_i^\gamma}\right) = \tilde{\omega} + \tilde{\beta}\log\left(\frac{h_{i-1}}{x_{i-1}^\gamma}\right) + \xi|z_{i-1}| + \tilde{\alpha}z_{i-1}. \tag{4.2.12}$$

This method is similar to Engle's (2000) UHF-GARCH model that specifies a GARCH component for the volatility of returns per unit time, which is h_i/x_i. Under Engle's framework, the conditional variance of return from one transaction to the next ($h_i = \mathrm{V}_{i-1}(r_i|x_i)$, where $\mathrm{V}_{i-1}(\cdot)$ denotes the conditional variance upon information at time t_{i-1}) equals the duration between the two consecutive trades (x_i) times the variance per second ($\sigma_i^2 = \mathrm{V}_{i-1}(r_i/\sqrt{x_i}|x_i)$). Although it seems natural to model the variance as a function of time when using irregularly-spaced transaction data, the modeling for the unit of time might be quite restrictive for some empirical data. The conditional heteroskedasticity in the returns could depend on time in a more complicated way due to the fact that the impact of the trading volume on volatility following the news events depends on the trading behavior of different type of investors, some are more sophisticated than others. Therefore, in the DDP method, the parameter γ specifies the duration weighting for the volatility of a particular stock, which has to be estimated for each stock. When $\gamma = 1$, the model is similar, though not equivalent, to the UHF-GARCH model; when $\gamma = 0$, it becomes the standard EGARCH model.

If the durations are weakly exogenous with respect to the processes for the returns, then the two parts of the log-likelihood function could be maximized separately, which simplifies the estimation (see, for instance, Engle (2000)). After the parameters are determined, we use simulation to forecast the price movements and the volume durations over a certain fixed time interval. In this chapter, the empirical return distribution of the forecast interval is estimated by 5,000 replications, and the $\mathrm{IVaR}_t(p)$ is then computed as the absolute value of the p-quantile of the empirical distribution.

The parameters estimated from the estimation part will be used to forecast the durations and returns within the forecast interval. For instance, to compute the IVaR between 10:10 - 10:45, the initial value of x_0, ψ_0, z_0 and h_0 can be collected from the information before 10:10. First, we collect the volume-event information between 9:30 - 10:10 and then compute the volume duration series $\{x_i^b\}_{i=1}^m$ and return duration series $\{r_i^b\}_{i=1}^m$ during 9:30 - 10:10 according to the pre-determined threshold $\bar{v}$.[4] Second, calculate the $\{\psi_i^b\}_{i=1}^m$ by equation (8), and calculate $\{z_i^b\}_{i=1}^m$ and $\{h_i^b\}_{i=1}^m$ by equation (11). Third, set the initial value of x_0, ψ_0, z_0 and h_0 to be x_m^b, ψ_m^b, z_m^b and h_m^b respectively. Finally, the simulation algorithm can be summarized as follows,

1. For $i = 1$, set the initial value for x_0 and obtain ψ_1 from equation (4.2.8).

2. Draw random number ε_i from the Weibull distribution.

3. Compute $x_i = \psi_i \, \varepsilon_i$, and ψ_{i+1} from equation (4.2.8).[5]

4. Conditional on the simulated value of x_i, h_i is computed using equation (4.2.11).

5. Draw random noise z_i from the standard normal distribution, and compute $R_i = z_i\sqrt{h_i}$. Calculate $t_i = t_{i-1} + x_i$ and $\log p_i = \log p_{i-1} + R_i\sqrt{\varphi(t_i)}$.

6. Set $i = i+1$ and iterate Steps 2, 3, 4 and 5 until t_i exceeds the pre-set forecast time interval.

4.3 AACD approach

Although the modified DDP method does not only take into account the irregularly-spaced information of transaction data but also takes account of noise filtering, the modeling of intraday returns and their associated volatility remain an important component in evaluating IVaR. While the problem arising from modeling the discreteness of the price process at the transaction level maybe solved to some extent by filtering microstructure noise, price changes are still modeled under the continuous-distribution framework. Models treating the price movement as a discrete variable and modeling the joint density of duration and price movement are needed to solve this problem. Bauwens and Giot (2003) propose an Asymmetric ACD model (AACD) in which the duration and quote revision are modeled *jointly*. Russell and Engle (2005) further propose an Autoregressive Conditional Multinomial - Autoregressive Conditional Duration (ACM-ACD) model to study the joint distribution of the duration and price change.

This section proposes a Monte Carlo simulation method to estimate IVaR, which is based on the AACD model. The AACD model has also been studied by Tay, Ting, Tse and Warachka (2011) to investigate the effect of trade volume, trade direction

[4] x_i^b and r_i^b denote the duration and return series before 10:10, in order to distinguish from the series in forecasting segment.

[5] x_1 should be larger than time between 10:10 and the last event time before 10:10.

and trade durations in explaining price dynamics and volatility. We use the AACD model to forecast price movements and price durations *jointly* for a pre-determined time interval and employ an intraday Monte Carlo simulation to obtain the empirical return distribution. Our method differ from the ACM-ACD model, which generates the trade duration and price change sequentially. The AACD model generates trade duration and price change synchronously, and is more convenient to estimate. $\mathrm{IVaR}_t(p)$ can then be computed by the p-th quantile of the empirical distribution. We first outline the AACD model, following by a description of the Monte Carlo simulation methodology.

4.3.1 The AACD model

Bauwens and Giot (2003) propose the AACD model, which extends the ACD model of Engle and Russell (1998) and allow the duration process to depend on the state of the price process. The Asymmetric ACD model is also called two-state ACD model; If the price has increased, the parameters of the ACD model can differ from what they are if the price has decreased. In other words, the AACD model allows the conditional expected duration to be depended not only on the previous duration, but also on the previous state of the price movement. Instead of *jointly* modeling trade durations and *tick-by-tick* price movements, in this chapter, we consider a two-state AACD model with possible price movements of a pre-determined threshold δ, either upwards or downwards, and price durations *jointly*.

Let $t_0, t_1, \cdots, t_N$ denote a sequence of times in which t_i is the time of the ith price event, to be defined below. Thus, $x_i = t_i - t_{i-1}$, for $i = 1, 2, \cdots, N$, are the intervals between consecutive price events, called price durations. A price event occurs if the cumulative change in the logarithmic transaction price since the last price event is at least of a preset amount δ, called the price threshold.[6] Thus, from time t_{i-1} to t_i, the price changes by an amount of at least δ, whether upwards or downwards. Let y_i denote the price movement direction of the ith price event, where y_i may take the values $j = -1,\ 1$ representing downward price movement and upward price movement, respectively.

Conditional on the ith price event, there are two possible end states: $y_i = -1$ or $y_i = 1$. Assume x_{ji}, $j = -1, 1$, to be two latent variables, which are unobservable, representing the durations with the two possible end states, respectively. Let $\psi_{ji} = \mathrm{E}(x_{ji} | \Phi_{i-1})$ be the conditional expected duration for latent variable x_{ji}, $j = -1, 1$, with Φ_{i-1} being the information set up to time t_{i-1}, which not only includes the previous duration x_{i-1} and lagged expected duration $\psi_{j,i-1}$ but also the price-movement direction y_{i-1}. Let

$$\varepsilon_{ji} = \frac{x_{ji}}{\psi_{ji}}, \qquad i = 1, 2, \cdots, N, \tag{4.3.1}$$

[6] δ is usually set to obtain an average duration of 5 min.

be the standardized price duration, for $j = -1,\ 1$. While there are two possible end states at the end of ith price event, there is only one realized state. As in the standard framework of a competing risk model, only the shortest duration from the two possible durations is observed (realized). Accordingly, x_i can be treated as the outcome variable of the function $x_i = \min(x_{1i},\ x_{-1,i})$. For instance, if an upward price movement is observed and the realized duration is x_i, then $x_{1i} = x_i$. Bauwens, Giot, Grammig and Veredas (2004) find that the Logarithmic ACD, if based on a flexible standardized duration distribution, provides a quite robust and useful framework for the modeling of price duration processes. Following their work, we assume that $\{\varepsilon_{1i}\}_{i=1}^{N}$ and $\{\varepsilon_{-1,i}\}_{i=1}^{N}$ are independently Weibull distributed with unit mean and finite variance. The density function of ε_j is given by

$$f(\varepsilon_j;\lambda_j,\phi_j) = \frac{\phi_j}{\lambda_j}\left(\frac{\varepsilon_j}{\lambda_j}\right)^{\phi_j-1} \exp\left[-\left(\frac{\varepsilon_j}{\lambda_j}\right)^{\phi_j}\right], \qquad \varepsilon_j > 0, \qquad j = -1, 1, \quad (4.3.2)$$

with $\lambda_j = 1/\Gamma(1+1/\phi_j)$ to ensure unit mean.

Our basic model is

$$\log\psi_{ji} = \sum_{k=-1,\ 1} (v_{jk} + \alpha_{jk}\log x_{i-1}) D_k(y_{i-1}) + \beta_j \log\psi_{j,i-1}, \qquad j = -1,\ 1, \quad (4.3.3)$$

where $D_k(z) = 1$, if $z = k$ and 0 otherwise. We let the expected conditional duration to depend on not only the previous durations, but also the previous state of price movement. For the upward price movement process, the expected conditional duration ψ_{1i} of the latent variable x_{1i} depends on the previous realized duration x_{i-1} and the previous expected conditional duration $\psi_{1,i-1}$ as well as the previous price-movement state y_{i-1}. Thus,

$$\log\psi_{1i} = \begin{cases} v_{1,1} + \alpha_{1,1}\log x_{1,i-1} + \beta_1 \log\psi_{1,i-1}, & \text{if } y_{i-1} = 1, \\ v_{1,-1} + \alpha_{1,-1}\log x_{-1,i-1} + \beta_1 \log\psi_{1,i-1}, & \text{if } y_{i-1} = -1. \end{cases} \quad (4.3.4)$$

Similarly, for the downward price movement process, $\psi_{-1,i}$ can be modeled as

$$\log\psi_{-1,i} = \begin{cases} v_{-1,1} + \alpha_{-1,1}\log x_{1,i-1} + \beta_{-1} \log\psi_{1,i-1}, & \text{if } y_{i-1} = 1, \\ v_{-1,-1} + \alpha_{-1,-1}\log x_{-1,i-1} + \beta_{-1} \log\psi_{1,i-1}, & \text{if } y_{i-1} = -1. \end{cases} \quad (4.3.5)$$

Furthermore,

$$x_{i-1} = \begin{cases} x_{1,i-1}, & \text{if } y_{i-1} = 1, \\ x_{-1,i-1}, & \text{if } y_{i-1} = -1. \end{cases} \tag{4.3.6}$$

Combining the two-state competing risk model with the Log-ACD model yields the asymmetric Log-ACD model.

As in the standard framework of a competing risk model, the joint conditional bivariate probability function - probability density function (pf-pdf) for x_i and y_i are given by

$$f(x_i, y_i|\Phi_{i-1}) = \prod_{j=-1,\ 1} h_{x_{ji}}(x_i|\Phi_{i-1})^{D_j(y_i)} S_{x_{ji}}(x_i|\Phi_{i-1}), \tag{4.3.7}$$

where $h_{x_{ji}}$ and $S_{x_{ji}}$ denote the conditional hazard function and conditional survival function of x_{ji}, with the form:

$$h_{x_{ji}}(x_i|\Phi_{i-1}) = -\frac{\phi_j}{\psi_{ji}\lambda_j}\left(\frac{x_i}{\psi_{ji}\lambda_j}\right)^{\phi_j-1}, \tag{4.3.8}$$

$$S_{x_{ji}}(x_i|\Phi_{i-1}) = \exp\left\{-\left(\frac{x_i}{\psi_{ji}\lambda_j}\right)^{\phi_j}\right\}. \tag{4.3.9}$$

The duration that is realized (observed) contributes to the joint conditional pf-pdf given by equation (4.3.9) via the conditional density function, whereas the unrealized duration contributes to it via the conditional survival function. For example, if a upward price movement state is observed at t_i, the conditional bivariate pf-pdf of the pair $\{x_i, y_i = 1\}$ is given by:

$$\begin{aligned} f(x_i, y_i = 1|\Phi_{i-1}) =& h_{x_{1i}}(x_i|\Phi_{i-1}) S_{x_{1i}}(x_i|\Phi_{i-1}) S_{x_{-1,i}}(x_i|\Phi_{i-1}) \\ =& f_{x_{1i}}(x_i|\Phi_{i-1}) S_{x_{-1,i}}(x_i|\Phi_{i-1}). \end{aligned} \tag{4.3.10}$$

Therefore, if the duration x_i ends up with an upward price movement ($y_i = 1$), x_i contributes to the pf-pdf via: (1) the conditional density of x_{1i} evaluated at x_i, i.e. $f_{x_{1i}}(x_i|\Phi_{i-1})$ and (2) the conditional probability that the duration $x_{-1,i}$ ended up with downward price movement is longer than the realized duration x_i, i.e. $S_{x_{-1,i}}(x_i|\Phi_{i-1})$.

Assuming the Weibull distribution for ε_{ji}, for $j = -1,\ 1$, the log-likelihood function can be derived as:

$$\log L(\Theta) = -\sum_{i=1}^{N}\left(\sum_{j=-1,\ 1}\left(\frac{x_i}{\psi_{ji}\lambda_j}\right)^{\phi_j} - \log\left(\sum_{j=-1,\ 1} D_j(y_i)\frac{\phi_j}{\psi_{ji}\lambda_j}\left(\frac{x_i}{\psi_{ji}\lambda_j}\right)^{\phi_j-1}\right)\right), \tag{4.3.11}$$

and the parameter vector Θ can be estimated by maximizing the log-likelihood.

4.3.2 Intraday Monte Carlo simulation

We now describe our Monte Carlo simulation procedure used for computing the IVaR based on high-frequency data. The estimated AACD model will be used to simulate the price events and event times for the pre-set intervals. For illustration, suppose we want to compute the IVaR between 10:10 - 10:45, the initial value of x_0, $\psi_{-1,0}$, ψ_{10}, and y_0 are calculated using the information before 10:10. First, we collect the price-event information between 9:30 - 10:10 and then compute the price duration series $\{x_i^b\}_{i=1}^m$ and price-movement direction series $\{y_i^b\}_{i=1}^m$ during 9:30 - 10:10 according to the pre-determined threshold δ.[7] Suppose the last event occurs at 10:08:32. Second, calculate the $\{\psi_{ji}^b\}_{i=1}^m$ by equation (4.3.3). Third, set the initial values of x_0, $\psi_{-1,0}$, ψ_{10}, and y_0 to be x_m^b, $\psi_{-1,m}^b$, ψ_{1m}^b and y_m^b respectively. We also set t_0 to be 10:08:32 and p_0 to be the price occurs at 10:08:32. Finally, the simulation algorithm is summarized as follows:

1. For $i = 1$, set the initial value as describe above and obtain ψ_{j1}, for $j = -1, 1$.
2. For $j = -1,\ 1$, draw two values of ε_{ji} from independent Weibull distributions with shape parameters ϕ_1 and ϕ_{-1} respectively.
3. Compute $x_{ji} = \psi_{ji}\ \varepsilon_{ji}$ and $\psi_{j,i+1}$ using equation (15), for $j = -1, 1$.
4. Set $y_i = j$ and $x_i = \min\{x_{-1,i}, x_{1i}\}$.
5. Compute $t_i = t_{i-1} + x_i$, and $\log p_i = \begin{cases} \log p_{i-1} + \delta, & \text{if } y_i = 1, \\ \log p_{i-1} - \delta, & \text{if } y_i = -1. \end{cases}$
6. Set $i = i + 1$ and iterate Steps 2, 3, 4 and 5 until t_i exceeds the pre-set time intervals to forecast.

Note that the first simulated duration should be larger than 88 seconds (the difference between time 10:10:00 and 10:08:32), otherwise, another price event has to be generated as the starting observation. Let $t_1, \cdots, t_n$ denote a sequence of times in which t_i is the time of the ith forecasted event time as illustrated in Figure 4.1. The return δ from time t_0 to t_1 is split into two parts. The part contributing to the interval 10:10 - 10:45, denoted as δ_1, is linearly approximated as proportional to the fraction within 10:10 - 10:45. The same procedure applies to the last simulated price event, the return contributing to the forecasted interval is denoted as δ_n. The simulated returns are computed as $r = \delta_1 + \log p_{n-1} - \log p_1 + \delta_n$. Then we repeat the whole procedure 5000 times to obtain an empirical distribution of the returns,

[7] x_i^b and y_i^b denote the duration and return series before 10:10, in order to distinguish them from the data in the estimation period.

and the absolute value of the p-quantile will be computed as IVaR(p). Note that, if some news hits the market before 10:10 and triggers a price event after 10:08:32, this will have impact on the initial value for the simulation. Thus, a shorter simulated duration x_1 may occur and may result in a larger IVaR. In practice, the starting time and ending time of the IVaR interval could be any time of the day. Since our method takes account of news happening before the interval of the IVaR, we name it "near" real-time IVaR.

4.4 IVaR backtesting

To evaluate the IVaR forecast capabilities of our model, we will perform an out-of-sample backtesting analysis. Christoffersen (1998) points out that the problem of determining the accuracy of a VaR model can be reduced to the problem of determining whether the hit sequence, $[\mathrm{I}_t(\alpha)]_{t=1}^{t=T}$, where

$$\mathrm{I}_t(\alpha) = \begin{cases} 1, & \text{if} \quad r_t < -\mathrm{IVaR}_t(\alpha) \\ 0, & \text{otherwise} \end{cases}, \tag{4.4.1}$$

satisfies two properties: unconditional coverage property and independence property,[8]

1. Unconditional coverage property - The probability of realizing a loss in excess of the reported VaR, VaR_α, is $100\alpha\%$, i.e. $\Pr(\mathrm{I}_t(\alpha) = 1) = \alpha$. If violation occurrence (the failure rate) deviate from $100\alpha\%$ significantly, the model should be considered problematic;

2. Independence property - while the unconditional property places a restriction on how often VaR violations may occur, the independence property places restrictions on the ways in which these violations may occur. VaR violations observed at two different dates for the same coverage rate must be distributed independently.

4.4.1 Kupiec test

Kupiec (1995) focuses exclusively on the property of unconditional coverage, a proportion of failures (POF) test is proposed, which examines how many times a financial institution's VaR is violated over a given span of time at VaR level of α . If the IVaR estimates are accurate, the failure rate $\hat{\alpha}$ should be approximately equal

[8] Conditional coverage property is satisfied if both unconditional coverage property and independence property are fulfilled.

to α. Under the null hypothesis that the model is correct, the Kupiec test statistic takes the form

$$\mathrm{LR_{POF}} = 2[\log((1-\hat{\alpha})^{T-\mathrm{I}(\alpha)}\hat{\alpha}^{\mathrm{I}(\alpha)}) - \log((1-\alpha)^{T-\mathrm{I}(\alpha)}\alpha^{\mathrm{I}(\alpha)})], \tag{4.4.2}$$

where

$$\hat{\alpha} = \frac{1}{T}\mathrm{I}(\alpha) = \frac{1}{T}\sum_{t=1}^{T}\mathrm{I}_t(\alpha), \tag{4.4.3}$$

and T is the sample size. Under the null hypothesis, $\mathrm{LR_{POF}}$ is asymptotically χ^2 distributed with one degree of freedom. If the value of the $\mathrm{LR_{POF}}$ statistic exceeds the critical value of the χ_1^2 distribution, the null hypothesis is rejected and the model is deemed inappropriate.

4.4.2 Dynamic quantile test

Engle and Manganelli (2004) suggest using a linear regression model that links current violation to past violations. If the IVaR forecast is correct, the violations should not be serially correlated. Let $\mathrm{Hit}_t(\alpha) = \mathrm{I}_t(\alpha) - \alpha$ be the de-meaned process on α associated with $\mathrm{I}_t(\alpha)$,

$$\mathrm{Hit}_t(\alpha) = \begin{cases} 1-\alpha, & \text{if} \quad r_t < -\mathrm{IVaR}_t(\alpha), \\ -\alpha, & \text{otherwise.} \end{cases} \tag{4.4.4}$$

Consider the following linear regression model:

$$\mathrm{Hit}_t(\alpha) = \omega + \sum_{k=1}^{K}\beta_k\mathrm{Hit}_{t-k}(\alpha) + \delta\mathrm{VaR}_t(\alpha) + \varepsilon_t, \tag{4.4.5}$$

where ε_t is an i.i.d. process. The null hypothesis of conditional coverage corresponds to the joint nullity of the coefficients β_k and δ as well as the constant ω, i.e.,

$$\mathrm{H}_0 : \omega = \delta = \beta_1 = \cdots = \beta_K = 0, \tag{4.4.6}$$

where $\beta_1 = \cdots = \beta_K = \delta = 0$ reflects the independence hypothesis, $\omega = 0$ reflects the unconditional coverage hypothesis. Indeed, under the null hypothesis $\mathrm{E}[\mathrm{Hit}_t(\alpha)] = \mathrm{E}(\varepsilon_t) = 0$, which implies that $\Pr[\mathrm{I}_t(\alpha) = 1] = \mathrm{E}[\mathrm{I}_t(\alpha)] = \alpha$. The joint nullity test of all coefficients, including the constant, therefore corresponds to a conditional coverage test. The LR statistic or the Wald statistic can then be used to test the simultaneous nullity of these coefficients. If we let $\Psi = (\omega, \delta, \beta_1, \ldots, \beta_K)'$ be the vector of the $K+2$ parameters in this model and let Z be the matrix of explanatory variables of equation (4.4.6), then the Wald statistic $\mathrm{DQ}_{\mathrm{CC}}$ associated with a test of conditional coverage is

$$\mathrm{DQ}_{\mathrm{CC}} = \frac{\hat{\Psi}' Z' Z \hat{\Psi}}{\alpha(1-\alpha)} \xrightarrow{D} \chi^2(K+2) \tag{4.4.7}$$

Following Engle and Manganelli (2004)'s framework, we use 5 lags ($K = 5$) in this chapter.

4.4.3 GMM duration-based test

Under the null that IVaR forecasts are correctly specified, the violations should occur at random time intervals. Suppose the duration between two violations is defined as

$$d_i = t_i - t_{i-1}, \tag{4.4.8}$$

where t_i denotes the violation number i. The duration between violations of the IVaR should be completely unpredictable. Under the conditional convergence hypothesis, the duration variable d_i follows a geometric distribution with parameter α and a probability mass function given by

$$f(d;\alpha) = \alpha(1-\alpha)^{d-1}, \qquad d \in \mathbb{N}. \tag{4.4.9}$$

Christoffersen and Pelletier (2004) and Haas (2005) independently propose backtesting statistics employing the properties of the geometric distribution, and call them duration-based backtests. However, some limitations of their test resulted in the lack of popularity among practitioners. For example, these tests exhibit low power for realistic backtesting sample size. Recently, Bertrand, Gilbert, Christophe and Sessi (2011) propose an GMM duration-based backtest which tackles these issues.

The expectation of some particular orthonormal polynomials associated with the geometric distribution is equal to 0, and these polynomials are used as special moment conditions to test for the geometric distribution. The orthonormal polynomials $\mathbf{M}_{j+1}(d,\alpha)$ associated to the geometric distribution with success probability α can be defined by the following recursive relationship:

$$\mathbf{M}_{j+1}(d,\alpha) = \frac{(1-\alpha)(2j+1)+\alpha(j-d+1)}{(j+1)\sqrt{1-\alpha}}\mathbf{M}_j(d,\alpha) - \left(\frac{j}{j+1}\right)\mathbf{M}_{j-1}(d,\alpha), \quad \forall d \in \mathbb{N} \tag{4.4.10}$$

for any order $j \in \mathbb{N}$, with $\mathbf{M}_{-1}(d,\alpha) = 0$ and $\mathbf{M}_0(d,\alpha) = 1$.[9]

If the true distribution of d is a geometric distribution with success probability

[9] Please see details in Bertrand, Gilbert, Christophe and Sessi (2011).

α, it follows that

$$\mathrm{E}[\mathbf{M}_j(d;\alpha)] = 0 \quad j \in \mathbb{N}. \tag{4.4.11}$$

The GMM duration-based backtest procedure exploits these moment conditions. More precisely, we define $\{d_1, \ldots, d_N\}$ as a sequence of N durations between IVaR violations, computed from the sequence of the hit variables $\{\mathrm{I}_t(\alpha)\}$. Under the conditional coverage assumption, the distributions of $d_i, i = 1, ..., N$, are i.i.d. and have a geometric distribution with a success probability equaling to the coverage rate α. Hence, the null of conditional convergence can be expressed as follows:

$$\mathrm{H}_{0,\mathrm{CC}} : \mathbf{M}_j(d;\alpha)] = 0, \quad j = \{1, \ldots, m\}, \tag{4.4.12}$$

where m denotes the number of moment conditions.

We denote $\mathrm{J}_{\mathrm{CC}}(m)$ as the conditional convergence test statistic associated with the first m orthonormal polynomials. Assume that the duration process is stationary and ergodic. Under the null hypothesis of conditional coverage, we have

$$\mathrm{J}_{\mathrm{CC}}(m) = \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{M}(d_i;\alpha) \right)^T \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{M}(d_i;\alpha) \right) \xrightarrow{D} \chi^2(m). \tag{4.4.13}$$

where $\mathbf{M}(d_i;\alpha)$ denotes a $(m,1)$ vector whose components are the orthonormal polynomials $\mathbf{M}_j(d_i;\alpha)$, for $j = 1, \ldots, m$. In this chapter, we consider 5 moment conditions $(m = 5)$.

Kupiec's test is probably the first and the most popular test to evaluate VaR performance, it only focuses on the unconditional coverage property and leaves the independence property on the air. DQ test is also widely used as an evaluation tool for VaR models, since it is easy to implement. The GMM duration-based test is the most recent test, and it tackles the sample size problem existed for the duration-based backtests. In this chapter, we use the Kupiec test, DQ test and GMM duration-based test to evaluate the IVaR performance.

4.5 Data

The data used in this chapter were extracted and compiled from the TAQ database provided through the Wharton Research Data Services. We downloaded the following variables from the Consolidated Trade (CT) file: date, time, price, and trade size. We obtained data for all stocks that were components of the S&P 500 index over 3 different periods and were traded on the New York Stock Exchange (NYSE). Period 1 covers the 2008 global financial crisis, from 2008/09/01 to 2008/12/31, which is a period of bearish market with prices generally falling. Period 2 is from 2010/01/01 to 2010/04/30, which is a post-financial crisis period. Period 3 is a more recent period, from 2012/01/01 to 2012/04/30. We excluded the days with trading time less than 2/3 of the total trading time (9:30-16:00) for each stock. For each period, the stocks with stock splits are excluded, and stocks with less than 80 trading days are also excluded. Table 4.2 summarizes some key statistics for our stock sample. In Period 3, there are 497 stocks which stayed as a member of the S&P 500 index, out of which 107 stocks were either not traded on the NYSE or had less than 80 trading days. There were two stocks with stock splits during this period. After excluding the stocks that do not meet our selection criteria, 388 stocks remain for investigation in Period 3. We select these 3 periods (during and after the 2008 global financial crisis) to evaluate the IVaR models under various different market environments. Furthermore, we short-list 10 stocks with large capitalization and trade intensities for reporting convenience. These are Exxon Mobil Corporation (XOM), General Electric (GE), Procter & Gamble Co. (PG), Merck (MRK), Johnson & Johnson (JNJ), AT & T (T), Chevron (CVX), JP Morgan (JPM), Wal Mart (WMT), IBM (IBM), and Pfizer (PFE). We also obtained data for these 10 stocks from 2008/01/01-2010/12/31 for further study, which is detailed later.

With the growing development of high-frequency trading, the probability of multiple trades with the same time stamp is increasing. As illustrated in the introduction, transactions can be done within a second and there are multiple trades with the same time stamp, unless we can access Millisecond Trade and Quote (MTAQ) database. We combine the trades with the same time stamp using the size weighted method. We consider three ways of filtering noise. First, we sample the tick data at large regularly-spaced (say, 30-min) intervals to reduce the effect of microstructure noise. The sampled data are then fitted to Giot's method after deseasonalizing the return series. Second, tick data are thinned using volume duration. The thinned duration and return series are then modeled using the DDP method. Third, tick data are thinned by price duration. The price change direction and duration series are then *jointly* modeled by the AACD model.

The sample size of different stocks in different periods is different, although there are at least 80 trading days for each stocks. We use 21 days as the estimation period and employ the estimated models to forecast the next day's price movements and hence IVaR. We then move the estimation period by including one day forward and excluding the first day to keep the estimation length to be 21 days. We repeat this procedure for the whole sample. Thus, there are at least 59 $(80 - 21)$ days and 59×13 30-min intervals, and at most 62 $(83 - 21)$ days and 62×13 30-min intervals in Period 1.

4.5.1 Duration Seasonality Adjustment

To take into account the time-of-day effect, Engle and Russell (1998) suggest computing diurnally-adjusted durations by dividing the raw durations by a seasonal deterministic factor. This factor can be jointly estimated using maximum likelihood method or by regressing the calendar-time durations on the time-of-day variables using a cubic spline function. However, as criticized in Wu (2012) and Tse and Dong (2012), the duration adjusted method is dependent on the specific smoothing method, and one also need to decide which time point best represents the duration. Wu (2012) and Tse and Dong (2012) propose a method for diurnally adjusting the intraday periodicity through time transformation.[10]

The theoretical underpinning of diurnal adjustment through time transformation is that the unconditional distribution of the duration process should be evenly distributed throughout the trading day under the assumption of no intraday periodicity. Therefore, for the diurnally-adjusted price durations, the price events should occur evenly across the trading day; and similarly for the diurnally-adjusted volume durations, the trading volume should be distributed evenly across the trading day. Let the strictly increasing time points $t_0,\ t_1,\cdots,t_{23400}$ separately denote 9:30:00, 9:30:01, $\cdots$, 16:00:00. Let $n_1,\cdots,n_{23400}$ $(n_0 = 0)$ and $v_1,\cdots,v_{23400}$ $(v_0 = 0)$ denote the total number of price events and total volume traded at time 9:30:00, 9:30:01, $\cdots$, 16:00:00, respectively, over all trading days in the sample. We compute

$$N_{t_k} = \sum_{i=0}^{k} n_i, \qquad k = 0,1,\cdots,23400, \quad \text{and} \ N_T = \sum_{i=0}^{23400} n_i,$$

$$V_{t_k} = \sum_{i=0}^{k} v_i, \qquad k = 0,1,\cdots,23400, \quad \text{and} \ V_T = \sum_{i=0}^{23400} v_i.$$

The two functions are further smoothed by linear interpolation in the neighborhood of t_k if $n_k = 0$ or $v_k = 0$. The time-transformation function $\tilde{Q}(t_k)$ and $\hat{Q}(t_k)$ are then computed as

$$\tilde{Q}(t_k) = \frac{N_{t_k}}{N_T}, \qquad k = 0,1,\cdots,23400,$$

$$\hat{Q}(t_k) = \frac{V_{t_k}}{V_T}, \qquad k = 0,1,\cdots,23400.$$

The diurnally transformed time $\tilde{t}_k$ and $\hat{t}_k$ in accordance with $\tilde{Q}(t_k)$ and $\hat{Q}(t_k)$ are then given by

$$\tilde{t}_k = 23400\tilde{Q}(t_k) = 23400\left[\frac{N_{t_k}}{N_T}\right], \qquad k = 0,1,\cdots,23400.$$

$$\hat{t}_k = 23400\hat{Q}(t_k) = 23400\left[\frac{V_{t_k}}{V_T}\right], \qquad k = 0,1,\cdots,23400.$$

The diurnally transformed duration of any two calendar-time points t_i and t_j is calculated as

$$\tilde{t}_i - \tilde{t}_j = 23400[\tilde{Q}(t_i) - \tilde{Q}(t_j)], \tag{4.5.1}$$

[10] Named time change method in Wu (2012).

$$\hat{t}_i - \hat{t}_j = 23400[\hat{Q}(t_i) - \hat{Q}(t_j)]. \tag{4.5.2}$$

The two time transformations with respect to $\tilde{Q}(\cdot)$ and $\hat{Q}(\cdot)$ are named TTM1 and TTM2 for short, respectively. In this chapter, we employ TTM1 to compute the diurnal adjustment of price durations for the AACD model. Also, we adopt the TTM2 method for the adjustment of volume duration for the DDP method. The most important advantage of the time transformation method is that the switch between calendar time and diurnally-adjusted time can be easily performed. Given any two calendar-time points $t_i < t_j$, the diurnally-adjusted duration between these two time points can be computed by equation (4.5.1) and equation (4.5.2). Likewise, given any two diurnally-adjusted time points $\tilde{t}_i < \tilde{t}_j$ (or $\hat{t}_i < \hat{t}_j$), the corresponding duration in calendar time is

$$\tilde{Q}^{-1}\left(\frac{\tilde{t}_j}{23400}\right) - \tilde{Q}^{-1}\left(\frac{\tilde{t}_i}{23400}\right) \tag{4.5.3}$$

where $\tilde{Q}^{-1}$ is the inverse function of $\tilde{Q}$. This facilitates the simulation of the AACD approach and DDP method. For example, the simulated durations from the AACD approach are all diurnally-adjusted. However, the IVaR (or return distribution) must be specified in calendar time. Time transformation methods are convenient to use, as the calendar time and diurnally-adjusted time can be easily converted from one another.

Figure 4.2 shows the average price durations and Figure 4.3 shows the volume durations for the 10 selected stocks during period 2008/01/01 - 2010/12/31, respectively. We set the price threshold δ and volume threshold $\bar{v}$ so as to obtain an average duration of about 5 min. Average durations over 30-minute regularly-spaced intervals are calculated and then smoothed using cubic splines to obtain the seasonal factor. As we can see, there exist significant intraday periodicity patterns for both raw price durations and raw volume durations. However, after we adjust the price and volume durations using the TTM1 and TTM2 methods, there is no clear intraday periodicity for the price and volume durations.

4.5.2 Volatility Seasonality Adjustment

When it comes to intraday returns and the associated volatility modeling, intraday volatility seasonality is a crucial factor that must be taken into account. For regularly-spaced returns, we assume a deterministic seasonality factor $\phi(t_i)$ to deseasonalize the intraday returns, like in equation (4.2.2). In this chapter, we consider two methods to calculate $\phi(t_i)$. Firstly, we calculate 30-min squared returns, and move the 30-min window 5 min forward until the last 30-min interval in each trading day. The deterministic seasonality factor $\phi(t_i)$ is computed by averaging the 30-min squared returns and then smoothed employing cubic splines. Secondly, we compute the intraday variance by the integrating conditional variance (ICV) method proposed by Tse and Yang (2012). The high-frequency volatility (over a day or shorter intervals) is captured by integrating the instantaneous conditional return variance per unit time obtained from the ACD models, called the ACD-ICV method. The ACD-ICV estimate has clear advantage in capturing volatility over short intervals such as 15 min or 30 min. We estimate the intraday variance over 30-min intervals by the ACD-ICV method. The intraday seasonality factor is then calculated as the average of 30-min intraday variance and then smoothed using cubic splines. The deterministic seasonality computed using squared returns and ACD-ICV method are named $\phi_1(t_i)$ and $\phi_2(t_i)$, respectively. Giot's method with $\phi_1(t_i)$ and $\phi_2(t_i)$ employee are subsequently named G1 and G2 method, respectively. Figure 4 shows the deterministic seasonality factor $\phi_1(t_i)$ and $\phi_2(t_i)$ for the 10 selected stocks. As we can see, both $\phi_1(t_i)$ and $\phi_2(t_i)$ exhibit a U-shape and $\phi_2(t_i)$ is much smoother than $\phi_1(t_i)$. The patterns are analogous to what has been found in previous studies.

We also consider two methods of estimating the deterministic seasonality factor of intraday volatility $\varphi(t_i)$ in the DDP model. Firstly, similar to Dionne, Duchesne and Pacurar (2009) but use 30-min squared returns, and then move the 30-min window 5 min forward until the last 30-min in each trading day. We denote this seasonality factor by $\varphi_1(t_i)$, which is the same as $\phi_1(t_i)$ in Giot's method. Second, since the returns in the DDP method are sampled using volume duration, another way is to model the variance per duration. Figure 4.5 shows the average variance and average number of volume events for the 13 30-min regularly-spaced intervals over all trading days for period period 2008/01/01 - 2010/12/31. We can see that the intraday variance is at its highest level after the market opens, and then gradually decreases to its lowest level around 12:30-13:00. Subsequently it gradually increases until the market closes. In contrast, the average number of volume events exhibit a different pattern. Specifically, it is higher after the market opens, and gradually decreases to its lowest level around 12:30-13:00. As investors close their positions before market closes, the average number of volume events reaches its highest level at the market close. As h_i in equation (4.2.11) is the variance of the returns per volume duration, we assume the deterministic seasonality factor $\varphi_2(t_i)$ to be the variance per duration. It is measured by the intraday variance estimated using ACD-ICV over 30-min interval divided by the number of volume events over that interval and then smoothed using cubic splines. Figure 4.6 presents the deterministic seasonality $\varphi_2(t_i)$ for the 10-listed stocks in period 01/01/2008 - 31/12/2010. We can see that the variance per duration is high during the first several hours after the market opens and is low during the time towards the market close. We name the DDP method employing $\varphi_1(t_i)$ and $\varphi_2(t_i)$ by DDP1 and DDP2, respectively.

In this chapter, we consider the AACD, DDP1, DDP2, G1 and G2 methods for the estimation of IVaR.

4.6 Empirical Results

We set the price threshold δ and volume threshold $\bar{v}$ to obtain an average duration of about 5 min in the estimation sample. For the G1 and G2 methods, the intraday returns are sampled at 30-min intervals in order to obtain 30-min IVaR estimates.

Table 4.3 presents the non-overlapping consecutive 30-min IVaR backtesting results of the AACD approach for the 3 different periods. The left-hand panel shows the p-values of the Kupiec test for 10 stocks with IVaR at the 5%, 2.5% and 1% levels, the middle panel shows the p-values of the DQ test of Engle and Manganelli (2004) and the right-hand panel presents the p-values of the GMM duration-based test of Bertrand, Gilbert, Christophe and Sessi (2011). Bolded entry denotes a failure of the IVaR model at the 95% confidence level, since the p-values are less than 0.05. Tables 4.4 through 4.7 present the backtesting results of the DDP1, DDP2, G1 and G2 methods, respectively. We select these 10 large-cap stocks for reporting convenience. The backtesting results of other stocks are summarized in Table 4.8.

Table 4.8 summaries the 30-min IVaR backtesting results of all 5 models for all selected stocks traded on the NYSE. The figure in the table is the percentage of stocks with IVaR backtesting p-values larger than 0.05 under each backtesting model. For the Kupiec test at IVaR level 5%, there are 93.67% out of 379 stocks with p-value larger than 0.05 in Period 1. However, the ratios are 63.59%, 84.96%, 52.77% and 66.23% for DDP1, DDP2, G1 and G2, respectively. Bolded entries present the highest percentages and italic entries present the second highest percentages. We can see that IVaR estimated by the AACD approach performs the best, while the DDP2 method is the second best. The DDP2 method performs better than DDP1 method under all the cases, which indicates that $\varphi_2(t_i)$ works better than $\varphi_1(t_i)$ as the deterministic intraday seasonality factor. The G2 method performs better than the G1 method in all the cases, which maybe due to the better performance of the ACD-ICV method as an estimate of intraday variance. We also see that the figures decrease as the IVaR levels decrease for both the G1 and G2 methods, which is due to the continuous-distribution assumption of the regularly-spaced intraday returns and the modeling of its associated volatility. Introducing irregularly-spaced information to the modeling of intraday return and its associated volatility, as in the DDP method, has alleviated the problem to a certain extent. However, the AACD approach, without doubt, has made significant improvements over the DDP methods.

To examine the intraday pattern of IVaR, we compute the mean of IVaR over each of the 13 intervals from 9:30 to 15:30 for the 10 selected stocks over the three sample periods. Figure 4.7 presents the intraday average IVaR at 5% level. To avoid jamming the figures, only the estimates of the AACD, DDP2 and G2 methods are presented. In particular, there is an "IVaR smile", with IVaR being the lowest in the 11:00-14:30 interval for most stocks. The three measures trace each other very well, however, with little difference between the first several and last several 30-min intervals. IVaR at the 2.5% and 1% levels share similar intraday patterns with some quantitative difference. Also, the lower IVaR level is associated with the higher IVaR value.

We further examine the percentages of the different IVaR methods in different 30-min intervals of the day. To this effect, we focus on the 10 selected stocks over the period 2008/01/01 to 2010/12/31. The results for the first two and last two 30-min intervals of each day are shown in Table 4.9. The entries present the number of stocks (out of 10) with IVaR backtesting p-values larger than 0.05. It can be seen that the numbers are the largest for the AACD approach. Other 30-min intervals are also computed and the backtesting results are not presented here. IVaR by the AACD approach performs well against the other two methods.

Indeed, IVaR can be computed for any time horizon once the AACD and DDP models have been estimated without requiring new sampling and estimation when the time horizon changes, due to the flexibility of irregularly-spaced information. Table 4.10 summaries the 60-min IVaR backtesting results of the AACD, DDP1 and DDP2 models for all selected stocks traded on the NYSE, the AACD approach perform the best among all the models.[11]

To further examine the effect of δ on the estimation of IVaR through AACD model, we perform a robustness check by varying the target average duration. Our robustness check shows that the AACD approach is not sensitive to the choice of the price range δ, provided that the price events sampled are not too infrequent.

Overall, our results show that IVaR estimated through the AACD approach is the most accurate among all the methods considered.

4.7 Conclusion

In this chapter, we propose a new method of computing the IVaR using high-frequency transaction data. Intraday directional price movements and price durations are *jointly* modeled by employing the AACD model. We adopt an intraday Monte Carlo simulation approach to estimate IVaR, which enables us to forecast high-frequency returns for any arbitrary interval. In our setup, price durations yield the consecutive steps in time while the price movements allow us to simulate the corresponding returns. Regularly-spaced intraday returns are simply the sum of the price movements simulated conditional on the given horizon. We also modify the DDP method of Dionne, Duchesne and Pacurar (2009) by filtering noise employing volume durations.

Using high-frequency data of all the S&P 500 component stocks traded on the NYSE over three different periods, our results show that the IVaR estimates computed using the AACD approach track closely to the DDP and Giot methods. IVaR backtesting results show that the AACD approach performs the best over other methods, based on the results on backtesting p-values. Our robustness check shows

[11] The first 15 min and last 15 min of each day are excluded, the remaining 6 hours are split into 6 60-min intervals each day.

that the AACD approach is not sensitive to the choice of the price range δ, provided that the price events sampled are not too infrequent.

Table 4.1: Summary statistics for tick price movements and trade durations.

Code	XOM	GE	PG	JNJ	T	CVX	JPM	WMT	IBM	PFE
No. of days	757	757	757	757	757	757	757	757	757	757
Frequency (%) of price movements										
5 ticks up or more	1.05	0.08	0.55	0.46	0.12	1.75	0.46	0.38	2.99	0.03
4 ticks up	0.77	0.06	0.45	0.36	0.13	1.19	0.54	0.35	1.89	0.02
3 ticks up	1.59	0.18	1.08	0.89	0.39	2.44	1.17	0.87	3.58	0.08
2 ticks up	3.94	0.87	3.55	2.83	1.60	5.64	3.12	2.73	6.81	0.54
1 ticks up	14.27	11.29	16.92	14.41	12.74	16.04	13.96	14.21	13.58	11.45
0 tick, no price change	56.81	75.01	54.97	62.07	70.04	45.78	61.52	62.98	42.23	75.72
1 tick down	14.27	11.30	16.82	14.48	12.71	16.14	13.92	14.16	13.64	11.49
2 tick down	3.91	0.88	3.57	2.81	1.61	5.65	3.12	2.73	6.86	0.53
3 tick down	1.58	0.18	1.09	0.88	0.39	2.45	1.18	0.87	3.56	0.08
4 tick down	0.76	0.06	0.45	0.36	0.13	1.19	0.54	0.35	1.88	0.03
5 ticks dowm or more	1.06	0.09	0.55	0.46	0.12	1.73	0.47	0.38	2.97	0.03
statistics of trade durations										
No. of trades[1]	21295	17684	11171	11236	11974	13128	23872	13341	10774	11280
No. of trades[2]	8208	6313	5692	5285	5294	6296	8458	6204	5366	5161
Avg duration per trade[1]	1.10	1.32	2.09	2.08	1.95	1.78	0.98	1.75	2.17	2.07
Avg duration per trade[2]	2.85	3.71	4.11	4.43	4.42	3.72	2.77	3.77	4.36	4.53
Avg trade size	310	867	283	299	491	221	392	341	201	890

Notes: Price movement of 1 cent is the standardized to 1 tick. Trade duration denotes the time between two consecutive transactions. The sample period is from 2008/01/01 to 2010/12/31.

Table 4.2: Summary of the components of the S&P 500 index stocks.

Period	N1	N2	N3	N4	D1	D2	T1	T2
Period 1: 2008/09/01 - 2008/12/31	480	101	0	379	80	83	1000	12350
Period 2: 2010/01/01 - 2010/04/30	493	106	0	387	81	82	493	7322
Period 3: 2012/01/01 - 2012/04/30	497	107	2	388	80	82	333	4841

Notes: N1 denotes the number of stocks that stayed as a component of the S&P 500 index and traded on the NYSE for each entire sample period. N2 denotes the number of stocks with less than 80 trading days during each entire sample period. N3 denotes the number of stocks with stock splits during each entire sample period. N4 denotes the number of stocks remaining for our study. D1 denotes the minimum trading days among all sampled stocks during each sample period. D2 denotes the maximum trading days among all sampled stocks during each sample period. T1 denotes the minimum average transactions per day among all the selected stocks. T2 denotes the maximum average transactions per day among all selected stocks.

Table 4.3: 30-min IVaR backtesting results for the AACD approach.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1: 2008/09/01 - 2008/12/31									
XOM	0.9613	0.6214	0.7011	0.9587	0.9787	1.0000	0.8097	0.3136	0.8908
GE	**0.0349**	**0.0380**	0.7439	0.2676	0.2640	0.6173	0.0863	0.1118	0.9210
PG	0.4794	0.9730	0.1113	0.8585	0.9819	0.9990	0.8644	0.9127	0.3033
JNJ	0.7849	0.9730	0.1934	0.5917	0.7029	0.3286	0.2960	0.8432	0.4576
T	0.3813	0.6214	0.1113	0.9722	0.6148	0.9976	0.7442	0.7484	0.3711
CVX	0.3813	0.6214	0.4450	0.9528	0.9872	1.0000	0.5025	0.2162	0.9304
JPM	0.7044	0.8501	0.7085	0.1678	0.9884	1.0000	0.3569	0.6715	0.6020
WMT	0.8327	0.7934	0.4450	0.5018	0.9978	0.9897	0.5013	0.6611	0.8450
IBM	0.8327	0.6214	0.9830	0.9157	0.9355	1.0000	0.8682	0.6173	0.6983
PFE	0.8327	0.6808	0.5080	0.7364	0.9007	0.9629	0.4734	0.7169	0.2538
Period 2: 2010/01/01 - 2010/04/30									
XOM	0.9546	0.7912	0.0976	0.8544	0.9917	0.9529	0.6293	0.9248	0.1648
GE	0.3459	0.4811	0.4777	0.9891	0.8972	0.9993	**0.0298**	0.1461	0.5036
PG	0.6625	0.6268	0.3006	0.9953	0.8893	0.5086	0.8759	0.7114	0.6920
JNJ	0.5460	0.3577	**0.0247**	0.5482	0.7638	0.4280	0.1501	0.2211	0.1011
T	0.1411	0.6733	0.7085	0.8218	0.9726	0.9752	0.5299	0.9862	0.6464
CVX	0.8268	0.3577	0.3006	0.6919	0.8934	0.9960	0.3299	0.5844	0.5044
JPM	0.1045	0.0797	0.1769	0.3424	0.5204	0.9508	0.1191	0.2383	0.3577
WMT	0.1045	**0.0312**	0.1769	0.7778	0.0614	0.4157	0.2031	0.2077	0.4285
IBM	0.3930	0.2577	0.0506	0.2145	0.7394	**0.0008**	0.6551	0.4929	0.0559
PFE	0.7867	0.6268	0.4777	**0.0048**	**0.0211**	**0.0000**	0.0791	0.2281	0.1444
Period 3: 2012/01/01 - 2012/04/30									
XOM	0.7044	0.3683	0.4717	0.4980	0.9812	0.9959	0.4520	0.5193	0.6991
GE	0.8268	0.1216	0.0976	0.9994	0.6866	0.8565	0.4530	0.3423	0.2987
PG	0.1872	0.0797	0.0976	0.1408	**0.0276**	**0.0000**	**0.0203**	**0.0322**	**0.0306**
JNJ	0.3354	**0.0419**	0.0877	0.7669	0.7889	0.7116	0.5458	0.2174	0.2768
T	0.2440	0.1216	0.7085	0.1733	**0.0056**	**0.0079**	0.2719	0.2264	0.6373
CVX	0.3930	0.1216	0.3006	0.8152	0.8744	0.9926	0.2197	0.4069	0.5406
JPM	0.3930	0.2577	0.0976	0.9962	0.9986	0.6137	0.5191	0.6302	0.2597
WMT	0.1872	0.3577	**0.0021**	0.2683	0.5770	**0.0002**	0.1461	0.3099	**0.0215**
IBM	0.6262	**0.0419**	0.9429	0.7229	0.4096	0.9970	0.6314	0.1838	0.7302
PFE	0.1411	0.3577	0.7085	**0.0469**	0.9165	1.0000	0.2152	0.3228	0.3550

Notes: This table presents the p-values for the Kupiec test, the Engle-Manganelli DQ test using 5 lags with the current IVaR as explanatory variables, and the GMM duration-based test for conditional coverage with 5 moment conditions.

Table 4.4: 30-min IVaR backtesting results for the DDP1 method.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1: 2008/09/01 - 2008/12/31									
XOM	0.5889	0.1424	0.4450	0.3587	0.1667	0.1425	0.4971	**0.0092**	0.4965
GE	0.3672	**0.0013**	**0.0025**	**0.0006**	**0.0000**	**0.0000**	**0.0451**	**0.0140**	**0.0088**
PG	0.9102	0.2913	0.7011	**0.0157**	0.1053	0.2372	0.0676	0.0525	0.3424
JNJ	0.1732	**0.0044**	**0.0004**	**0.0015**	**0.0000**	**0.0000**	**0.0251**	**0.0128**	**0.0127**
T	0.6658	0.3989	0.5080	**0.0485**	**0.0361**	0.9739	0.1473	0.1560	0.1509
CVX	0.3813	0.4654	0.9830	0.1822	0.8853	0.4300	0.0612	0.9556	0.9119
JPM	0.3930	0.1216	**0.0247**	**0.0008**	**0.0003**	**0.0291**	**0.0230**	**0.0319**	**0.0279**
WMT	0.2960	0.8489	0.9830	0.3438	**0.0364**	0.2387	0.6648	0.9063	0.8999
IBM	0.5889	0.7934	0.9830	0.9986	0.9948	1.0000	0.9390	0.6622	0.8910
PFE	0.4555	0.1415	0.1934	0.0650	0.1708	0.1428	0.2057	0.3669	0.4668
Period 2: 2010/01/01 - 2010/04/30									
XOM	0.7044	0.7912	0.7085	0.7399	0.1125	0.2558	0.4318	0.0779	0.4776
GE	0.3124	**0.0018**	**0.0003**	**0.0014**	**0.0000**	**0.0001**	0.1696	**0.0260**	**0.0157**
PG	0.7867	0.7912	0.4777	0.4443	0.8923	0.9470	0.8949	0.7249	0.4143
JNJ	0.8268	**0.0187**	**0.0050**	**0.0060**	**0.0004**	**0.0079**	0.2193	0.0791	**0.0274**
T	**0.0263**	**0.0034**	**0.0003**	**0.0008**	**0.0009**	**0.0001**	0.0788	0.0602	**0.0218**
CVX	0.5900	0.0507	**0.0247**	0.2978	**0.0188**	**0.0432**	0.1946	0.2232	0.1250
JPM	0.1411	0.1798	0.0506	0.1829	**0.0009**	**0.0001**	0.1660	0.0711	0.0669
WMT	0.3930	0.0797	**0.0114**	0.0739	**0.0296**	**0.0000**	0.3525	0.2982	0.0795
IBM	0.2440	0.0507	0.3006	0.0845	0.1886	0.7303	0.6648	0.2314	0.3696
PFE	0.7044	0.8501	0.0506	0.1164	0.4061	**0.0011**	0.9300	0.3837	**0.0364**
Period 3: 2012/01/01 - 2012/04/30									
XOM	0.4399	0.5103	0.2619	0.5974	0.9456	0.9996	0.7257	0.6151	0.7282
GE	0.1872	0.0797	**0.0114**	0.1509	0.1883	**0.0027**	0.4645	0.1801	0.0853
PG	0.2440	0.0797	**0.0050**	**0.0000**	**0.0000**	**0.0000**	**0.0327**	**0.0180**	**0.0196**
JNJ	0.2023	**0.0150**	**0.0215**	0.0983	**0.0113**	**0.0207**	0.5063	0.0687	0.1148
T	0.9154	0.3577	0.0976	0.3826	0.2430	0.0937	0.9040	0.6316	0.1426
CVX	0.3930	0.6268	0.3006	0.4468	0.8554	0.1818	0.1179	0.7724	0.2741
JPM	0.9154	0.3577	0.3006	0.5589	0.4966	0.4271	0.7374	0.7466	0.3016
WMT	0.7044	0.4811	0.0976	0.2330	0.4451	0.2784	0.2626	**0.0215**	0.1586
IBM	0.8700	0.3189	0.2781	**0.0036**	**0.0182**	0.1122	0.0646	0.3449	0.4347

Notes: This table presents the p-values for the Kupiec test, the Engle-Manganelli DQ test using 5 lags with the current IVaR as explanatory variables, and the GMM duration-based test for conditional coverage with 5 moment conditions.

Table 4.5: 30-min IVaR backtesting results for the DDP2 method.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1: 2008/09/01 - 2008/12/31									
XOM	0.0818	**0.0241**	0.1113	0.1405	0.6361	0.9990	**0.0150**	**0.0470**	0.3941
GE	0.4555	0.2913	**0.0282**	**0.0002**	**0.0000**	**0.0000**	0.0811	0.0741	**0.0333**
PG	0.9613	0.3989	0.7439	**0.0365**	0.2532	0.0528	0.0995	0.1754	0.4373
JNJ	0.1732	**0.0137**	0.1934	**0.0011**	**0.0000**	**0.0000**	**0.0251**	**0.0144**	**0.0285**
T	0.2960	0.2239	0.4450	0.9650	0.8967	0.9880	0.5422	0.5751	0.5465
CVX	0.9613	0.8489	0.7011	0.4931	0.9985	1.0000	0.2923	0.8358	0.5994
JPM	0.5900	0.2577	0.4777	0.1047	0.2932	0.6874	**0.0082**	0.0681	0.1637
WMT	0.3813	0.7934	0.5080	0.5113	0.3226	**0.0000**	0.2206	0.5794	0.1172
IBM	0.1647	0.2239	0.1113	0.9802	0.9871	0.9989	0.1254	0.5251	0.1044
PFE	0.7075	0.7934	0.7011	0.7917	0.5511	0.2295	0.9855	0.8511	0.8671
Period 2: 2010/01/01 - 2010/04/30									
XOM	0.6625	0.5103	0.4777	0.8028	0.4685	**0.0000**	0.4600	**0.0470**	**0.0184**
GE	0.5460	0.4811	**0.0114**	**0.0051**	**0.0134**	0.2883	**0.0299**	0.4282	0.0709
PG	0.3459	0.2518	0.4777	0.1292	0.9934	0.9906	0.1389	**0.0169**	0.1286
JNJ	0.7867	0.3577	0.0506	0.1308	0.5252	0.3164	0.1993	0.2366	0.0857
T	0.2440	0.0507	0.0506	0.1687	0.3001	0.7414	0.3771	0.2182	0.1973
CVX	0.2440	0.2577	0.3006	0.0520	0.2187	0.6385	0.1374	0.2668	0.5422
JPM	0.2440	0.1798	**0.0247**	0.1866	0.1990	**0.0291**	**0.0470**	0.4528	0.0667
WMT	0.5460	0.7912	0.3006	0.9130	0.2038	0.5537	0.4149	0.9550	0.5556
IBM	0.9154	0.4811	0.9801	0.5859	0.8038	0.4815	0.7607	0.5652	0.6628
PFE	0.3459	0.3577	0.1769	0.9206	0.1774	0.1377	0.5553	0.6649	0.3358
Period 3: 2012/01/01 - 2012/04/30									
XOM	**0.0456**	0.3683	0.4717	0.6935	0.9938	0.9988	0.0557	0.3423	0.7712
GE	0.4858	0.1798	0.9801	0.8921	0.8985	0.9853	0.6944	0.1770	0.6820
PG	0.3459	0.1798	0.1769	**0.0275**	**0.0002**	0.3615	**0.0226**	**0.0190**	0.0934
JNJ	1.0000	0.9091	0.9429	0.9754	0.9868	1.0000	0.6724	0.8957	0.2363
T	0.1010	0.9683	0.7085	0.7009	0.8416	0.4599	0.2385	0.8959	0.4540
CVX	0.8268	0.2577	0.0976	0.7914	0.7194	0.1340	0.2992	0.6149	0.2468
JPM	0.1978	0.2518	0.7349	0.9700	0.9668	0.3584	0.5637	0.2793	0.8092
WMT	0.9546	0.8501	**0.0247**	0.9866	0.8188	**0.0004**	0.8461	**0.0180**	**0.0177**
IBM	0.1239	0.5579	0.9429	0.3016	0.7351	0.7736	0.2069	0.1992	0.9269
PFE	0.7044	0.9683	0.7085	0.7884	0.7806	1.0000	0.5365	0.7257	0.2633

Notes: This table presents the p-values for the Kupiec test, the Engle-Manganelli DQ test using 5 lags with the current IVaR as explanatory variables, and the GMM duration-based test for conditional coverage with 5 moment conditions.

Table 4.6: 30-min IVaR backtesting results for the G1 method.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1: 2008/09/01 - 2008/12/31									
XOM	0.3813	0.4654	0.9830	0.9492	0.9248	**0.0415**	0.7550	0.0639	0.1571
GE	0.5554	0.0941	**0.0010**	**0.0356**	**0.0445**	**0.0002**	0.0763	0.2024	**0.0303**
PG	0.4794	0.2913	0.0569	0.1754	0.0873	**0.0498**	0.4198	0.1952	0.0960
JNJ	**0.0240**	**0.0006**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0180**	**0.0140**	**0.0070**
T	0.9613	0.0941	**0.0025**	0.4810	**0.0120**	**0.0004**	0.7637	**0.0306**	**0.0115**
CVX	**0.0080**	0.0848	0.2441	0.2050	0.8440	0.9948	**0.0009**	0.1239	0.2514
JPM	0.1045	**0.0187**	**0.0008**	0.1714	0.2012	**0.0032**	0.0516	**0.0405**	**0.0212**
WMT	0.7849	**0.0380**	**0.0001**	0.0732	0.0687	**0.0000**	0.2584	0.1081	**0.0180**
IBM	0.9613	0.0941	**0.0059**	0.7044	0.2715	**0.0001**	0.2894	0.1697	**0.0332**
Period 2: 2010/01/01 - 2010/04/30									
XOM	0.6625	0.3577	0.0506	0.9778	0.9773	0.3206	0.0869	0.8030	0.0940
GE	0.5900	0.1798	**0.0114**	0.5810	0.1728	**0.0001**	0.2569	0.1041	**0.0440**
PG	0.9154	0.2577	**0.0050**	0.4729	**0.0439**	**0.0001**	**0.0074**	0.4131	0.0505
JNJ	0.3930	0.1216	**0.0050**	0.1882	**0.0069**	**0.0001**	0.0864	0.0702	**0.0326**
T	**0.0050**	**0.0000**	**0.0000**	**0.0044**	**0.0000**	**0.0000**	**0.0394**	**0.0123**	**0.0069**
CVX	0.2440	**0.0312**	**0.0114**	0.6651	**0.0467**	**0.0339**	0.6544	0.1037	0.1224
JPM	**0.0263**	**0.0000**	**0.0000**	**0.0007**	**0.0000**	**0.0000**	**0.0174**	**0.0089**	**0.0047**
WMT	0.3124	0.1216	**0.0050**	0.7959	0.5335	**0.0004**	0.5759	0.3969	0.0612
IBM	**0.0381**	**0.0109**	**0.0000**	**0.0225**	**0.0001**	**0.0000**	0.0748	**0.0335**	**0.0095**
PFE	0.8268	0.6268	0.0976	0.3539	0.0933	**0.0157**	0.0603	0.2220	0.0629
Period 3: 2012/01/01 - 2012/04/30									
XOM	0.7044	0.4811	0.0506	0.3512	0.9618	0.9463	0.7854	0.1750	0.2291
GE	0.8268	**0.0187**	0.0506	0.9871	0.4352	0.5974	0.4121	0.1376	0.1943
PG	0.1411	**0.0061**	**0.0003**	**0.0017**	**0.0000**	**0.0000**	0.0501	**0.0123**	**0.0091**
JNJ	0.5177	0.0670	**0.0215**	0.8912	0.6972	0.6127	0.8465	0.2525	0.1177
T	0.3930	0.3577	0.7085	0.3632	0.5469	0.5367	0.7673	0.5479	0.4142
CVX	0.1411	**0.0009**	**0.0000**	0.8281	0.1368	**0.0000**	0.3332	**0.0488**	**0.0125**
JPM	**0.0263**	**0.0005**	**0.0001**	**0.0059**	**0.0000**	**0.0000**	0.1519	**0.0352**	**0.0107**
WMT	0.4858	0.4811	**0.0114**	0.5905	0.2876	**0.0003**	0.4169	**0.0058**	**0.0251**
IBM	0.8700	0.5741	0.2781	0.5571	0.3065	0.9633	0.8726	0.7628	0.4917
PFE	0.8268	0.1798	**0.0050**	0.3321	**0.0003**	**0.0029**	0.4251	0.2175	**0.0483**

Notes: This table presents the p-values for the Kupiec test, the Engle-Manganelli DQ test using 5 lags with the current IVaR as explanatory variables, and the GMM duration-based test for conditional coverage with 5 moment conditions.

Table 4.7: 30-min IVaR backtesting results for the G2 method.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1: 2008/09/01 - 2008/12/31									
XOM	0.6658	0.0941	**0.0059**	0.6677	**0.0002**	**0.0000**	0.2147	0.0667	**0.0384**
GE	0.0698	**0.0013**	**0.0010**	0.0601	**0.0001**	**0.0000**	0.1486	**0.0143**	**0.0154**
PG	**0.0498**	**0.0078**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0130**	**0.0210**	**0.0112**
JNJ	**0.0007**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0114**	**0.0066**	**0.0045**
T	0.1732	**0.0380**	**0.0004**	0.1554	**0.0004**	**0.0000**	0.5065	**0.0235**	**0.0062**
CVX	0.1302	**0.0231**	**0.0004**	0.0668	**0.0065**	**0.0000**	0.0665	**0.0405**	**0.0131**
JPM	**0.0050**	**0.0002**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0109**	**0.0072**	**0.0073**
WMT	0.5554	0.0607	**0.0001**	0.2963	0.0766	**0.0000**	0.7399	0.1072	**0.0125**
IBM	0.0698	0.0607	**0.0059**	0.1518	**0.0213**	**0.0000**	0.3636	0.2960	0.0612
PFE	0.2909	**0.0024**	**0.0000**	**0.0005**	**0.0000**	**0.0000**	**0.0338**	**0.0153**	**0.0077**
Period 2: 2010/01/01 - 2010/04/30									
XOM	0.7867	0.6733	0.4777	0.9991	0.9982	0.9998	0.4281	0.9049	0.1174
GE	0.7867	0.2577	**0.0114**	0.8612	**0.0086**	**0.0001**	**0.0136**	0.0682	**0.0470**
PG	0.6625	0.3577	0.1769	0.6679	0.9935	0.9741	0.3235	0.5303	0.3029
JNJ	0.9546	0.4811	0.0976	0.6915	0.5927	0.2678	0.1255	0.5848	**0.0244**
T	0.0760	**0.0187**	**0.0021**	0.2024	0.9014	0.3536	0.1676	0.0742	**0.0189**
CVX	0.0760	0.0507	0.4777	0.0781	0.2332	0.9974	0.2758	0.2155	0.1641
JPM	0.1872	**0.0001**	**0.0000**	0.6299	**0.0000**	**0.0000**	0.1768	**0.0137**	**0.0075**
WMT	0.4858	0.0507	**0.0050**	0.9958	0.2415	**0.0005**	0.7779	0.1955	**0.0341**
IBM	0.3930	0.1216	0.3006	0.8456	0.6062	0.9971	0.7090	0.4077	0.5029
PFE	0.5460	0.6268	0.4777	0.9999	0.9076	0.6725	0.8762	0.1582	0.6658
Period 3: 2012/01/01 - 2012/04/30									
XOM	0.9546	0.4811	0.3006	0.9466	0.9942	0.9957	0.6433	0.5743	0.5602
GE	0.5900	0.1216	**0.0008**	0.9220	0.7401	**0.0283**	0.4485	0.3661	**0.0341**
PG	0.3930	**0.0109**	**0.0008**	**0.0359**	**0.0001**	**0.0001**	0.2868	0.0501	**0.0158**
JNJ	0.8690	0.2265	0.2781	0.8654	0.5849	0.9966	0.8650	0.5568	0.5595
T	0.5460	0.1216	0.0976	0.3596	0.1798	0.2662	0.4640	0.2263	0.3012
CVX	0.2440	**0.0034**	**0.0050**	0.9347	0.1497	**0.0011**	0.1777	0.0731	0.0559
JPM	0.9546	0.3577	0.1769	0.9997	0.9955	0.9828	0.9828	0.7562	0.3112
WMT	0.7044	**0.0109**	**0.0021**	0.8163	**0.0186**	**0.0000**	0.9247	0.0675	**0.0143**
IBM	0.6177	0.5579	0.6734	0.9989	0.9646	0.9996	0.3542	0.9307	0.4476
PFE	0.6625	0.6733	0.4777	0.9981	0.9857	0.6037	0.2632	0.2876	0.7819

Notes: This table presents the p-values for the Kupiec test, the Engle-Manganelli DQ test using 5 lags with the current IVaR as explanatory variables, and the GMM duration-based test for conditional coverage with 5 moment conditions.

Table 4.8: 30-min IVaR backtesting results for all three periods.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1 for 379 stocks									
AACD	**93.67**	**92.08**	**95.78**	**93.67**	**93.40**	**89.97**	**86.54**	**92.35**	**96.31**
DDP1	63.59	61.48	56.73	37.47	39.58	38.79	58.05	60.42	65.44
DDP2	*84.96*	*78.63*	*80.74*	*72.82*	*72.56*	*64.91*	63.59	*71.50*	*78.89*
G1	52.77	27.70	13.98	45.38	30.61	16.62	64.12	47.76	22.69
G2	66.23	40.90	22.69	55.15	41.69	27.44	*65.96*	50.66	34.56
Period 2 for 387 stocks									
AACD	**92.76**	**86.82**	**81.65**	**92.25**	**86.56**	**79.59**	**82.95**	**84.24**	**81.40**
DDP1	68.48	52.45	43.15	47.03	38.50	31.78	60.41	53.49	44.70
DDP2	*86.05*	*74.68*	*60.72*	*80.88*	*66.93*	*53.49*	*66.47*	*65.89*	*54.01*
G1	57.36	20.41	6.72	60.72	31.52	12.40	58.40	35.66	13.70
G2	72.87	41.34	14.73	74.42	46.77	23.77	63.57	49.10	24.81
Period 3 for 388 stocks									
AACD	**93.81**	**93.04**	**89.69**	**93.56**	**90.98**	**81.44**	**90.98**	**93.04**	**93.30**
DDP1	80.41	69.33	57.22	57.22	50.00	41.49	82.73	79.12	70.36
DDP2	*93.04*	*91.24*	*80.93*	*89.43*	*79.38*	*67.53*	82.99	*88.66*	*85.05*
G1	78.09	53.09	32.47	79.12	60.31	39.18	85.82	73.71	47.42
G2	88.92	67.01	41.24	86.86	67.27	52.58	*89.95*	81.70	60.82

Notes: Summary of 30-min IVaR backtesting results for all selected stocks in three different sample periods. The figures are the percentages of stocks with IVaR backtesting p-value larger than 0.05 over each period. In each column, the bold figures represent the highest percentage and the italic figures represent the second largest.

Table 4.9: IVaR 30-min backtesting summary of the 10 selected stocks.

	Kupiec test			Dynamic quantile test			Duration based test		
IVaR level	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
9:30-10:00									
AACD	6	5	6	6	6	6	7	7	7
DDP2	5	4	3	6	5	4	5	6	5
G2	0	0	0	2	0	0	1	0	0
10:00-10:30									
AACD	10	9	10	10	8	8	9	9	10
DDP2	8	9	10	10	9	9	8	9	10
G2	10	6	2	9	8	5	9	8	5
15:00-15:30									
AACD	4	9	9	10	10	10	6	10	9
DDP2	8	8	8	10	10	10	8	9	8
G2	4	6	10	10	9	9	5	8	10
15:30-16:00									
AACD	9	8	8	9	9	9	9	8	9
DDP2	8	9	8	10	9	8	7	9	9
G2	6	3	2	7	5	1	4	4	2

Notes: Entry of figures are the number of stocks with IVaR backtesting *p*-value larger than 0.05 for different time intervals in sample period 2008/01/01-2010/12/31.

Table 4.10: 60-min IVaR backtesting summary for all 3 periods.

VaR level	Kupiec test			Dynamic quantile test			Duration based test		
	5%	2.5%	1%	5%	2.5%	1%	5%	2.5%	1%
Period 1 for 379 stocks									
AACD	**88.92**	**76.78**	**76.25**	**82.59**	**73.88**	**60.16**	**89.71**	**83.91**	**62.80**
DDP1	48.81	42.74	51.98	37.99	33.77	26.65	54.88	48.28	33.77
DDP2	66.75	59.37	62.27	58.84	50.13	36.68	66.23	63.06	43.80
Period 2 for 387 stocks									
AACD	**90.44**	**79.84**	**70.54**	**90.96**	**82.43**	**73.39**	**89.66**	**87.86**	**71.83**
DDP1	76.23	65.12	50.13	69.25	60.47	47.55	72.61	66.41	43.93
DDP2	82.69	67.70	58.40	81.65	72.35	56.85	75.45	74.16	50.90
Period 3 for 388 stocks									
AACD	**93.56**	**92.01**	**81.70**	92.01	**88.66**	**76.55**	**92.53**	**93.56**	**81.44**
DDP1	83.25	76.80	64.95	81.19	68.81	59.02	90.21	86.08	69.07
DDP2	92.53	89.43	80.67	**92.78**	85.57	71.13	**92.53**	**93.56**	78.35

Notes: Summary of 60-min IVaR backtesting results for all selected stocks in three different sample periods. The figures are the percentages of stocks with IVaR backtesting *p*-value larger than 0.05 over each period. In each column, the bold figures represent the highest percentage.

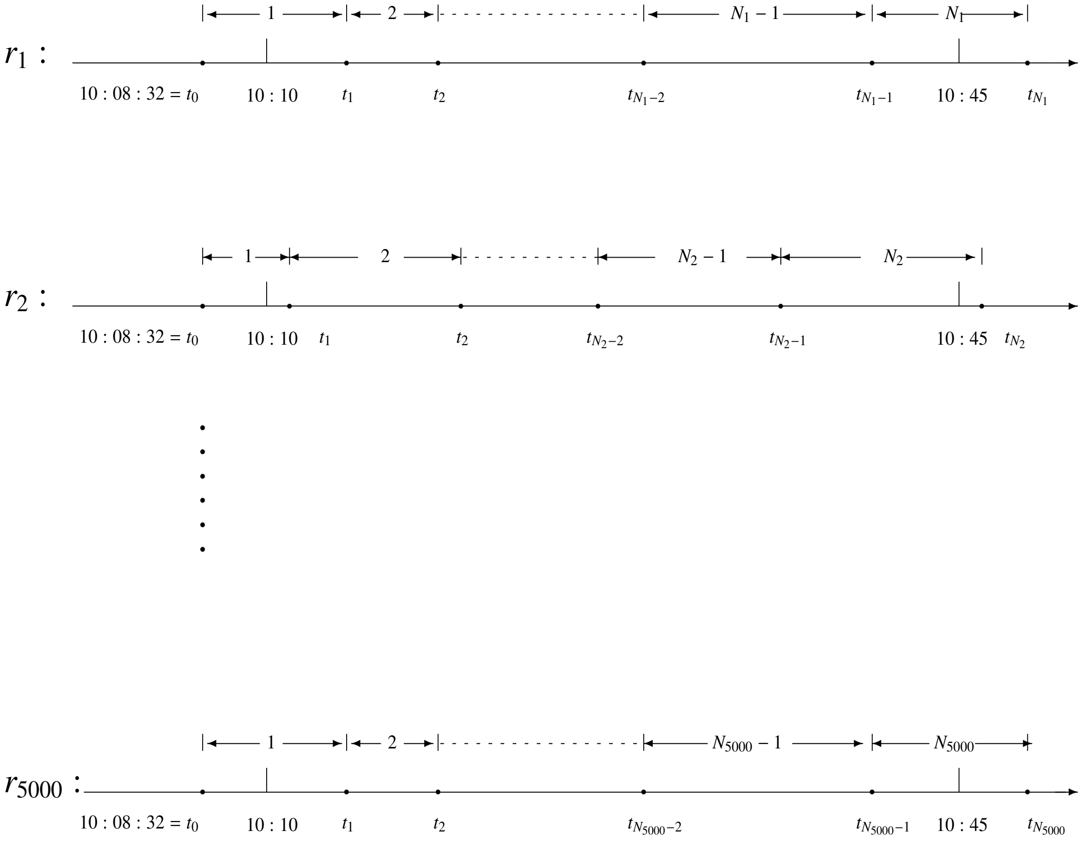


Figure 4.1: Illustration of intraday Monte Carlo simulation procedure.

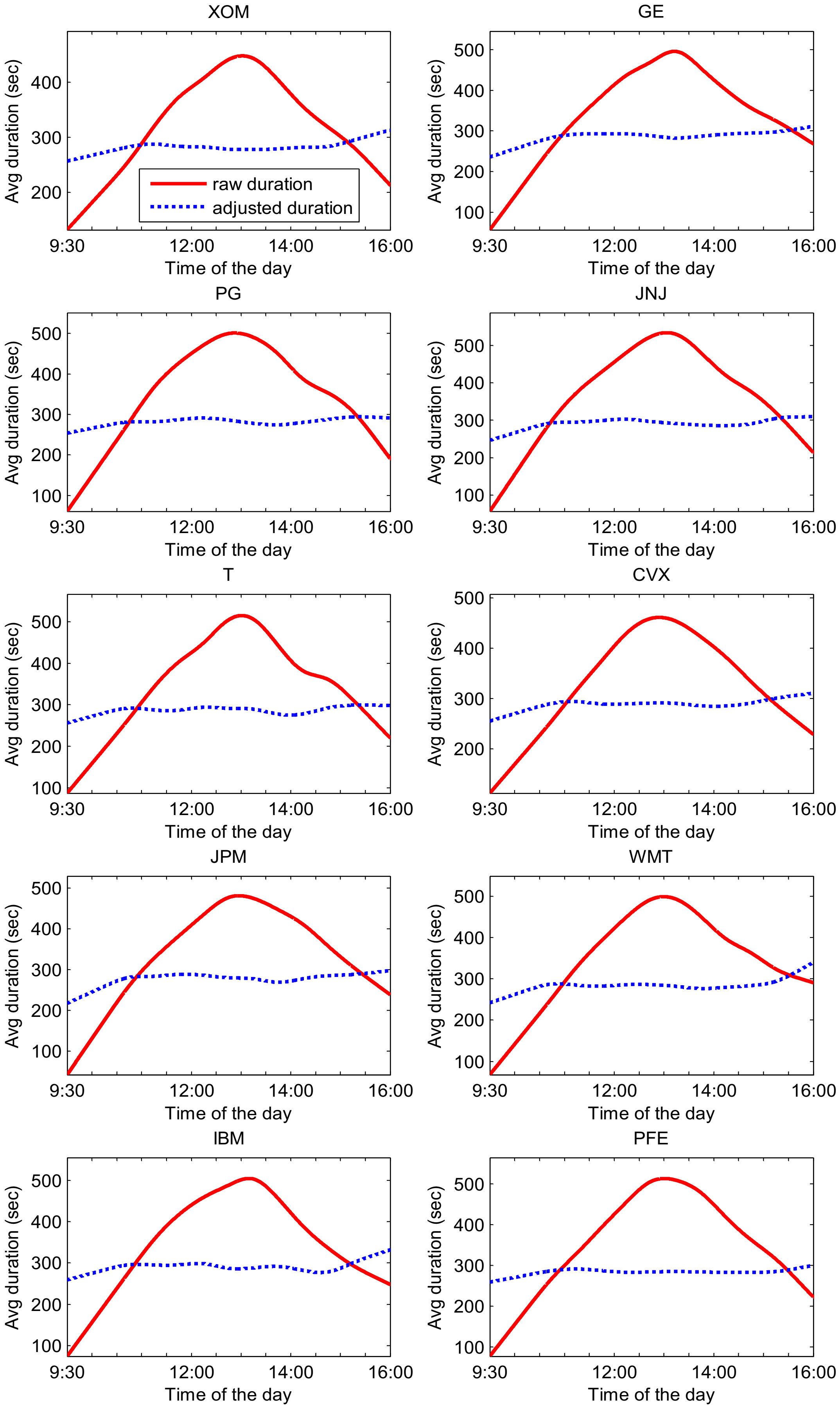


Figure 4.2: Average raw price durations & diurnally adjusted price durations for period 2008/01-2010/12.

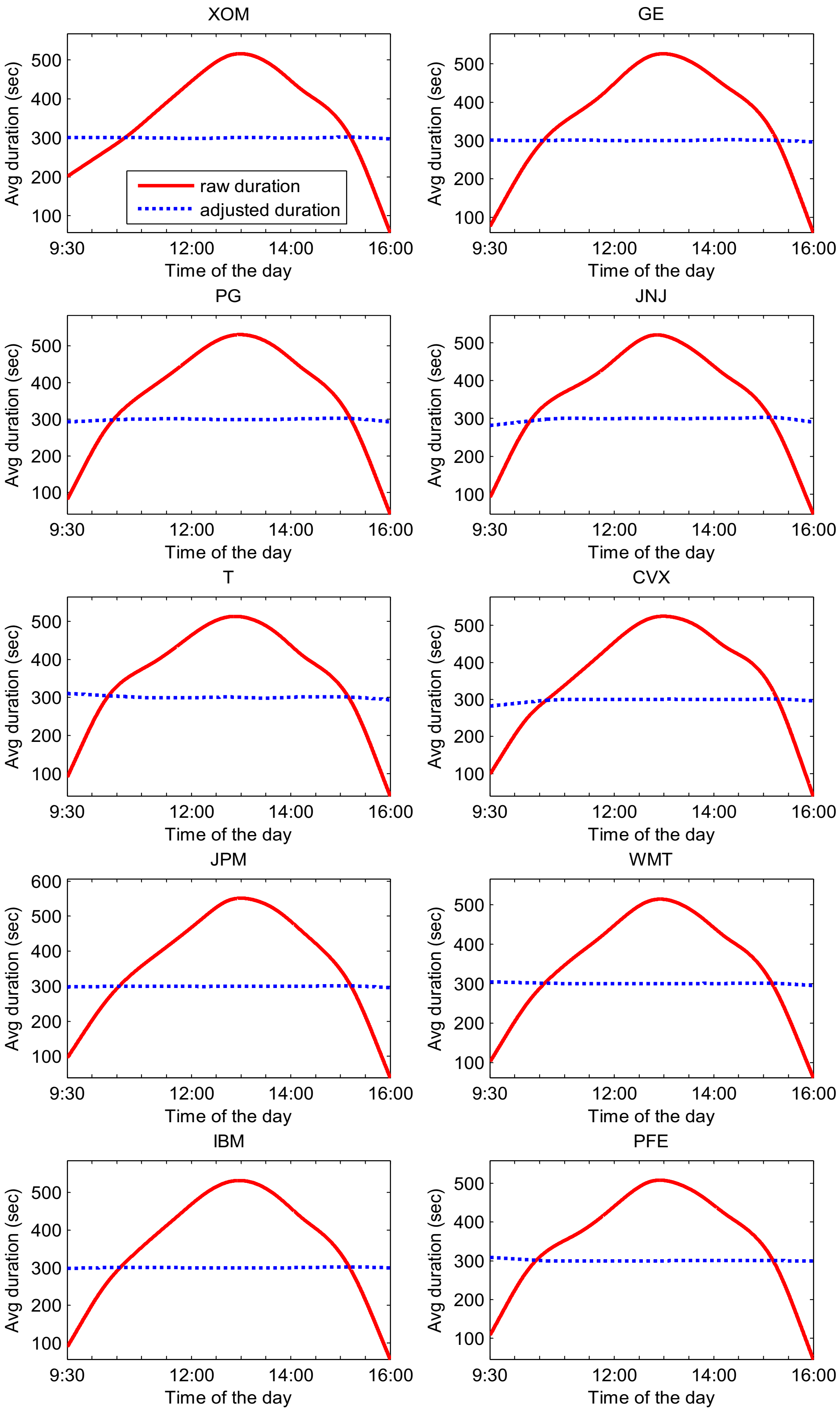


Figure 4.3: Average raw volume durations & diurnally adjusted volume durations for period 2008/01-2010/12.

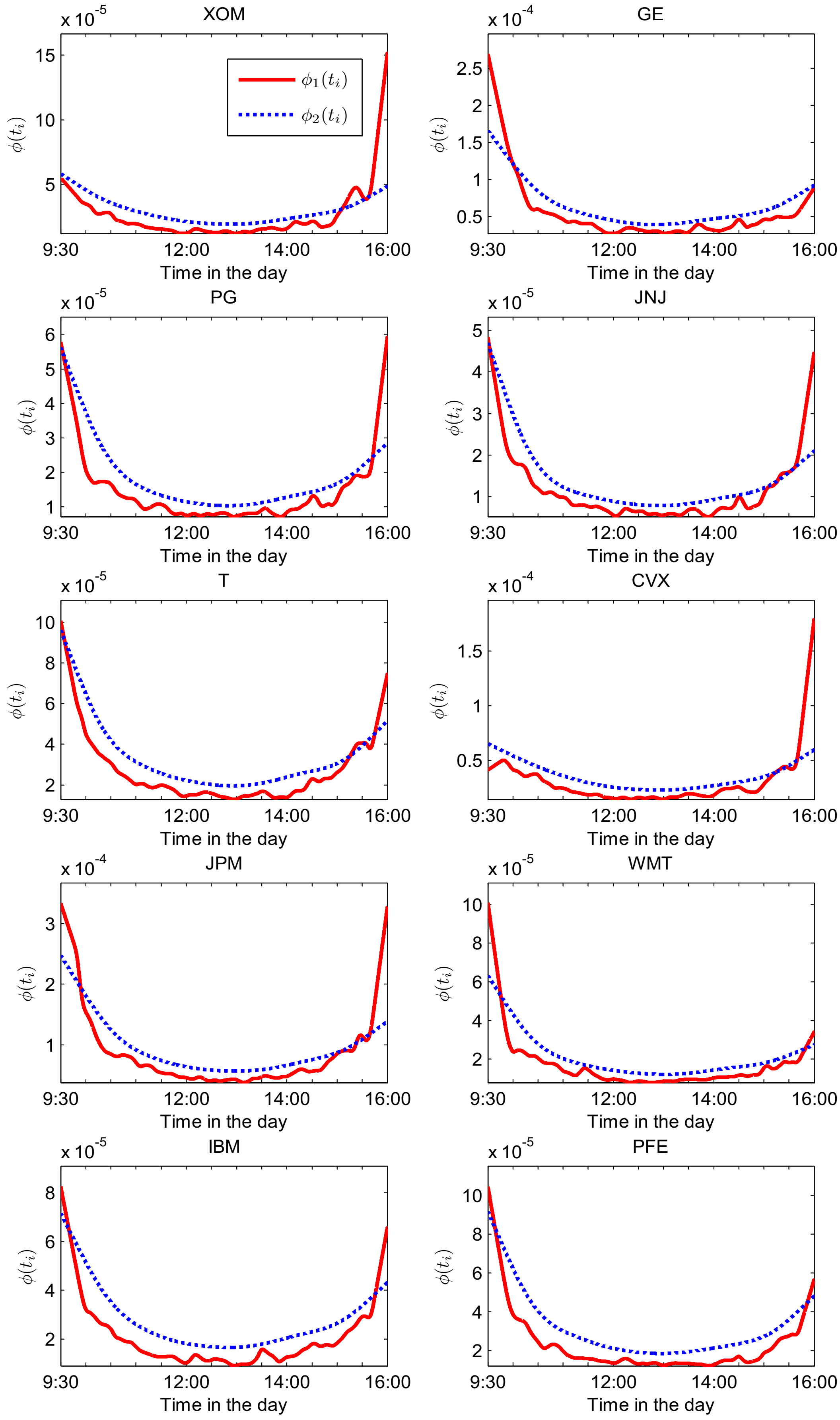


Figure 4.4: The deterministic intraday variance seasonality $\phi(t_i)$ for period 2008/01-2010/12.

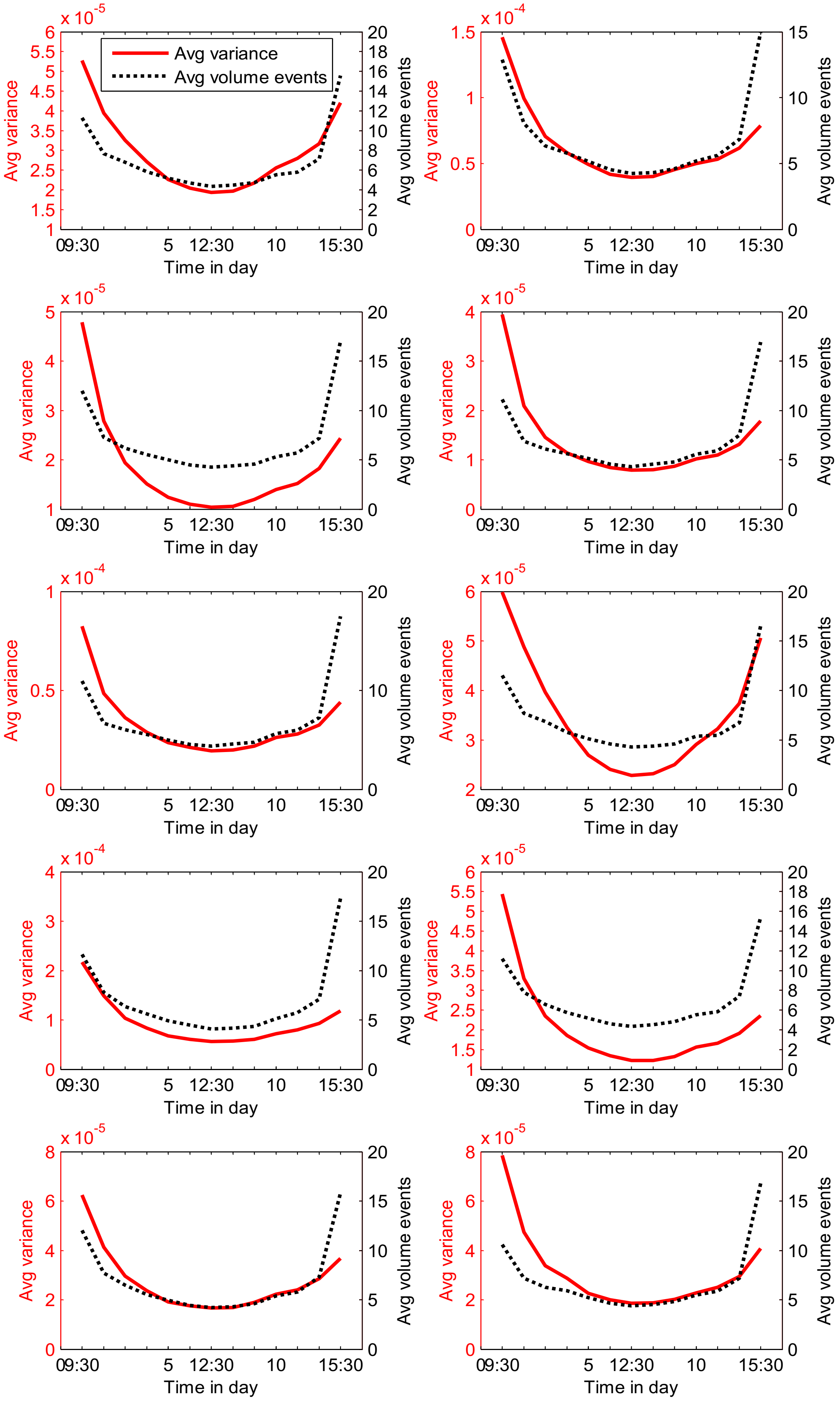


Figure 4.5: The deterministic intraday variance seasonality and the intraday seasonality of volume event numbers for period 2008/01-2010/12.

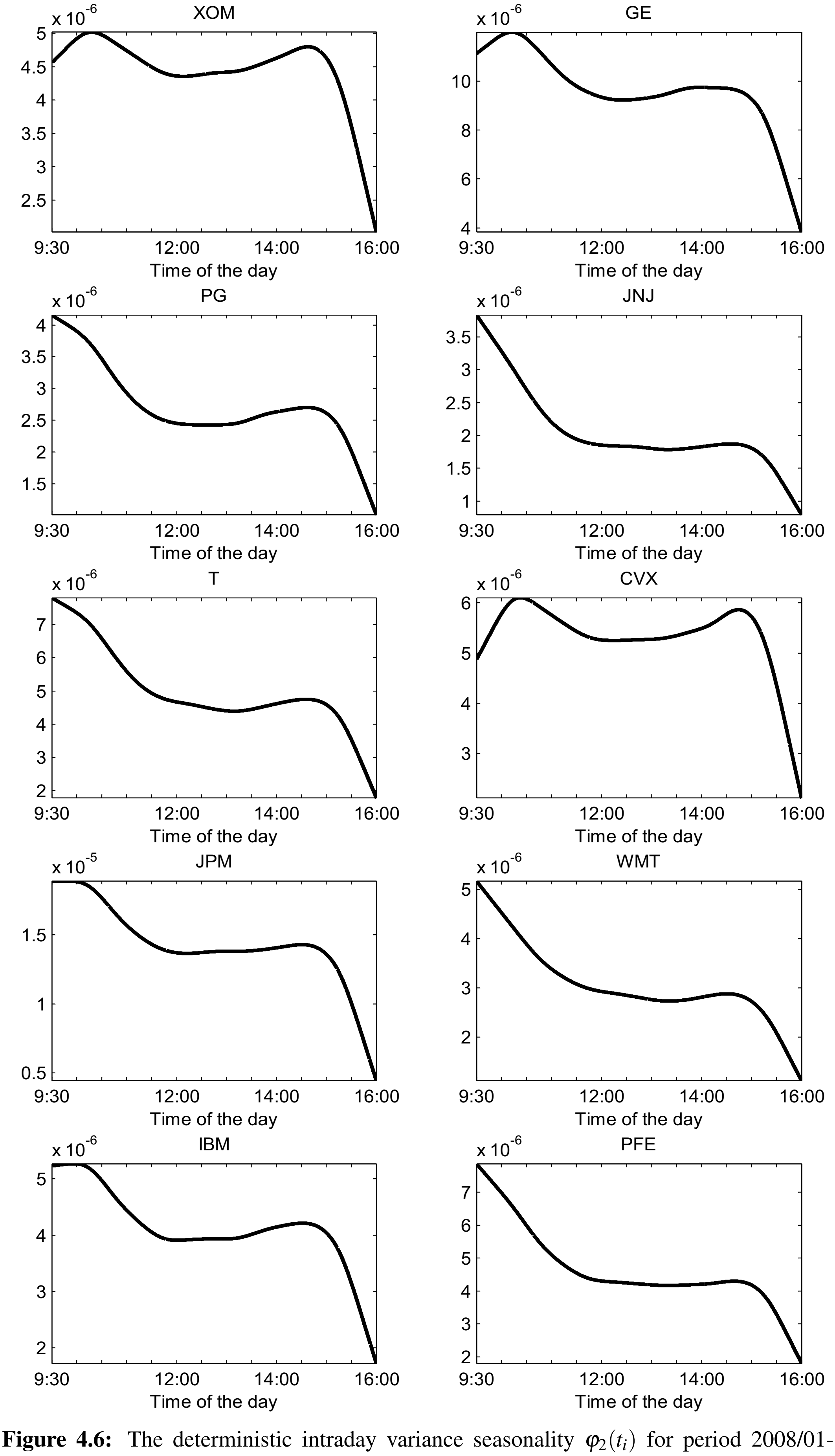


Figure 4.6: The deterministic intraday variance seasonality $\varphi_2(t_i)$ for period 2008/01-2010/12.

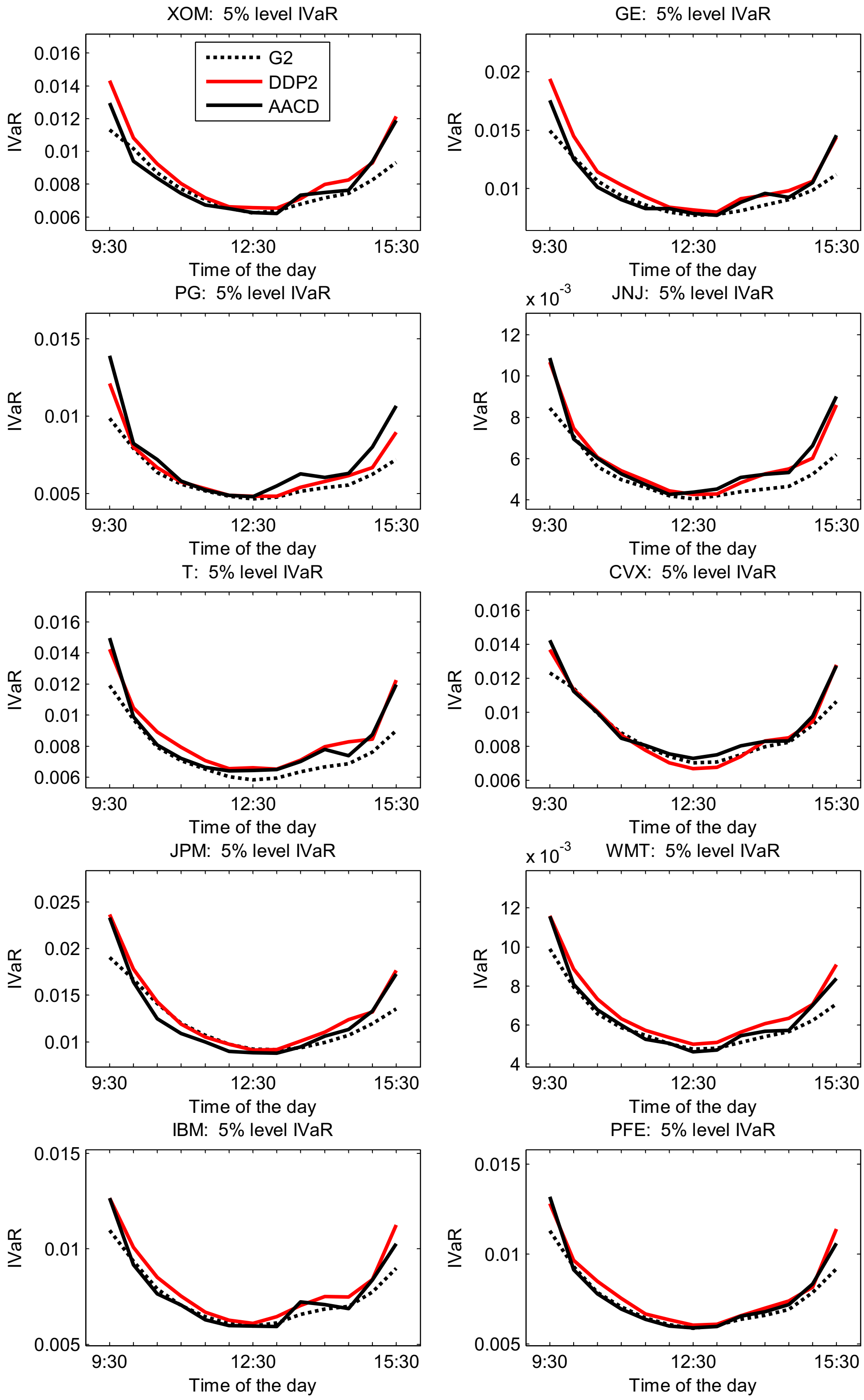


Figure 4.7: 5% level IVaR smile for 3 periods.

Chapter 5 Summary of Conclusions

We have extended the ACD-ICV method proposed by Tse and Yang (2012) to estimate stock volatility over longer intervals such as a month in the chapter 2. Estimation of low-frequency volatility is important for studies involving macroeconomic data that are available only monthly or quarterly. In addition, returns over longer intervals are less susceptible to the contamination of noise over short intervals and may be preferred in studies on asset pricing. Our MC study suggests that price events defined by return of about 0.15% to 0.35% are appropriate for the ACD-ICV method. Based on the transaction data, the ACD-ICV method outperforms the RV method in our MC experiments. On the other hand, if daily data are used, the GARCH method based on aggregating the daily estimates of conditional variance is superior to the RV method, which is widely used in the literature. Our empirical results using ten NYSE stocks show that the ACD-ICV, RV and GARCH estimates track each other quite closely. The RV estimates, however, have larger fluctuations and exhibit occasionally extreme volatility estimates. Co-movements of volatility across different stocks are highest according to the ACD-ICV estimates. Our empirical study on VIX and the S&P500 index shows that VIX is a more successful predictor of future volatility if volatility is estimated by the ACD-ICV method than the RV method.

Chapter 3 proposed to model the aggregate trade volume of stocks in a quote-driven market using a compound Poisson distribution. In our model trades may be initiated by informed or uninformed traders, differentiated by their motivation of trade. We assume that the aggregate volume of each group of traders follow a compound Poisson distribution, with the parameters for the distribution of trades due to informed traders dependent on some information variables. We use two approaches to estimate the model. First, we use tick imbalance as proxy for information variable. Conditional on the tick imbalance, MLE method is used to estimate the model; Second, we treat the information variable random and unobservable, GMM method is used to estimate the model. We then calibrate the model and propose measures of relative intensity of informed trading based on trade frequency and trade volume. Our model treats volume endogenously and does not assume *a priori* that volume and volatility are related. Our empirical analysis of the daily volatility estimates of 50 NYSE stocks confirm that trade frequency dominates trade volume and trade size in affecting volatility. Yet trade volume and trade size have incremental information for volatility beyond that contained in trade frequency. Tick imbalance is an appropriate information proxy under our compound Poisson distribution assumption. Our results also show that informed trading volume increase volatility, while uninformed trading volume reduce volatility. However, for both informed and uninformed traders, the disaggregated effect of trade frequency is to increase volatility. The converse effects of liquidity traders on volatility remains the future research.

Chapter 4 proposed a new method of computing the IVaR using high-frequency transaction data. Intraday directional price movements and price durations are *jointly* modeled by employing the AACD model. We adopt an intraday Monte Carlo simulation approach to estimate IVaR, which enables us to forecast high-frequency returns for any arbitrary interval. In our setup, price durations yield the consecutive steps in time while the price movements allow us to simulate the corresponding returns. Regularly-spaced intraday returns are simply the sum of the price movements simulated conditional on the given horizon. We also modify the DDP method of Dionne, Duchesne and Pacurar (2009) by filtering noise employing volume durations. Using high-frequency data of all the S&P 500 component stocks traded on the NYSE over three different periods, our results show that the IVaR estimates computed using the AACD approach track closely to the DDP and Giot methods. IVaR backtesting results show that the AACD approach performs the best over other methods, based on the results on backtesting p-values. Our robustness check shows that the AACD approach is not sensitive to the choice of the price range δ, provided that the price events sampled are not too infrequent.

Bibliography

Aït-Sahalia, Y., and L. Mancini, 2008, Out of sample forecasts of quadratic variation, Journal of Econometrics, 147, 17-33.

Akgiray, V., 1989, Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts, Journal of Business, 62, 55-80.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001, The distribution of Realized volatility, Journal of Financial Economics, 61, 43-76.

Andersen, T. G., 1996, Return volatility and trading volume: An information flow interpretation of stochastic volatility, Journal of Finance 51, 169-204.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2001, The distribution of exchange rate volatility, Journal of American Statistical Assoction, 96, 42-55.

Andersen, T.G., D. Dobrev, and E. Schaumburg, 2008, Duration-based volatility estimation, working paper, Northwestern University.

Angelidis, T., and A. Benos, 2006, Liquidity adjusted value-at-risk based on the components of the bid-ask spread, Applied Financial Economics, 16, 835-851.

Bali, T.G., N. Cakici, X.S. Yan, and Z. Zhang, 2005, Does idiosyncratic risk really matter? Journal of Finance, 60, 905-929.

Barndorff-Nielsen, O.E., and N. Shephard, 2002, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, Journal of the Royal Statistical Society B, 63, 167-241.

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., P.R. Hansen, A. Lunde, and N. Shephard, 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, Econometrica, 76, 1481-1536.

Bauwens, L., and P. Giot, 2003, Asymmetric ACD models: introducing price information in ACD models, Empirical Economics, 28, 709-731.

Bauwens, L., P. Giot, J. Grammig, and D. Veredas, 2004, A comparison of financial duration models via density forecasts, International Journal of Forecasting, 20, 589-609.

Becker, R., A.E. Clements, and S.I. White, 2007, Does implied volatility provide any information beyond that captured in model-based volatility forecasts? Journal of Banking and Finance, 31, 2535-2549.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307-327.

Candelon, B., G. Colletaz, C. Hurlin, and S. Tokpavi, 2011, Backtesting value-at-risk: a GMM duration-based test, Journal of Financial Econometrics, 9, 314-343.

Chan, C. C., and W. M. Fong, 2006, Realized volatility and transactions, Journal of Banking and Finance 30, 2063-2085.

Chan, K., and W. M. Fong, 2000, Trade size, order imbalance, and the volatility-volume relation, Journal of Financial Economics 57, 247-273.

Chung, S.L., W.C. Tsai, Y.H. Wang, and P.S. Weng, 2011, The information content of the S&P500 index and VIX options on the dynamics of the S&P500 index, working paper, 21st Asia-Pacific Futures Research Symposium.

Coroneo, L., and D. Veredas, 2011, A simple two-component model for the distribution of intraday returns, The European Journal of Finance, 10, 1-23.

Dionne, G., P. Duschesne, and M. Pacurar, 2009, Intraday value at risk (IVaR) using tick-by-tick data with application to the Toronto Stock Exchange, Journal of Empirical Finance, 16, 777-792.

Drost, F.C., and B.J.M. Werker, 2004, Semiparametric duration models, Journal of Business and Economic Statistics, 22, 40-50.

Dumitrescu, A., 2010, The strategic specialist and imperfect competition in a limit order market, Journal of Banking and Finance, 34, 255-266.

Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, Is information risk a determinant of asset returns?, Journal of Finance 57, 2185-2221.

Easley, D., N. Kiefer, M. O'Hara, and J. B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, Journal of Finance 51, 1405-1436.

Engle, R., and A. Lunde, 2003, Trades and quotes: a bivariate point process, Journal of Financial Econometrics, 1, 159-188.

Engle, R., 2000, The econometrics of ultra-high-frequency data, Econometrica, 68, 1-22.

Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica, 50, 987-1007.

Engle, R., and S. Manganelli, 2004, CAViaR: conditional autoregressive value-at-risk by regression quantiles, Journal of Business & Economic Statistics, 22, 367-381.

Engle, R.F., and V.K. Ng, 1993a, Measuring and testing the impact of news on volatility, Journal of Finance, 48, 1749-1778.

Engle, R.F., and V.K. Ng, 1993b, Timing-varying volatility and the dynamic behavior of the term structure, Journal of Money, Credit & banking, 25, 336-349.

Engle, R.F., and J.R. Russell, 1998, Autoregressive conditional duration: A new model for irregularly spaced transaction data, Econometrica, 66, 1127-1162.

Fama, E.F., 1976, Inflation uncertainty and expected returns on Treasury bills, Journal of Political Economy, 84, 427-448.

Fernandes, M., and J. Grammig, 2005, Non-parametric specification tests for conditional duration models, Journal of Econometrics, 127, 35-68.

Fernandes, M., and J. Grammig, 2006, A family of autoregressive conditional duration models, Journal of Econometrics, 130, 1-23.

Fisher, L., 1966, Some new stock-market indexes, Journal of Business, 39, 191-225.

French, K.R., G.W. Schwert, and R.F. Stambaugh, 1987, Expected stock returns and volatility, Journal of Financial Economics, 19, 3-29.

Fu, F., 2009, Idiosyncratic risk and the cross-section of expected stock returns, Journal of Financial Economics, 91, 24-37.

Ghysels, E., C. Gourierous, and J. Jasiak, 2004, Stochastic volatility duration models, Journal of Econometric, 119, 413-435.

Giot, P., 2005, Market risk models for intraday data, The European Journal of Finance, 11, 309-324.

Giot, P., and J. Grammig, 2006, How large is liquidity risk in an automated auction market? Empirical Economics, 30, 867-887.

Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance, 48, 1979-1801.

Goettler, R. L., C. A. Parlour, and U. Rajan, 2009, Informed traders and limit order markets, Journal of Financial Economics 93, 67-87.

Gourieroux, C., J. Jasiak, and G. Le Fol, 1999, Intraday market activity, Journal of Financial Markets, 2, 193-226.

Goyal, A., and P. Santa-Clara, 2003, Idiosyncratic risk matters! Journal of Fiance, 58, 975-1007.

Grammig, J., and K.O. Maurer, 2000, Non-monotonic hazard fuctions and the autoregressive conditional duration Model, Econometrics Journal, 3, 16-38.

Groth, S.S., and J. Muntermann, 2011, An intraday market risk management approach based on textual analysis, Decision Support Systems, 50, 680-691.

Guo, H., and R. Savickas, 2008, Average idiosyncratic volatility in G7 countries, Reviews of Financial Studies, 21, 1259-1296.

Harris, M., and A. Raviv, 1993, Differences of opinion make a horse race, Review of Financial Studies 6, 473-506.

Hendershott, T., and P. C. Moulton, 2011, Automation, speed, and stock market quality: The NYSE's hybrid, Journal of Financial Markets 14, 568-604.

Heston, S.L., 1993, A closed-form solution for options with stochastic volatility with application to bond and currency options, Review of Financial Studies, 6, 327-343.

Huang, R. D., and R. W. Masulis, 2003, Trading activity and stock price volatility: Evidence from the London Stock Exchange, Journal of Empirical Finance 10, 249-269.

Jiang, G.J., and Y.S. Tian, 2005, The model-free implied volatility and its information content, Review of Financial Studies, 18, 1305-1342.

Jiang, G.J., and Y.S. Tian, 2010, Forecasting volatility using long memory and comovements: An application to option valuation under SFAF 123R, Journal of Financial and Quantitative Analysis, 45, 502-533.

Jones, M. J., G. Kaul, and M. L. Lipson, 1994, Transactions, volume, and volatility, Review of Financial Studies 7, 631-651.

Kupiec, P., 1995, Techniques for verifying the accuracy of risk measurement models, The Journal of Derivatives, 3, 73-84.

Lee, C. M. C., and M. A. Ready, 1991, Inferring trade direction from intraday data, Journal of Finance 46, 733-746.

Li, J., and C. Wu, 2006, Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume, Journal of Business 79, 2697-2739.

Ludvigson, S.C., and S. Ng, 2007, The empirical risk-return relation: A factor analysis approach, Journal of Financial Economics, 83, 171-222.

Merton, R.C., 1980, On estimating the expected return on the market: An exploratory investigation, Journal of Financial Economics, 8, 323-361.

Nelson, D.B., 1991, Conditional heteroskedasticity in assert returns: a new approach, Econometrica, 59, 347-370.

Officer, R., 1973, The variability of the market factor of the New York Stock Exchange, Journal of Business, 46, 434-454.

Pacurar, M., 2008, Autoregressive conditional duration (acd) models in finance: A survey of the theoretical and empirical literature, Journal of Economic Surveys, 22, 711-754.

Pagan, A., and G.W. Schwert, 1990, Alternative models for conditional stock volatility, working paper, NBER.

Pascual, R., and D. Veredas, 2010, Does the open limit order book matter in explaining informational volatility? Journal of Financial Econometrics 8, 57-87.

Pacurar, M., 2008, Autoregrrive conditional duration (ACD) models in finance: a survery of the theoreticial and empirical lituature, Journal of Economic Surverys, 22, 711-751.

Roşu, I., 2009, A dynamic model of the limit order book, Review of Financial Studies 22, 4601-4641.

Roşu, I., 2010, Liquidity and information in order driven markets, working paper.

Sakalauskas, V., and D. Kriksciuniene, 2006, Short-term investment risk measurement using VaR and CVaR, Computational Science-ICCS 2006, 316–323.

Scholes, M., and J. Williams, 1977, Estimating betas from nonsynchronous data, Journal of Financial Economics, 5, 309-327.

Schwert, G.M., 1989, Why does stock market volatility change over time? Journal of Finance, 44, 1115-1153.

Schwert, G.M., 1990a, Stock market volatility, Financial Analysts Journal, 46, 23-34.

Schwert, G.M., 1990b, Stock Volatility and the Crash of '87, Review of Financial studies, 3, 77-102.

Schwert, G.M., and P.J. Seguin, 1990, Heteroskedasticity in stock returns, Journal of Finance, 45, 1129-1155.

Sentana, E., 1995, Quadratic ARCH models, Review of Economic Studies, 62, 639-661.

Tay, A.S., C. Ting, Y.K. Tse, and M. Warachka, 2011, The impact of transaction duration, volume and direction on price dynamics and volatility, Quantitative Finance, 11, 447-457.

Tse, Y. K., 2009, *Nonlife Actuarial Models*: *Theory, Methods and Evaluation*, Cambridge University Press.

Tse, Y. K., and T. T. Yang, 2012, Estimation of high-frequency volatility: An autoregressive conditional duration approach, Journal of Business and Economic Statistics 30, 533-545.

Tse, Y.K., and Y. Dong, 2012, Intraday periodicity adjustments of transaction duration and their effects on high-frequency volatility estimation, Singapore Management University, working parper.

Wang, J., 1993, A model of intertemporal asset prices under asymmetric information, Review of Economic Studies 60, 249-282.

Wang, J., 1994, A model of competitive stock trading volume, Journal of Political Economy 102, 127-168.

Whaley, R.E., 2009, Understanding the VIX, Journal of Portfolio Management, 35, 98-105.

Wu, Z., 2012, On the intraday periodicity duration adjustment of high-frequency data, Journal of Empirical Finance, 19, 282-291.

Xu, X. E., P. Chen, and C. Wu, 2006, Time and dynamic volume-volatility relation, Journal of Banking and Finance 30, 1535-1558.

Zhang, M.Y., J.R. Russell, and R.S. Tsay, 2001, A nonlinear autoregressive conditional duration model with application to financial transaction data, Journal of Econometrics, 104, 179-207.

Zhang, C., 2010, A Reexamination of the Causes of Time-Varying Stock Return Volatilities, Journal of Financial and Quantitative Analysis, 45, 663-684.

Zhang, L., P.A. Mykland, and Y. Aït-Sahalia, 2005, A tale of two time scales, Journal of the American Statistical Association, 100, 1394-1411.