Essays on Realized Measures of Volatility

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Peiran Shi

Department of Economics

Lancaster University

October 2013

父兮生我，母兮鞠我。拊我畜我，长我育我，顾我复我，出入腹我。

——《诗经·小雅·蓼莪》

This thesis is dedicated to

My parents.

Abstract

This thesis investigates the stylized facts of realized measures of volatility in 10 different market sectors. Traditionally, studies in the area have addressed the issues by either using a single measure on a number of stocks or indices, or a number of measures on a given stock or an index. This usually provides results that cannot be generalized; hence does not allow for discussing these measures comparatively, nor fully quantifies the gains from using high frequency data in general.

Using 100 stocks from 10 sectors over the period 2000 - 2010, we investigate topics within the high frequency context of various realized volatility measures.

In Chapter 1, we investigate whether the stylized facts of different realized measures vary across sectors. To this end, our work could be seen as an extension of Andersen *et al* (2001), Luu and Martens (2003), Andersen *et al* (2010), Fleming and Paye (2011), and Giot *et al* (2010). Our findings here are of interest as it provides guidance as whether certain realized measures are best suited to address specific queries relative to others.

In Chapter 2, we revisit the volatility-volume (number of trades) relation. The literature takes it as a task to establish as which is a better measure of the market activity. Despite numerous studies, this remains an open question, a query that we will address as a part of our investigation. We revisit this relation within the context of what is known as the mixture of distributions hypothesis. We aim to investigate whether this relation is stable across different sectors and whether it is measure dependent. We also aim to show that the information content between the two activity measures is distinct. We find that on average, the number of trades is a better proxy

for market activity. We also show that a trade that accompanies a price change is more important than one which takes place at the same price.

In Chapter 3, we address the issue of recovering returns normality using parametric and non-parametric measures of volatility. Returns are not normal, as evident from the vast number of empirical studies that investigate their stylized facts. The finding that returns normality could be achieved through standardization is based on the assumption that any semi-martingale process could be written as a time-changed Brownian motion. The aim in this chapter is to highlight the important factors that may affect recovering returns normality. We look at factors such as the frequency at which the realized measures are estimated, the level of stock activity, the effect of jumps and microstructure noise. We find that the most dominant factors are the sampling frequency and microstructure noise.

Overall, this thesis seeks to investigate the outlined topics to check whether the extensively reported findings still hold by using a very refined data.

Acknowledgements

I would like to take this opportunity to thank for the help I received during the journey of my Ph.D. First and for most, I would like to express my deepest gratitude to my supervisors, Dr Marwan Izzeldin, Professor Ivan Paya and Professor David Peel. Thank you so much for your kind and constant inspiration, help and support in the past four years. I would never have fulfilled this dream without you.

I thank Mr Gerald Steele for his patience and for his kindness in helping go through the mistakes in my English. I am also grateful to the faculty and staff of the Department of Economics. It is really a pleasure to work with you.

Last but not least, I am thankful to my colleagues and my friends. The years with you shall never fade away.

Contents

Introduction

The discussion and subject matter of this thesis is motivated by the recent development of realized volatility measures. Volatility is fundamental to asset pricing, risk management, and portfolio allocations. It plays a crucial role in financial investment. Modelling volatility has drawn great interest from both scholars and practitioners in recent decades. With the availability of high frequency data and high performance computers, the estimation of volatility is now *ex-post*. The so-called realized volatility measures, which are based on continuous time jump diffusion frameworks, have been analyzed and modelled extensively. It has been generally accepted in the literature that realized volatility measures provide better in-sample fitness and out-of-sample forecasting than traditional volatility measures, such as GARCH, absolute returns and squared returns. The superior performance of various realized volatility measures has been discussed and reported using a wide range of financial, or even non-financial, data.

Although realized volatility measures have been studied extensively, a gap currently existing in the literature is that relatively few studies look at the comparative performance of different realized volatility measures. Using four extensively studied realized volatility measures which are calculated from 100 stocks traded in the US equity market spanning an over 10-year time period, this thesis aims to fill the gap. Our data are collected from a bias/error corrected database and show a more accurate estimation of the measures than the commonly used TAQ database, especially for realized range. The 100 stocks are further segmented into 10 market sectors to check the extent to which the validity of models tested varies across sectors. To the best of our knowledge, this work is one of the earliest empirical investigations on modelling realized volatility measures at the market sector level.

The thesis is organized as follows:

Chapter 1: On the Stylized Facts of Realized Measures of Volatility in Different Market Sectors. This chapter investigates the stylized facts of four popular realized volatility measures: realized variance, realized range, realized power variation and realized bipower variation. Volatility measures are compared and valued using various commonly adopted econometric techniques. We address several stylized facts under the continuous time jump-diffusion model framework: a) optimal sampling frequency; b) impact of the presence of jumps; c) pair-wise correlations; d) volatility regimes under Markov-Switching dynamic model; e) leverage and feedback effects; f) long memory; g) volatility-volume relation; and h) the distributional properties of volatility and volatility standardized return. This chapter compares the performance of different realized volatility measures and attempts to draw a conclusion whether any measure produces most consistently best results and hence can be concluded as the most accurate estimator of the true latent volatility. This chapter also aims to draw comparisons between different market sectors.

Chapter 2: On the Significance of Trading Volume and Number of Trades in High Frequency Data. This chapter addresses the volatility-volume (number of trades) relation from the conclusions reached in Chapter 1. We attempt to determine which market activity measures best explain volatility. To address the issue, a series of econometric techniques have been adopted. These include looking at: a) the long memory properties of the market activity measures; b) correlation structure linking volatility and activity measuring variables; c) common structural breaks in volatility measures and activity variables; d). regression analysis; e) Granger causality; f) The performance of GARCH augmented with market activity measures; and g) Estimating the moments of information flow using GMM. By decomposing number of trades according to prices change, we find that the stronger explanatory power of number of trades on volatility reported in the literature should be attributed to the number of trades that is happened when prices change. The number of trades happened when prices remain the same does not necessarily carry more market latent information flow than trading volume. The findings of this chapter also suggest that the MDH is better supported by using more accurate volatility measures.

Chapter 3: Factors Affecting Recovering Returns Normality Using Parametric and Non-Parametric Volatility Measures. This chapter discusses the issue of recovering returns normality using both parametric and non-parametric volatility measures. We investigate the impact of stock type and activity level on the capacity of various volatility measures to achieve return normality. The second task is to look at the impact of a) sampling frequency; b) jumps; and c) market microstructure noise on the distributional properties of standardized returns. We provide the comparison between different volatility measures (parametric and nonparametric) and between different sampling frequencies. We recognize the impact of microstructure noise on recovering returns normality and suggest a moving average filtration approach that applies to all realized volatility measures and that is capable of providing improvements over the jumps robust measure such as the bipower variation.

Chapter 1

On the Stylized Facts of Realized Measures of Volatility in Different Market Sectors

Abstract

We investigate the stylized facts of four extensively studied realized volatility measures, namely, realized variance, realized range, realized power variation and realized bipower variation, within the context of the continuous-time jump diffusion model. Different realized volatility measures are compared and evaluated using various econometric techniques which are commonly adopted in the literature. We look at optimal sampling frequency, impact of jumps, distributional properties of returns and volatility, long memory, volatility regimes, the volatility-volume relation and the recovering returns normality. We use a data set of 100 stocks, representing 10 sectors over the period 2000-2010. To the best of our knowledge, this is the largest data set ever investigated in empirical research in this area.

Our findings show the properties of the realized measures vary widely across sectors. The results obtained add to our understanding about how different sectors operate, especially during the financial crisis.

1.1 Introduction

Volatility modelling and estimation are widely addressed in the literature, especially in relation to its importance in pricing risk and the desire to understand how financial markets operate. Modelling volatility accurately is vitally important for option pricing, risk management and portfolio selection. Early studies use absolute returns, squared returns, stochastic volatility, and (G) ARCH family models, and so on, as volatility measures and model them using various datasets. Although estimated differently, a set of statistical facts have emerged from the empirical studies of volatility measures, which are common to a variety of financial assets and markets. These common properties of volatility are known as stylized facts and have been extensively discussed. Several studies investigate the stylized facts of volatility measures particularly. To cite a few: Karpoff (1987) discusses the volatility-volume relation and finds a positive relation to hold in both equity and future markets; Granger and Ding (1995) investigate the properties of absolute returns and find (i) volatility has a long memory and decays slowly, (ii) the moments of absolute returns are exponentially distributed; Malmsten and Teräsvirta (2004) investigate three popular volatility models (GARCH, EGARCH and Autoregressive Stochastic Volatility) and show how these volatility measures are more or less capable of reproducing the observed stylized facts of financial assets. Their main findings document the presence of high kurtosis and slow decaying autocorrelation functions in all volatility measures.

These above studies were based on either daily data or monthly data. In the past few decades, the growth of financial markets, advances in computer power and the availability of high frequency financial data have given scholars and practitioners new motivation to model and forecast volatilities. The availability of high frequency data also allows market information to be gauged at intraday levels, allowing more precise measurements of volatility. Before the availability of high frequency data, the sampling frequency spanned from daily, weekly, monthly to even quarterly and annually. Nowadays, prices can be collected at ultra-high frequency level, often termed, tick-by-tick level. This development in the dataset has pushed the volatility modelling on to new ground. There are many drawbacks in using low frequency data to construct volatility measure. One of which, is that *"the standard latent volatility models fail to describe in an adequate manner is the low, but slowly decreasing, autocorrelations in the squared returns that are associated with high excess kurtosis of returns"* (McAleer *et al*, 2008).

Measures based on high frequency data are usually classified as non-parametric. The current empirical literature focuses on four high frequency measure, namely realized volatility (rv_t), realized bi-power variation (bv_t), realized power variation (pv_t) and realized range (rr_t). See Andersen and Bollerslev (1998), Christensen and Podolskij (2007), Martens and van Dijk (2007), and Barndorff-Nielsen and Shephard (2003, 2004, and 2006). Realized measures of volatility assume continuity of the underlying volatility process which contradicts some of the empirical findings documented in the recent literature. Empirical findings have also showed that the continuity assumptions are more likely to be met in active stocks than in least active stocks. See for example, Aït-Sahalia and Jacod (2009a, b, 2010).

The advantages of using high frequency data to estimate volatility measures are soundly based. For instance, high frequency volatility measures, or 'realized' volatility measures, do not require explicitly modeling the intraday data. Most of the realized volatility measures are treated as observed rather than latent, and hence are easier to estimate relative to parametric volatility models such as stochastic volatility. Moreover, those model-free estimators produce significant improvements in both in-sample fitness and out-of-sample forecasting. They are unbiased and highly efficient estimators of the integrated volatility under certain conditions.

The superiority of realized volatility measures over other low frequency volatility measures have been discussed and confirmed empirically. A number of studies have addressed the stylized facts of volatility measures in general (Ghysels *et al* (2006) & Fuertes *et al* (2009)), and of the high frequency measures in specific (ABDL (2001&2003); Christensen and Podolskij (2007), Martens and van Dijk (2007), and Barndorff-Nielsen and Shephard (2004 and 2006). When compared with low frequency volatility measures, high frequency volatility measures show a better performance. In addition, a number of stylized facts have emerged for the high frequency measures. Giot and Laurent (2004) summarize the stylized fact of realized volatility as follows: realized volatility is highly skewed and kurtosed, yet the logarithmic realized volatility is Gaussian. Both realized volatility and logarithmic realized volatility appear to be fractionally integrated and they both show long-range dependence, as well as slowly decreasing autocorrelation functions. There are both leveraged and feedback effects presented in the realized volatility. Furthermore, Corsi *et al* (2012) add that the presence of jumps in realized volatility is relatively infrequent and unpredictable but has a strong impact on future volatility. Realized volatility measures are also found to possess the stylized facts of low frequency volatility measures and, to an even greater extent, such as the positive relation between volatility and volume, and the power to recover normality. (Chan and Fong (2006), ABFN (2010)).

Previous studies have either considered a single measure on a number of stocks and indices or a number of measures on a single stock and index. However, very few studies discuss the stylized facts of different realized volatility measures comparatively. This chapter aims to fill that gap. In the first chapter, we address the stylized facts of four realized volatility measures across 10 different market sectors. The idea is to check the extent to which such stylized facts vary across sectors. We generalize previous studies by considering all four realized measures using a diversified data set which looks at 100 stocks representing 10 sectors.

Our main contributions can be summarised as follows:

1. Our data set provides more robust results when compared to earlier studies. For example, most studies have used data from the TAQ database. Unlike TAQ, tick data adjusts for stock splits and dividend payments, ignoring which can result in greater variation in the results obtained, especially for some measure such as the realized range.
2. We provide a systematic ranking for 100 stocks in 10 sectors according to returns, realized volatility measures, daily and intraday jumps and recovering return normality. Trading volume and realized measures provide different rankings for stocks in the sectors under consideration. This suggests that the nature of information content of trading volume and realized measures are distinct. For example, and according to the volatility ranking provided by realized variance, we find that "materials" ranks top whereas "consumer staples" ranks last.

3. All realized volatility measures are shown to provide similar ranking for the least active sectors and to show diversified ranking for the more actively traded sectors.
4. The 5 minute is the optimal sampling frequency for all realized measures except in some cases realized power variation. Measures diverge at the 1 second frequency and tend to converge at the 5 minute frequency. The result is independent of both stock activity level and volatility.
5. The stylized facts of realized measures tend to vary across sectors. For example, the positive relation between trading volume and realized measures of volatility is best represented in the "materials" sector and to a lesser extent the in "consumer discretionary" sector. Power and bi-power measures of volatility are found the most capable in recovering normality with their performance varying across sectors.
6. Realized measures are highly correlated across sectors, nevertheless they display different regimes.
7. Jumps are directly related to both level of stock activity and the sampling frequency. Active stocks show fewer jumps and the higher the sampling frequency the higher the number of jumps detected. Volatility and Jumps are not correlated.

This chapter is presented in 12 sections: Section 1 gives a brief introduction of the chapter. Section 2 provides a literature review. In section 3, we provide the theoretical framework of the realized volatility measures. Section 4 describes and discusses the data. Section 5 discusses the descriptive statistics of realized measures of volatility measures and of jumps. In Section 6, Leverage and feedback effect is discussed. Sections 7 and 8 respectively provide the correlations between realized volatility measures and volatility regimes. The long-memory property of realized volatility measures is shown in Section 9. The regression results of volatility- volume relation is given in Section 10. Section 11 summarizes the results of the normality recovery power by different volatility measures and the distributional properties of realized volatility measures. Section 12 provides rankings for sectors by the results of different tests. The final section, Section 13, presents our conclusions.

1.2 Literature Review

Measurement and estimation of volatility has undergone many changes in the last three decades. This is mainly attributable to the development of the theoretical framework, advances in data and computer technology and the growth of financial markets.

The literature defines and estimates volatility in diverse ways. Early studies consider both parametric and non-parametric estimates. Some very early papers consider mainly non-parametric ways, such as price change or absolute price change (see Ying (1966), Clark (1973) for instance). Some later papers also use squared returns (Andersen (1996)) and absolute returns (Andersen (1996)). Another stream in the literature estimates volatility parametrically and semi-parametrically. For instance, stochastic volatility (SV) is developed by Taylor (1986), and by Hull and White (1987). Another important parametric volatility measure is the (G) ARCH families firstly introduced by Engle (1982) and Bollerslev (1986).

However, the stylized facts and properties of volatility are very much data dependent. Earlier non parametric measures were noisy and non-reflective when compared to the standard adopted measures such as stochastic volatility and GARCH (1,1) which were both unbiased and easy to estimate in case of the GARCH. Noisy measures fail to capture information contained in the data and hence are less effective in out-of-sample forecasting exercises. The emergence of high frequency data in the past two decades has made the volatility now "observable", so that it can be modeled directly. Volatility non-parametrically calculated based on high frequency data is now known as realized volatility, for which there are various measurements. Below we aim to give a review of the literature highlighting the stylized facts of high frequency measures.

Realized variance, or realized volatility first appears in the family of realized volatility measures. Realized variance, the sum of intra-day squared returns, is also the most widely used and discussed volatility measure within the high frequency literature. Under weak regularity conditions, realized variance can be constructed for the integrated variance which is asymptomatically unbiased and converges to the true volatility as the sampling frequency tends to infinity (Andersen *et al* (2001), Barndorff-Nielson *et al* (2002)).

Andersen, Bollerslev, Diebold and Labys (ABDL, 2001) analyze the distribution and correlation of realized volatility using a foreign exchange dataset of 10-year DM/USD and Yen/USD returns at 5-minute frequency. Unconditional distributions (univariate and multivariate) and conditional distributions are examined. The distributions of FX realized variance, standard deviations and covariance all exhibit right skewness and leptokurtosis. The normal distribution is rejected. Only the correlations appear to be close to normal. However, the distributions of log standard deviations and correlations are approximately normal. The correlation between the two FX series increases with the realized variance as well. The authors also look at temporal aggregation effects. When the realized volatility measure is temporally aggregated at different return frequencies, realized variance shows strong volatility clustering effects. The persistence in realized variance is evident at the monthly level. Finally, the paper finds that realized variance is stationary, but fractionally integrated and slowly mean-reverting. In a later paper by ABDE (2001), which examines the distributions of returns of 30 stocks, the conclusions about distribution properties and correlations are in line with the previous authors' findings which use foreign exchange rate data. Furthermore, the paper by ABDE confirms that an asymmetric relation exists between stock returns and realized volatility. When returns are negative, the volatility innovations appear to be higher than the innovations associated with positive returns. There also exists a volatility-in-correlation effect of data, which shows the strong positive correlations between stock volatilities and between contemporaneous stock correlations.

ABDL (1999) studies the unconditional and conditional distributions and the correlations of realized volatility over 1,000 days. They conclude that realized volatility changes from day to day and displays substantial persistence. The correlation is always positive and highly correlated with the realized variance. In an attempt to reduce the microstructure effects in the high frequency data, the authors optimize realized volatility and correlation by looking at the sample frequency and then use volatility signature plots to decide the sampling time frequency. For a liquid asset sample and a less liquid asset sample, 20-minute and 15-minute frequency are chosen respectively according to the volatility signature plots. The forecasting of realized volatility and realized correlation is also discussed in the paper.

ABD (2003) conduct additional empirical research on realized volatility forecasting using spot foreign exchange data of USD, DM and JPY, from 1986 to 1999. The authors first review the quadratic variation theory and realized volatility theory under the assumption of frictionless market. The statistical properties of realized volatilities are then summarized. The long memory in realized volatility suggests a long memory Gaussian vector autoregression (VAR) for forecasting. Furthermore, the authors compare the VAR-RV forecasts and the forecasts of other traditional volatility models, such as VAR-ABS (absolute returns), VAR-RV, GARCH, Risk Metrics, daily FI-EGARCH and intraday FI-EGARCH, both at one-day and ten-day time length. In the one-day horizon, the results of one-day-ahead out-of-sample forecasting show that VAR-RV forecasts considerably outperform other volatility measures forecasts. The R-squared of the regressions of VAR-RV are always the highest. VAR-RV still outperforms most of the other volatilities except for intraday FI-EGARCH in the out-of - sample forecasting. However, even FI-EGARCH shows higher R-squared than realized volatility, the null hypothesis of the forecasting model is rejected for FI-EGARCH, but not for VAR-RV. The forecasting results of ten-day horizon are similar: VAR-RV forecasts are the best in most of the cases. Furthermore, VAR-RV also produces very promising density forecasts and associated quintile predictions (Value at Risk, or VaR).

Independently from the studies by ABDL and ABDE, another important theoretical work of realized variance is that of Barndorff-Nielsen and Shephard (BN-S, 2002), which looks at the properties of realized volatility under the content of stochastic volatility model. It analyzes the asymptotic distribution of the so called realized volatility error, which is defined as the difference between realized volatility and integrated volatility. Unlike previous researches, this framework provides model-based estimation of integrated volatility using realized volatility. In addition, it allows the estimation of SV parameters.

Since realized variance is constructed from high frequency data, the presence of microstructure noise is inevitable. A number of studies are focused on the impact of microstructure noise on realized variance. Hansen and Lunde (2006) propose a Newey-West type correction of realized variance that cleans the noise and yields an unbiased RV estimator. Bandi and Russell (2006 &2008) look at the relationship between realized variance, microstructure noise and optimal sampling. Both papers use US equity data and separate the unobservable microstructure noise from the realized variance. Another important finding of the Bandi and Russell's papers is that, in re-examining the optimal sampling frequency first proposed by ABDL (1999), they conclude that 5-minute is an empirically satisfactory frequency.

Even though realized variance is considered to be a more efficient volatility measure comparing with other volatility measures, such as GARCH and squared returns, it has still certain drawbacks, one of which is the presence of jumps in the volatility series. Therefore, BN-S propose two new realized volatility measures, namely realized power variation and realized bipower variation, both of which are robust to jumps.

Realized power variation (PV), which is calculated as sums of absolute powers of increments, is first introduced by BN-S (2003). The paper derives the theorems of realized power variation and also provides empirical applications from Monte Carlo stimulations. The limiting distribution theory presented in the paper further explains the variability of the difference between the realized power variation and the actual power variation. When the logarithm transformation of realized power variation is applied, the QQ-plots show improvement of the normality.

BN-S (2004) extend the volatility measure of realized power variation and derive the theoretical properties of a new estimation, realized bipower variation (BV). BV is calculated from products of powers of absolute returns. Both PV and BV are robust to rare jumps, especially in the case of BV, as the time frequency tends to infinite, RV converges to quadratic variation and BV converges to the integrated variance. Therefore, the quadratic variation of the jump component is the difference between RV and BV. Hence, quadratic variation may be divided into the continuous component of log-prices and the component of jumps. The paper also reviews the probability limit of RV and PV. In the simulation and empirical parts, the theory is tested and the results confirm that RV and BV may be combined to estimate the jumps as the probability limit of BV is unaffected even in the presence of jumps in stochastic volatility model.

Another paper also by BN-S (2006) uses BV to test jumps in the high frequency financial time series. Recalling BN-S (2004), RV can in theory decompose the components of quadratic variation into jumps part and continuous part of log-prices. In that case, BV can be consistently estimated. The paper also derives the asymptotic distribution theory for nonparametric tests of jumps under very weak conditions. The test is applied both to simulated data and to real foreign exchange data over 10 years. Given the null hypothesis of no jumps, the simulation experiment suggests the rejection is heavily influenced by the variance of jumps, rather than the frequency or size of the jumps. It also shows that an adjusted ratio jump statistic can be used to test jumps where the intraday time period is reasonably small and the sample period is small (for instance one day). The test outcome of FX data fails to accept the null that there is no jump in the data. The rejection of no jump is attributed to the breaking macroeconomic news in some studies.

Christensen and Podolskij (2007) derive the theoretical properties of another realized volatility measure using high and low price range, which they call realized range-based variance (later known as realized range).The main contributions of this paper include: the theoretical framework of realized range (RR) estimation is built. It is proved to be consistent with the integrated variance and to be less noisy than realized variance. Several probabilistic laws for sampling intra-day high-low price ranges are derived. The downward bias, reported in a number of daily range papers, is also adjusted and removed by introducing a new scaling factor. In the empirical part, the authors conduct both Monte a Carlo experiment and an empirical test for 4-year data for General Motors. Monte Carlo examines the normal and log-normal distribution of realized range. Both distributional results are consistent with the CLM while the log-normal outperforms. In respect of comparisons between realized range and realized variance, the mean and variance of RR are lower than for RV. RR is less skewed, more persistent and shows lower kurtosis. Moreover, the test results confirm that RV has lower sampling errors than RV. For their general conclusion, the authors claim RR to be a less volatile and more efficient volatility estimator than RV.

Martens and Dijk (2007) further investigate the properties of RR. The paper tests the realized range and uses both Monte Carlo simulation and empirical data from S&P500 index-futures and individual stocks from the S&P100 index. Before the simulation and empirical tests, the authors conduct a bias correcting procedure for realized range and realized variance: realized range is more affected by the market microstructure as it is constructed from high-low prices. Two methods are discussed. The first is based on the derivation of expressions for the expected difference between the observed and unobserved high-low prices. The second method replaces the scaling factor *4log2* by the expected value of the squared range of a Brownian motion which is the number of observations during the ith intra-day interval. However, both methods are less adequate to deal with the upward bias of the realized range due to bid-ask bound, the authors alternatively correct the bias by scaling the realized range with the ratio of the average level of the daily price range and of the price range over the q previous trading days, given q as long as possible under certain conditions. The main findings of the comparisons between realized range and realized variance include: the Monte Carlo simulation is consistent with the theory that the realized range also converges to the integrated variance and is more efficient (5-times according to the empirical result) than realized variance; and that the realized range has a lower mean-squared error. At the same time frequency, both realized range and realized variance are upward biased when bid-ask bound exists. However, whereas realized range is downwards biased where there is infrequent trading, realized variance is unaffected. Empirical results using S&P500 data reach similar conclusions to the Monte Carlo simulations. They show that realized range significantly improves the Two Time Scale estimator, which is also viewed as a volatility measure. When S&P100 data are used, the results are more ambiguous but still confirm realized range a more efficient than realized variance at the 5 and 30 minute frequency. For their general conclusion, the authors claim the realized range to be a better volatility measure than the realized variance.

In addition to literature mentioned above, there are many other empirical studies in the area of high frequency finance, especially realized volatility measures. As the ABD (2003) paper compares realized variance with other parametric volatility measures, other papers comprise the comparative studies, in particular between different realized volatility measures.

Ghysels *et al* (2006) use Mixed Data Sampling (MIDAS) regressions for comparisons between different measures of volatility and volatility forecasting. MIDAS, defied as a reduced-form forecasting device of realized volatility, is valid for both in-sample fit and out-of-sample forecasting. It also allows parsimoniously parameterized regression of data at different time frequencies. Several volatility measures are used in the paper: squared returns, absolute returns, realized volatility, realized range and realized power variation. The authors first examine the forecasting power of the MIDAS of different volatility measures using both daily and 5-minute frequency data of 10-year Dow Jones Composite Portfolio and Dow Jones Index. The forecasting horizon ranges between 1day and 1-4 weeks. For both in-sample and out-sample cases, realized power variation is the best predictor of future realized volatility. The second best would be the realized range while the squared return performs most poorly. MIDAS regression is also examined by high frequency data (at 5 minute). The data is seasonally adjusted before the test. The results are very similar to those using daily data. The realized power variation remains to be the best performed volatility measure. The authors find that the use of high frequency data in the MIDAS does not directly improve the forecasting performance. It has very similar outcomes to daily frequency prediction. In the conclusion, the paper highlights some possible explanations for the remarkable forecasting power of realized power variation.

Fuertes, Izzeldin and Kalotychou (2009) compare the forecasting gains in GARCH (1, 1) when augmented by realized measures of volatility. Using 14 NYSE equity stocks for the period 1997-2003, they show that among different realized volatility estimators, realized range works best, according to the normality recovery by realized volatilities standardized returns. One-day-ahead forecasting is undertaken in the framework of GARCH and augmented GARCH models and judged by different loss functions. The realized power variation outperforms other estimators as it improves the forecasting power of GARCH by most after augmented in the GARCH model. When realized variance, realized range, realized power variation and realized bipower variation are combined in the forecasting model, forecast errors appear to be the smallest for almost half the sample data. This paper also takes trading volume into consideration and finds that, when trading volume is relatively low or the market is over-performing in day *t-1*, the volatility forecasts using data from day *t-1* to day *t* will more accurate outcomes.

Brownlees and Gallo (2010) conduct a comparison of volatility measures by forecasting Value-at-Risk (VaR). They use 4 blue chip companies in the US market to construct the realized volatility measures, namely realized volatility, realized bipower volatility, two-scaled realized volatility, and realized kernel and daily range, which is the only daily volatility measure. Both in-sample modeling and out-of-sample forecasting confirm that realized volatility measures (as well as daily range) are more efficient than absolute or squared returns. The results of in-sample modeling show that realized kernel outperforms other measures and provides the most accurate estimation of the variance of returns. The two-scaled realized volatility performs second best. Realized volatility provides better estimation than realized bipower volatility. In the cast of out-of-sample VaR forecasting, realized kernel again outperforms the rest. However, daily range also performs very close to realized kernel. The authors attribute the outstanding performance of realized kernel both in-sample and out-of sample to the fact that it is the most robust to market microstructure noise. Another interesting finding in this paper is that the means of RV and BV are substantially constant across sampling frequencies in excess of 30 seconds.

Fleming and Paye (2011) investigate the mixture of distribution hypothesis (MDH) by comparing the standardizing power of realized variance, realized bipower variation and realized kernel on daily returns. They use trade and quote US equity data of the most 20 heavily traded stocks in NYSE. The normality of returns standardized by realized variance is all rejected. The returns standardized by realized kernel are platykurtotic. However, the returns standardized by realized bipower variation show a great improvement in normality. As realized bipower variation is the only volatility estimator that is robust to jumps, Fleming and Paye (2011) conclude that the presence of jumps in the realized volatility violates the continuous price paths assumption and thus leads to the failure of standardization by realized variance. As long as the jump component is removed (for instance using BV), normality of standardized returns can be attained.

The realized volatility measures are found to be more efficient than the parametric counterpart. Nevertheless, empirical results based on realized volatility measures still appear to be diverged from what theoretical models suggest. Consequently, there has been a growing literature in modeling the jump component contained in RV. According to probability theory, quadratic variation can be divided into a continuous component (integrated variance), and a discrete, or jump component. Empirically, the detection of jumps relies on the RV, which is a proxy of quadratic variation and BV, a proxy of integrated variance. Huang and Tauchen (2005) propose a daily jump detection method which identifies the presence of jumps by testing the significance difference between RV and BV. Following Huang and Tauchen (2005) method, a series of studies attempt to ascertain the empirical performance of realized volatility measures which exclude jumps.

ABD (2007a) model and forecast RV taking jumps into consideration. The empirical work is based on the theoretical results of BN-S (2004, 2006). 5-minute returns for DM/USD foreign exchange market from December 1986 to June 1999, S&P 500 market index and 30-year US Treasury yields from January 1990 through December 2002, are used. The models implemented are known as HAR-RV-CJ forecasting models, which are based on earlier HAR-RV and HAR-RV-J models. The authors test the models both linearly and nonlinearly and find that the HAR-RV-CJ models eliminate most of the strong autocorrelation in the realized volatility series. Even jumps are more predictable, only the continuous sample paths have the forecasting power when both components are included in the model. Separating the jump components from the continuous sample paths significantly improves the out-of-sample forecasting. In the final part of the paper, the authors suggest some possible extensions, addressing the issue of integrated volatility forecasting in the presence of jumps.

Another recent paper concerning the effects of jumps in financial data is that of Andersen, Bollerslev, Frederiksen and Nileson (2010) (ABFN). It works on the distributional prosperities of daily returns and realized volatility in the presence of jump components, using individual stocks from the DJIA Index. The authors suggest a new sequential jump detection method which could identify multiple jumps over the same day. Together with the single jump detection method, leverage effects and feedback effects, daily return distributions (both unconditional and standardized) are tested. The test results confirm some earlier stylized facts that equity returns are rightly skewed and have tails. The normality hypothesis of GARCH standardized returns is also rejected by all 30 stocks, while the normality of RV standardized returns shows significant improvement compared to GARCH standardized. The authors also find that jump adjustments, both single and sequential, do not necessarily enhance the normality of stock returns. Nevertheless, when returns are standardized by realized volatilities that are not only adjusted by jumps, but also by event-time, or financial time sampling, the normality is dramatically restored. The authors claim that this result confirms that "*inter-daily stock prices may usefully be thought of as discretely sampled observations from an underlying continuous-time jump-diffusion model, but it is essential to also accommodate leverage and/or volatility feedback effects*".

The discussion of the related literature in this section is inevitably partial. We consider empirical studies most closely to ours and of significant interests in empirical research. There are numerous studies which cover every research area in the high frequency finance and realized volatility measures. In the next section, we discuss the theoretical framework of the realized measures of volatility that will be used in our empirical research.

1.3 Theoretical Framework

The construction of realized volatility measures is based on jump-diffusion process. Jump diffusion models model the asset price as a mixture of a continuous diffusion path and an occasional discontinuous jump path. The incorporation of jumps dates back to Merton (1976). Recent empirical evidence by Andersen *et al* (2011), Tauchen and Zhou (2011), and Dobrev and Szerszen (2010) all gave support to a jump-diffusion specification.

Assume price process P_t, which is semi-martingale, follows a geometric Brownian Motion,

$$\mathrm{d}p_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}B_t + c_t \mathrm{d}q_t \quad [1.1]$$

where μ_t denotes a continuous and locally bounded process, σ is the constant volatility parameter, B_t denotes a standard Brownian Motion, $\mathrm{d}q_t$ is the counting process with jump intensity λ_t and c_t the size of the corresponding jumps. Suppose that, in a given trading day, *t*, a set of *j* = M+1 intraday prices are available at equally spaced intervals of $\Delta = 1/M$. Denote *j*th intra-daily log-price for day *t* by $p_{t,j}$, where $j = 0,1,...,M$ and $t = 1,...,T$ .Then the M continuously compounded intra-day return for day *t* can be expressed as

$$r_{t,j} = p_{t,j} - p_{t,j-1}, \qquad j = 1,...,M,\ t = 1,...,T \quad [1.2]$$

1.3.1 Realized Variance (rv_t)

Realized variance is defined as the sums of squared intraday returns. (ABDL, 2001, BN-S, 2002). Mathematically, rv_t for day *t* is expressed as

$$rv_t = \hat{\sigma}^2_{t,N} = \sum_{j=1}^{M} r^2_{t,j} \quad t = 1,2,...T \quad [1.3]$$

The daily returns r_t which are calculated from the sum of M intraday returns $r_{t,j}$ are defined by

$$r_t = \sum_{j=1}^{M} r_{t,j}\ ,\ \mathrm{M} = 1, 2, 3.... \qquad [1.4]$$

It directly follows from the quadratic variation theory that rv_t converges uniformly in probability to Quadratic Variation (QV) when the sampling frequency increases. In a frictionless world without jumps, realized variance should converge for the Integrated Variance (IV).

According to ABDL (2001), rv_t is unbiased, consistent, highly efficient, yet jump-contained. It is also the most intensively studied realized volatility measure in the literature.

1.3.2 Realized Power Variation (pv_t)

Realized Power Variation was introduced by Barndorff-Nielsen and Shephard (2003, 2004) and is written as

$$pv_t(p) = \mu_p^{-1} \frac{t}{M}^{1-\frac{p}{2}} \sum_{j=1}^{M} \left| r_{t,j} \right|^p,\ \mu = \frac{1}{t}, and\ t = 1,2,\ldots T \qquad [1.5]$$

$$\mu_p = E|\mu|^p = 2^{\frac{p}{2}} \frac{\Gamma(0.5(\mathrm{p}+1))}{\Gamma(0.5)}, p > 0, \mu \sim N(0,1)$$

When $t \to \infty$,

$$pv_t(p) \xrightarrow{p} \begin{cases} \int_0^t \sigma^p(s)ds, & 0 < p < 2 \\ QV_t, & p = 2 \\ \infty, & p > 2 \end{cases} \quad [1.6]$$

Deciding the value of p is important. When$0 < p < 2$, realized power variation is robust to jumps and converges to the integrated variance. In the case of $p = 2$, the realized power variation becomes rv_t; and when $p = 1$, it is termed as absolute variation. In line with the literature, we use a value of p =1.5 which yields the lowest RMSE, according to Liu and Maheu (2005).

1.3.3 Realized Bipower Variation (bv_t)

Barndorff-Nielsen and Shephard (2004, 2006) define the Realized Bipower Variation as

$$bv_t = \mu_1^{-2} \sum_{j=2}^{M} |r_{t,j}||r_{t,j-1}| \quad [1.7]$$

$$\mu_1 = E(|\mu|) = \sqrt{2/\pi}, \mu \sim N(0,1)$$

bv_t is independent of assumptions concerning the distribution of the jumps or the relationship between the jump process and the stochastic volatility component. BN-S states that bv_t could be used to estimate both continuous and discontinuous components of quadratic variation (QV), which is given as,

$$qv_t = \int_{t-1}^{t} \sigma^2(s)\mathrm{d}s + \sum_{t-1<s\leq t, dq(s)=1} c_s^2 \quad [1.8]$$

When the time frequency becomes infinitely dense, the rv_t then converges to the quadratic variation and the realized bipower variation converges to the integrated variance.

$$rv_t \rightarrow qv_t \qquad [1.9]$$

$$bv_t \rightarrow \int_{t-1}^{t} \sigma_t^2(s)\mathrm{d}s \rightarrow iv_i \qquad [1.10]$$

Therefore, the jump component may be separately calculated as

$$J_t = rv_t - bv_t \qquad [1.11]$$

1.3.4 Realized Range (rr_t)

Christensen and Podolskij (2007) derive the Realized Range (rr_t) from earlier works of Parkinson (1980) which focuses on the high and low prices of stocks.

$$rr_t = \frac{1}{4log^2}\left[\sum_{j=1}^{M} 100 \times \left(\log\left(p_{t,j}^h\right) - \log\left(p_{t,j}^l\right)\right)^2\right], t \; = \; 1,2,3 \; ...\mathrm{T} \qquad [1.12]$$

Where $\left(p_{t,j}^h\right)$ are high prices and $(p_{t,j}^l)$ are low prices in the jth interval respectively, and $4log^2$ is a scaling factor that is used to correct biases of market microstructure effects such as bid-ask bounce due to second moment of the range of a standard Brownian Motion, B_t, that $E(s_B^2) = 4log2, where\ s_B = sup_{0\le t,s\le 1}(B_t - B_s)$.

In a world that is absent of jumps and microstructure noise, realized range converges to the integrated variance. Previous studies (Christensen and Podolskij (2007), Martens and van Dijk (2007)) suggest that in a frictionless world, realized range is 5 times more efficient than the realized variance and converges to the integrated variance. Martens and van Dijk (2007) conduct Monte-Carlo simulations and conclude that realized range is better than realized variance only under the same sample frequency. However, in the presence of microstructure frictions, infrequent trading leads to a downward bias in realized range alone, but both realized range and realized variance show upward bias.

1.4 Data

We consider transaction data of 100 US traded stocks from 10 sectors. The time period for the data is from 02/01/2000 to 31/12/2010, a total of 2767 trading days. Our selection criteria are made on the basis of market capitalisation and wide coverage sector representation whereby the following sectors are considered: Consumer Discretionary (CD), Consumer Staples (CS), Energy (ENG), Financials (FIN), Health Care (HC), Industrials (IND), Information Technology (IT), Materials (MAR), Telecommunications (TEL) and Utilities (UTL) according to the category from S&P 500. Different sectors present different degrees of volatility and liquidity. For instance, the IT sector is a heavily traded sector which includes very active companies such as CSCO, INTC, MSFT and ORCL. Materials, Telecommunications and Utilities sectors are much less actively traded.

Our 100 stocks consist mostly of DJIA stocks and of S&P 100 and S&P 500 stocks, with a few exceptions. We exclude KFT from CS sector and add BT, VOD and TEF in TEL sector for data consistency and full time coverage. In general, for each sector, we combine some of the largest market capitalization companies and median to large size companies within the sector. Each sector is ranked according the sector average trading volume. In our sample, IT is the most actively traded sector whereas UTL is the least.

[Table 1.1 of full list companies here]

All data are from Tick Data, which is sourced from the NYSE's TAQ (Trade and Quote) database. TAQ records intraday transaction data for all securities listed on the NYSE, AMEX and NASDAQ. A paper by Brownlees and Gallo (2006) discusses high frequency data handling concerns. The authors state that the TAQ data does not guarantee the accuracy as the NYSE itself does not. The database contains delayed and incorrect recordings, hence produces errors and bias.

The Tick Data are adjusted, cleaned and managed from the TAQ database using the following process: Ticker Mapping adjusts historical data for corporate actions such as M&A and symbol changes etc. Condition Code Filtering is a process whereby trade and quote data are filtered for various condition codes such as out of sequence trade and quote, cancelled trades and other conditions which require prior removal of data points. Price Filtering filters flag trades that are bad ticks and suggest corrected values; Data Validation uses third party data to ensure the accuracy of previous 3 processes. The final adjustment is to generate stock splits and cash dividends data and to allow the application of splits and dividend adjusted high frequency data, which is especially important for calculating realized range.

To show the advantage of using TICK instead of TAQ, **Table 1.2** presents a comparison of summary statistics of realized volatility measures of INTC calculated from TAQ and TICK database within the same time period.

[Table 1.2 here]

The summary statistics of volatility measures calculated from tick data appear to be more stable and to present less variation. The mean, maximum and minimum values are all lower than counterpart calculated from TAQ data, especially for realized range. The skewness, kurtosis, and JB statistic further suggest higher normality of returns of Tick data. This could be further evidence of a more bias-free dataset of Tick.

Finally, the time period we chose ranges from 2000 to 2010, which covers 11 years. The length also allows us to investigate the impact of the 2001 internet bubble crisis and the 2008-2009 financial crisis. From September to October 2008, volatility and trading activities appear to be very high. This is the period when US financial institutes were hit most severely by the crisis. Allowing the presence of crisis periods also enables us to test whether the crisis may override some of the stylized facts of the volatility measures.

1.4.1 Optimal Sampling Frequency

In this section, the aim is to identify the optimal sampling frequency across different volatility measures, by conducting the volatility signature plots (VSP) first proposed by ABDL (1999).

Finding an optimal aggregation level of the tick by tick data is vital to empirical research. Tick by tick data are less readily directly modelled, as they contain a high level of market microstructure noise. Market microstructure noise leads to the autocorrelation of intraday returns. The autocorrelation increases with frequency is higher and causes further bias to rv_t.

Hansen and Lund (2004) cite four reasons why rv_t might be biased due to the microstructure noise. First of all, lack of liquidity may cause the observed price to be different from the true price. Second, bid and ask spread, as well as the discrete nature of price data, could have rounding errors. Third, econometric methods which are used to construct artificial price data may lead to pricing errors. This is especially relevant to the construction of prices when no actual trading happen at a given price point. The final reason of pricing error is due to the quality of the data used. For instance, mis-recorded prices induce market microstructure noise inevitably. Not only do these four reasons explain the difference between volatility measures calculated form TICK and TAQ databases, but also show the inadequacy of using tick by tick data to construct the daily nonparametric volatility measures.

In summary, an optimal sampling frequency is needed to balance both the bias associated with high sampling frequency and the lost information due to the low sampling frequency.

The VSP is calculated as the "average realized variance against sampling frequency". (ABDL, 1999).

$$\overline{RV}^{(m)} = n^{-1}\sum_{t=1}^{n} RV_t^{(m)} \qquad [1.13]$$

Where m is the sampling frequency and n is the number of periods (days).

The 1999 paper recommends sampling until the point at which microstructure noise starts to be absorbed by the realized variance.

Several studies, (see Hansen and Lunde (2004), Bandi and Russell (2008), as well as Shephard (2010),) all follow the above method to determine the optimal sampling frequency using various data sources. Conclusions differ from one paper to another. For instance, using FX data, the pioneering paper by ABDL (1999) finds that the optimal frequency for rv_t should be 20 minute. Bandi and Russell (2008) find 5-minute to be a satisfactory frequency for IBM quote data. Although focusing on microstructure noise, Hansen and Lunde (2004), Fleming and Paye (2007), argue that, even at 5-minute frequency, there is an upward bias for rv_t. Both studies use US equity trade data. Andersen *et al* (2010) revisits the VSP by calculating both rv_t and bv_t at an expanded sampling interval span and database. According to the 2010 paper, VSPs should exhibit a decreasing tendency and will be flatter and become relatively constant after certain aggregation frequency as the overwhelming microstructure frictions at tick by tick level are gradually balanced by the effect of the aggregation process. Their finding that 5 min could be used as the optimal sampling frequency is based on results from 30 DJIA stocks.

In addition to the various data sources mentioned above, different "benchmark" volatility measures, which are assumed to be the least autocorrelated and therefore least biased, are also used. Andersen *et al* (1999) and Hansen and Lunde (2004) use rv_t aggregated at 30-minute, Fleming and Paye (2007) uses Newey-West rv_t and ABFN (2010) use bv_t aggregated at 30-minute.

However, most studies merely consider the optimal sampling frequency of rv_t (or modified rv_t) and leave other nonparametric volatility measures not investigated. Since the objective of this chapter is a comparison study of nonparametric volatility measures, the optimal sampling frequency of bv_t, pv_t and rr_t is also of interest. We will follow closely the method used by AFBN (2010) to construct the VSPs of all four volatility measures. We illustrate this issue by looking at two specific stocks in our sample[1].

Figure 1.1 shows the volatility signature plots for realized variance, realized range and realized power and bipower variations. The benchmark shown as the horizontal line is the realized bipower variation aggregated at 30-min frequency of FTR from TEL sector and of GS from FIN sector. We select these two stocks based on the number of daily jumps detected. FTR contains the highest number of jumps while GS contains the lowest.

[Figure 1.1 here]

In like fashion to ABFN (2010), we report the VSP from the frequency of 1 second to 1800 seconds (30 min equivalent). The VSPs show a number of interesting findings. First of all, all realized measures of volatility exhibit a gradual decreasing trend, with the exception of rr_t, which shows a weakly increasing trend. This result could be explained by the fact that rr_t is the only realized volatility measure calculated based on price high and low. When the sampling frequency is ultra-high, there might be little the price change within the sampling frequency and therefore the difference of highest and lowest prices is less obvious as compared with rr_t calculated at lower sampling

[1] The VSPs from other stocks in our sample reach the same conclusion. We report these two stocks due to the space limit and the representation they bare regarding to the number of jumps contained.

frequency. pv_t is the only measure which exhibits a constantly decreasing tendency even after 5 minutes. However, pv_t does not apply the same measurement units as other realized measures of volatility. This explains why it does not reach a stable level after 5 minutes. The plots of rv_t, bv_t and pv_t show maximum values at 1 second. rv_t and bv_t become satisfactorily stable after 5 minutes. The figure also illustrates a narrowing then a relatively constant difference between rv_t and bv_t. This difference also shows a direct relation between the sampling frequency and the level of jumps presented in rv_t. This finding is in line with ABFN (2010) which attributes the gap to the influence of the jump components. The larger gap between the two measures when the aggregate frequency is high is more likely to be attributed to the significant market noise (ABFN, 2010).

In general, the VSPs confirm that 5 minute frequency is an optimum level for rv_t and bv_t as these two measures start to stabilize after the 5 minute sampling aggregation level. It is suboptimal for pv_t, because pv_t seems to follow a decreasing trend with the aggregation frequency. For the stock with highest number of jumps, the continuous decreasing pattern is more obvious. rr_t, although has weakly upward trend, converges to stable value after 5 minutes. Thus 5 minute should be a reasonable sampling for rr_t.

To conclude, 5-minute sampling frequency will be used to construct our realized measures of volatility. Not only it is considered as the best trade-off between information accuracy and microstructure noise, but also it gives the comparison benchmark of different realized volatility measures.

Hence, all our data is aggregated at 5-minute interval every trading day from 9:30 to 16:00, 6.5 hours and 78 intraday periods within one trading day in total. 9:30 – 16:00 is the trading hours in NYSE and NASDAQ and is also the most frequently used in the literature. As the main task of this chapter is to disucss the stylized facts of different volatility measures and to compare them with those reported in the previous studies, we also follow this trading hour and do not consider the transactions before and after the market trading hours.

1.4.2 Constructing Daily Returns

In the previous section, we demonstrated that 5-minute is a generally accepted optimal sampling frequency for both intraday return and realized measures of volatility. **Table 1.3** reports the mean summary statistics of daily returns across sectors. Among all the sectors, the CS sector yields the highest average return (0.5%) and the MAR sector yields the lowest (-0.88%). At the individual stock level, all returns are approximately zero. Returns are not normally distributed, exhibiting excess kurtosis, which suggests fat tails. The JB statistics suggest that the return series of UTL is most distorted from the normal distribution. Returns of IND have the lowest JB statistics, which is still as high as 2738. All the properties on daily returns are in line with the stylized facts of stock returns.

[Table 1.3 here]

[Figure 1.2 here]

1.5 Descriptive Statistics of Realized Volatility Measures and Jumps

1.5.1 Realized Volatility Measures

Table 1.4a & 1.4b report the summary statistics of the daily realized measures of volatility, by sector. The outcome is in line with many stylized facts of the volatility measures driven by latent information flow from previous studies such as ABDL (1999, 2001), ABDE (2001) ABD (2003,2007), but not with the studies on rr_t (Martens and van Dijk, 2007).

[Table 1.4a and 1.4b here]

Although previous studies conclude that rr_t is more efficient than rv_t, our sample reaches slightly different conclusions. The overall average rr_t has a lower mean but a similar standard deviation to rv_t. On the other hand, the skewness and kurtosis of rr_t are higher. rr_t has the highest kurtosis, 162.5, among all realized volatility measures. This also makes rr_t appear the highest JB statistics. Furthermore, we compare these two measures at sector average level and individual stock level. IT, the most actively traded sector in our sample, has a higher rr_t than rv_t (rv_t: 5.722, rr_t: 6.046). The standard deviation, skewness and kurtosis statistics of rr_t are also higher. The standard deviation, skewness and kurtosis statistics of rr_t of the least actively traded sector UTL are lower than rv_t (rv_t: 3.419, rr_t: 2.836). Besides, least actively traded stocks tend to have lower mean of rr_t. OKE, the least actively traded stock, has rv_t of 3.136 and rr_t of 2.279 for instance. This result is in line with Martens and van Dijk (2007) who claim that infrequent trading leads to a downward bias to rr_t only. On the other hand, the most actively traded stocks, C, CSCO, INTC and MSFT all exhibit higher means of rr_t that rv_t[2]. There are several possible reasons for the contrasting finding in rr_t and rv_t between earlier papers and this chapter: Our sample contains two crises, the 2001 internet bubble crisis and 2008 financial crisis. During these two periods, there are more likely to have extreme prices. Since rr_t is constructed from intraday highest and lowest prices, it is more affected by the extreme prices and hence shows higher mean values and standard deviation. During a relatively calm period when both trading volume and volatility are low (2003 - 2006), rr_t usually has a lower mean than rv_t. Second, rr_t and rv_t converge to QV according to the quadratic variation theory and contain jumps. The level of jumps and microstructure noise contained in both measures will also inflate the values of the measures. rr_t is more affected by microstructure noise than rv_t (therefore it contains a scale to adjust the microstructure noise) and this is especially true for the stocks that are more actively traded. Most of stocks in our sample are from S&P 100 stocks and are sufficiently actively traded. Hence, in our sample, we find that rv_t a more efficient volatility measures than rr_t.

[2] C:rv_t:8.720, rr_t: 9.555; CSCO: rv_t:5.706, rr_t: 6.792; INTC: rv_t:5.359, rr_t:5.874; MSFT: rv_t:1.878, rr_t:2.236.

The mean of pv_t is higher than that of both rv_t and bv_t. However, pv_t does not apply the same measurement units as other realized measures of volatility. Realized power variation has the lowest skewness and kurtosis even though it has a higher mean value than the other three measures. bv_t has the second lowest mean and lowest standard deviation on average, suggesting the most consistent volatility estimator among all.

All four realized measures of volatility are severely right skewed as well as leptokurtic. Realized variance, realized power variation and realized bipower variation all show extremely strong serial correlations, with the realized power variation the strongest, as indicated the Ljung-Box statistics and Autocorrelation Function with up to 36 lags. We also compute the LB statistics of R^2, the squared daily return. The result shows R^2 to be serial correlated but the correlation is generally weaker than for the other realized measures. ABDL (2003) claim that the lower LB statistics of squared returns in relation to the LB of realized volatility suggests squared return is a very noisy volatility measure as the strong persistence in the latent volatility dynamics is erased by the noises in the volatility measures. The low serial correlation in the rr_t of some stocks can be attributed to the volatility construction method which is more exposed to the noise as due to the intraday highest and lowest prices.

All the volatility measures show similar pattern to each other (**Figure 1.3**). There are two peaks: the first is for 2001-2002, (the dotcom bubble), and the second is for 2008-2009 (the recent financial crisis). Comparing the volatility plots with the return plots in **Figure 1.2**, we find that high volatility is associated with large positive/negative returns; and that the high volatility days cluster. This is another well observed fact in the literature.

[Figure 1.3 here]

1.5.2 Jumps

In this section, we discuss the non-continuous component of realized variance; this is jumps. The presence of jumps draws a great deal of interests in the literature and is well observed with various financial dataset. Dated back to 1976, Merton states that *"since empirical studies of price series tend to show far too many outliers for a simple, constant-variance lognormal distribution, there is a 'prima facie' case for the existence of jump"*. However, only until the availability of high frequency data, the visual confirmation and stylized properties of jumps are better examined. The jumps are usually small and represent the uncertainty of the underlying financial market. As jumps may account for a significant proportion of sum of square intraday return, including them lead to the bias of realized variance. The occurrence of (large) jumps is mainly attributed to the unexpected macroeconomic news, abnormal trades, recording errors as well as the shocks in the asset liquidity. Eraker *et al* (2003) state that jumps should command relatively larger risk premia than the continuous variance because the contribution of jumps to period of market is greater. The risk arising from jumps also cannot typically be hedged away. Therefore, identifying jumps has practical implications for risk management and derivatives hedging. ABD (2007a) show that the continuous path and jump path of the prices are distinct to each other and hence should be modelled separately. In the same paper, they also show that excluding jumps from realized variance enhance both the in-sample estimation and the forecasting power of realized variance.

The possibility of separating jumps from rv_t provides new grounds for analyzing the properties of realized variance with/without jumps and the properties of jumps alone. Theoretically, the jump component should be strictly non-positive, as **Equation [1.8]** suggested. Nevertheless, ABD (2007a) note, "*nothing prevents the estimates of the squared jumps ... from becoming negative in a given finite sample*". In order to comply with the theory, different empirical studies apply different jump construction methods. The jumps constructed according different methods will naturally have different values from one to another. Here we follow the method by Bollerslev *et al* (2009) which accommodates the presence of both small and negative jumps.

$$J_t = \ln(rv_t) - \ln(bv_t) \qquad [1.14]$$

$$cv_t = rv_t - J_t \qquad [1.15]$$

According to Bollerslev *et al* (2009), this avoids "*the arbitrary choice of any pre-specified significance level affecting the selection of 'significant' jumps*". Wang and Huang (2012) also adopt this method to construct the jumps series to investigate the volatility-volume relationship.

In addition, we consider two nonparametric jump tests to detect the non-negative and significant jumps. There are (1) the *MaxZ* test of Huang and Tauchen (2005) for daily jumps and (2) the Lee and Mykland Test (2008) for intraday jumps. The *MaxZ* test builds on the asymptotic distribution theory of BN-S (2004, 2006) and the empirical evidence from Huang and Tauchen (2005). The test assesses the significance of the daily jump component according to the logarithmic test statistic given by

$$\max Z_t = \sqrt{M}\frac{\ln rv_t - \ln bv_t}{[(\mu_1^{-4} + 2\mu_1^{-2} - 5)\max(1, TQ_t bv_t^{-2})]^{1/2}} \rightarrow N(0,1) \quad [3] \qquad [1.16]$$

The intraday test of Lee and Mykland (2008) tests whether a given intraday return $r_{t,j}$ comes from a diffusion or a jump process. The statistic for the LEM test is given by

$$J_{t,j} = \frac{|r_{t,j}|}{\hat{\sigma}_{t,j}} \qquad [1.17]$$

where $\hat{\sigma}_{t,j}$ is an estimate of the local standard deviation and is usually replaced by $\hat{s}_{t,j} = \sqrt{\frac{1}{M}bv_t}$. Lee and Mykland (2008) show that the sample maximum of the absolute value of the J test follows a Gumbel distribution. The original Lee and Mykland (2008) test statistic does not allow for the periodicity that is usually encountered in high frequency data. For such purpose we adopt a modified version of the test (see Boudt *et al*, 2008) which replaces $\hat{\sigma}_{t,j}$ by $\hat{f}_{t,j}^{WSD}\hat{s}_{t,j}$, $\hat{f}_{t,j}^{WSD} = \frac{WSD_{t,j}}{(\frac{1}{M}\sum_{j=1}^{1} WSD_{t,j}^2)^{1/2}}$ where WSD stands for Weighted Standard Deviation filter, with $WSD_{t,j} = \sqrt{1.081 \times \frac{\sum_{l=1}^{n_{t,j}} \omega[(\overline{r}_{l;t,j} / \hat{f}_{t,j}^{ShortH})^2 \overline{r}_{l;t,j}^2}{\sum_{l=1}^{n_{t,j}} \omega[(\overline{r}_{l;t,j} / \hat{f}_{t,j}^{ShortH})^2]}}$. The threshold 6.635 equals the 99% quartile of the χ^2 distribution with 1 degree of freedom. The Weighted Standard Deviation in (17) has a 69% efficiency under the normality of the $J_{t,j}$'s. See Boudt *et al* (2008) for further details.

We consider the top 5% of large jumps at the 5% significance level. We also compute significant jumps at top 0.1%, 1% and 10%, for which the conclusion remains the same.

---

[3] $TQ_t = \frac{1}{M}\mu_{4/3}^3 \sum_{j=3}^{M} |r_{t,j}|^{4/3} |r_{t,j-1}|^{4/3} |r_{t,j-2}|^{4/3}$, $t = 1,\ldots,T$, $\mu_{4/3}^3 = 2^{2/3}\Gamma(7/6)/\Gamma(1/2)$ and $\Gamma(\cdot)$ is the gamma function.

Summary statistics of daily jump series is reported in **Table 1.5**. Average continuous variances of each sector for the overall sample all have slightly lower values than average rv_t. The same finding applies to the skewness and kurtosis. The maximum average rv_t shows great similarity to the average maximum cv_t, indicating that extremely volatile days are not necessarily associated with large jumps ($\max(rv_t - cv_t) < \max J_t$). A similar finding is also reported by Ané and Metáis (2010). The distributional properties of jumps are also very different from both rv_t and cv_t. Moreover, as **Figure 1.3** shows, changes in the level of rv_t are not being matched by the changes in the level of jumps. The clustering feature observed in the rv_t (and bv_t as well as cv_t) does not appear in jumps.

[Table 1.5 here]

Table 1.5 summarizes average daily and intraday jump intensity and the daily jump contribution to volatility at the 5% level. The detected intraday jumps are much less than daily jumps, not only in terms of numbers but also in terms of magnitude. This result holds across all sectors. For the overall sample, there are 124 intraday jumps detected on average while the number of daily jumps detected is 698. The average proportion of detected jumps is 0.06% for intraday level and 26.84% for daily level. This is in line with the findings of Eraker *et al* (2003) that intraday jumps are rarer than the daily jumps. The TEL sector records the highest number of detected jumps [intraday (153) - daily (823)] whereas the IT sector records the lowest [intraday (72) - daily (634)]. Active sectors/stocks feature fewer jumps. The ranking of intraday and daily jumps coincides for TEL, ENG and IT sectors. Furthermore, although average trading volumes vary significantly from one sector to another, the number and proportion of jumps detected do not appear to be largely differentiated. Eraker *et al* (2003) also find that intraday and daily jumps are distinct in nature which our rank results appear to confirm.

Along with the intensity of jump occurrences, the proportion of total realized variance that can be explained by the daily jumps is reported in **Table 1.6.** We also calculated the summary statistics for $J_{t,5\%} / rv_t$. The largest 5% of daily jumps on average contribute approximately 25% of rv_t . This proportion is fairly constant across 10 sectors. The standard deviation ranges between 0.09 and 0.11 only. Significant jumps, on the other hand, are important components of realized variance, as they may contribute as much as 87.1% to the total realized variance.

[Table 1.6 here]

1.6 Leverage Effect

Large negative returns have a greater tendency to be accompanied by higher future volatility than positive returns of the same magnitude. This well-documented phenomenon is known as the leverage effect, or asymmetric cross-correlations. Subsequently the increase in volatility results in negative returns and causes what is known as the feedback effect.

Bollerslev *et al* (2006) find the leverage effect to be significantly negative and the feedback effect to be usually negligible for a horizon over several days. Bollerslev *et al* (2009) segregate the continuous and jump components of realized variance and conclude that leverage effects works primarily via the continuous part of the variation process. Our extended data confirm this result. We illustrate our finding using the simple, yet straight forward method of ABFN (2010).

We plot the cross-correlations between returns, realized variance and its two components at the daily level. The correlations in lags are the graphical expression of the leverage effect whereas the correlations in the leads represent the feedback effect. **Figures 1.4a-c** show median values for all the 10 sectors. In the case of rv_t and cv_t, the plots exhibit a clear tendency for the correlations between rv_t (cv_t) and returns to be negative for negative i with a distinctive peak around zero for positive i. In the jump case, the cross-correlation plots show no clear negative or positive tendency and fluctuate around zero. **Figure 1.4d** shows the sector median plots at the intra-day level for the cross-correlations between intra-day returns and absolute intra-day returns. Again, the median plots show a very similar pattern to those of ABFN (2010) where it is shown that the leverage effect may exist at high frequencies but with little or no impact for the feedback effect.

To summarize, our findings suggest that the leverage and feedback effects only exist within the realized variance (as well as the continuous variance) but not within the jumps component. The plots from other realized volatility measure also support this conclusion[4]. The results add more on the findings of both Bollerslev *et al* (2009) and ABFN (2010) who report that the leverage and feedback effects exist only in the volatility not in jumps. The degree of the effects may vary from one stock to another yet hold relatively constant across different realized volatility measures.

[4] The plots of rest three measures are not provided in the main figures as the plots of different realized volatility measures estimated within the same stock are visually indistinguishable between the plots of realized variance.

[Figure 1.4a-d here]

1.7 Correlations

Several studies report high positively correlations between realized volatility measures. Fuertes *et al* (2009) examine correlations between volatilities using 14 US stocks and find that correlations between realized volatility measures often exceed 90%. In this section, we look at the correlation between realized volatility measures themselves, and the correlation between volatility and volume. Further, we decompose realized variance and investigate the correlation between rv_t and its continuous and non-continuous components.

1.7.1 Correlation Matrix of Realized Volatility Measures and Trading Volume

Overall and sector average level correlations between realized volatility measures and trading volume are reported in **Table 1.7a**. In line with the literature, all the realized volatility measures are highly positively correlated with each other. This finding holds for all the sectors. With few exceptions, the correlation coefficients between volatility exceed 90%. The correlations between rv_t , pv_t and bv_t are higher than the correlations between rr_t and other three measures. This is not surprising as rr_t is the only realized volatility measure that is not calculated from intraday returns. The construction of rr_t determines that the properties are most likely to be diversified from the rest three. So far we do not find any pattern between trading activeness and realized volatility measures properties from the level correlation. Whether they are highly actively traded or less actively traded, the level correlations remain high, suggesting a great similarity among the four realized volatility measures.

[Table 1.7a here]

In the same table, the correlations between different volatility measures with trading volume are reported. Trading volume is positively correlated with all the realized volatility measures, with the correlation coefficients ranging from 20% to 60%. The correlations between rr_t and volume and between pv_t and volume are higher than the other two volatility measures. rv_t consistently has the lowest correlations with volume in 9 of the 10 sectors.

1.7.2 Correlation Matrix of Continuous Variance, Jumps and Trading Volume

We also look at the correlations between rv_t, its two components and trading volume. In part, our findings differ from those of Giot *et al* (2010), who found negative correlations between jump components and trading volume, and between jump components and realized variance & continuous variance. The cross-sectional average correlations between jump components and volume are negative and are not significant at the 5% level for 9 of the 10 sectors (with the exception of HC).

The correlations between jumps and realized variance/continuous variance, which are reported in **Table 1.7b**, are non-negative in the most cases. With only one exception

(IT), values are positive. IT shows negative correlations between jumps and realized variance/continuous variance. In sector MAR, although the correlation between jumps and realized variance is positive, the jumps and continuous variance are negatively correlated.

The sector average p values suggest that the correlation between jumps, rv_t / cv_t, and volume are not significant. Therefore we further consider the correlations at individual stock level. Only 24 of 100 stocks [5] present negative correlations between jumps and realized variance/continuous variance. The IT sector has the highest number of stocks with negative correlations (6/10), while all stocks in HC and IND sectors have positive correlations between jumps and realized variance/continuous variance. For the remaining sectors, the negative correlation holds for 2 or 3 stocks (out of 10) in each sector. We also find this negative correlation to be more apparent in actively traded stocks. Again, the p values of individual stocks fail to accept the null that the correlation is significant at the 5% level.

Wang and Huang (2012), who use the same method as our paper to construct the jumps series, report this negative relationship between jumps and trading volume using Hu-Shen 300 index data. This negative relationship might be attributed to the "public information" contained in jumps, while the continuous variance (as well as the realized variance) is more likely to be driven by the "private information". When large part of trading is induced by private information, the relationship of these two series is inevitably negative. On the other hand, the difference between our results and those of Giot *et al* might be attributed to the identification of jumps. Giot *et al* (2010) identifies only significant jumps at the 0.01% level. In other words, they only consider very

[5] AA, AKS, AMZN, APPL, CEG, COST, CSCO, COST, CVX, DELL, DUK, FTR, MAR, MSFT, NUE, ORCL, OXY, Q, SUN, T, TEA, UL, WFC, and XOM.

large jumps in their dataset whereas we admit the presence of small positive jumps and even negative jumps.

Overall, we could only find negative correlations between jumps and volume but weak positive correlations between jumps and realized variance/continuous variance.

[Table 1.7b here]

1.8 Volatility Regimes

Section 1.7 findings along with the results from **Figure 1.4** demonstrate high correlation linking the various volatility measures. The results hint at a single regime governing the dynamics of the four realized measures. To investigate such a possibility, we adopt the Markov Switching (MS) model by Hamilton (1994). This aim is to investigate whether observed high correlations would lead to similar (different) regimes.

Markov switching models allow each observation to be assigned a probability of belonging to one of several Markov states. Here we apply Markov Switching Autoregression (MS-AR) modelling on the sector average and overall average rv_t, rr_t, pv_t and bv_t by specifying two-regime states, high and low.

$$V_t - \mu(s_t) = \sum_{j=1}^{p} \rho(V_{t-j} + \mu(s_{t-j}) + \varepsilon_t, \varepsilon \sim N[0, \sigma^2] \qquad [1.18.1]$$

where V_t denotes the four volatility measures. The unobserved random variable, s_t, is denotes the regime to which observations belong; then

$$\mu(s_t) = \begin{cases} \mu_0 & if\ s_t = 0\ (LOW) \\ \mu_1 & if\ s_t = 1\ (HIGH) \end{cases} \qquad [1.18.2]$$

s_t follows a Markov chain, defined by the transition probability between the N states:

$$p_{i|j} = P[s_{t+1} = i \mid s_t = j], \quad i, j = 0, \ldots N-1. \qquad [1.18.3]$$

Thus the probability of moving from state j in one period to state i in the next period depends on the previous state only. Since the system has to be in one of the N states, then

$$\sum_{i=0}^{N-1} p_{i|j} = 1 \qquad [1.18.4]$$

For our case of $i, j = 0,1$ and N=2, the Markov chain transition probability matrix P_t is

$$P_t = \begin{pmatrix} p_{0|0} & p_{0|1} \\ p_{1|0} & p_{1|1} \end{pmatrix} \qquad [1.18.5]$$

For instance, if $p_{0|1}$ is very small, the model is more likely to stay longer in the state 1 and *vice versa*.

In our sample, we first consider MS (2)-AR (1) model. MS (2)-AR (1) denotes two regimes (M=2) and 1 lag (p=1). Then, to account for the long memory property of the realized volatility measures, we also consider the MS-AR models with longer lags and find that the model fails to converge when the number of lags of exceeds 2. The MS-

AR requires a state vector of dimension $N = S^{(1+p)}$ to obtain the Markov representation for the likelihood evaluation for S regimes and p autoregressive model. When the autoregression orders are high, MS-AR becomes effectively infeasible (Doornik and Hendry, 2009). Therefore, we also consider the Markov Switching Dynamic (MS-DR) model, which has the same number of states and number of regimes ($N=S$). Both specifications reach the same conclusion, with slightly difference in the period of the high/low regimes.

Figure 1.5a plots the results of MS-AR model for overall sample average different realized volatility measures. The grey shaded area is regime 1, (the high volatility regime). It is clear that all four measures share something in common in high regimes. The high volatility regime detected by all four volatility measures is mainly concentrate on two periods: the internet bubble (2000-2001), and 2008, start of financial crisis. The high regime is similarly identified by rv_t, bv_t and rr_t In the case of pv_t, there are more days that are identified under the high volatility regime.

[Figure 1.5a here]

We then apply the MS-DR model on the sector average volatility measures. Although the overall average volatility measures show similar volatility regimes identification, the picture changes at the sector level. UTL is the only sector which is identified with low volatility regime, except a very short period of time in 2008 when testing pv_t. The 2008 financial crisis raises the market volatility overall, with the period from September to December of 2008 similarly identified as the high-regime across sectors and volatility measures. From September 2008 to December 2010, rv_t and rr_t are all in the high volatility regime. Periods which are in the high regime are shorter when we apply the model to pv_t and bv_t. High regime is only detected from September, 2008 to the end of the year or early the following year. The remaining 3 sectors which detect a high volatility regime in 2008/2009 period are ENG, IND and MAR. These are also the sectors which activities are less affected by the IT sector.

Unsurprisingly, the IT sector is more severely affected by the 2001 crisis, registering a high volatility regime by all volatility measures. Another sector which is greatly influenced by the 2001 crisis is TEL sector. The high volatility regime is first detected in the year 2001 to 2002, a year later than the peak of the dotcom bubble. TEL sector is heavily related to the IT sector and the factors which drive the volatility of IT sector up will also affect TEL sector, but in a lagging way. Another interesting finding is found in sector CD and CS. The period, identified as a high volatility regime by rv_t, pv_t and bv_t, is identified as a low volatility regime when applying the Markov switching model to rr_t.

[Figure 1.5b-f here]

To summarize, the results from the Markov Regime Switching models show that different volatility measures tent to have similar, yet not identical high/low volatility regimes. The regimes identified in rv_t and bv_t are most similar to each other whereas the regimes detected in pv_t are less like the others. The periods in high/low volatility regimes also differentiate from one sector to another.

1.9 Long Memory

The long-memory property in volatility is a well-documented stylized fact and features in many financial and macroeconomic series, see for example Robinson (1995), Ding, Granger and Engle (1993), Baillie *et al* (1996), ABDE (2001) and Bandi *et al* (2006)). The long-memory property feature applies to low and high frequency measures of volatility with a fractional differencing parameter "*d*" estimate in the range of (0.2 to 0.3) for low frequency measures and (0.3 to 0.4) for high frequency ones. The higher value in the high frequency measures is mainly attributed their less noisy feature.

Our interest is to check whether the "*d*" estimate of the various realized measures varies across different. There are several commonly used long memory tests in the literature such as Lo's modified rescaled arrange (Lo, 1991), the KPSS statistic (Kwiatkowski *et al*, 1992), the rescaled variance (Giraitis *et al*, 2003), the GPH statistic (Geweke and Porter-Hudak, 1983), and the *H* statistic in Robinson (1995) and Robinson and Henry (1999). Here we adopt the method of Robinson and Henry (1999). The bandwidth parameter m is 0.5, which is the most commonly used in the literature[6].

Table 1.8 reports average Robinson's "*d*" of different volatility measures and trading volumes at both the sector and overall levels.

[Table 1.8 here]

---

[6] We also consider $m = 0.3$ and $m = 0.4$, respectively. The conclusion remains. We also conduct the GPH and AFIMA (0, d, 0) long memory tests to estimate the long memory parameter d. Diffferent tests lead to the same conclusion.

Main findings are as follows. The Robinson's "d" estimates of various realized volatility measures tend not to differentiate much across sectors: they range from 0.3 to 0.4. There is variation between the different volatility measures. For example, pv_t shows the highest estimates and rv_t is the lowest. bv_t is higher than rv_t which might indicate that bv_t is robust to jumps. This result also applies to pv_t, which is also robust to jumps and shows a higher "d" estimate than bv_t.

We also find that realized measures of volatility of more actively traded sectors have greater persistence than those of less actively traded sectors. Thus, the most actively traded sector, IT, has the highest "d" estimates and the least active sector UTL shows the lowest persistence.

Figure 1.6a shows the Autocorrelation Functions (ACFs) of the average realized volatility measures. All the volatility measures of most stocks exhibit a strong hyperbolic decay pattern up to 50 lags. The decay pattern of realized measures of volatility is very close to each other. For every single stock, the ACFs of rv_t and bv_t are most similar to each other. This confirms the results from the Robinson's "d" as well. We also find that realized volatility measures of actively traded stocks sometimes show persistent yet periodic autovariance function. A similar feature was observed by Baillie and Bollerslev (1993) who use FX data and attribute this result to high trading activity. We also plot the ACFs of each realized volatility measures at sector level. **Plot 1.6b** shows the result for rv_t. The ACFs of rv_t vary from one sector to another in general. The ACF from UTL sector dies out more rapidly than any other sectors. UTL is the least actively traded sector. FIN sector, one of the most actively traded sector, exhibit periodic pattern.

[Figure 16a-e here]

The long memory presented in the realized volatility measures is strongly supported in our data. The lower "*d*" of rv_t than that of pv_t and bv_t supports the earlier findings that incorporating jumps in the realized volatility measures decreases the long-range dependence of the volatility. More persistent the realized volatility measure is, higher the predictive power of the realized volatility measure is likely to have. The highest "*d*" in pv_t suggests that pv_t may carry the strongest forecasting power than other three measures. Our research mainly focuses on the in-sample estimations, nevertheless, there are other studies which provide both direct and indirect supports for the superiority of more persistent volatility measures in forecasting. Ghysels *et al* (2006) propose a mixing frequencies model, which is known as MIDAS, and find that (logarithmic) pv_t produces the best out-of-sample forecasts among a series of volatility measures including rv_t. The indirect support is given by ABD (2007) and Corsi *et al* (2012) using heterogeneous autoregressive (HAR) model. Both papers find that excluding jumps enhances long-range dependence of rv_t and hence enhances the forecasting power of the measures. Moreover, MIDAS and HAR models are both developed to capture the long memory in volatility measures.

1.10 Volatility-Volume Relation

Clark (1973) provides theoretical foundations for the volatility-volume relation in the framework of what is now known as the Mixture of Distribution Hypothesis model. Following Clark's paper, an impressive body of literature has investigate the relation (see Epps and Epps (1976), Karpoff (1987), Harris (1987), Ané and Geman (2000), Martens and Luu (2003) and Chan and Fong (2000, 2006)). Advances in volatility estimation have led to improved measures of volatility and have further emphasized the relation between volatility and volume. Depending on the volatility measures adopted, trading volume is capable of explaining between 5 - 50% of the variation in the volatility. More recently, and given the empirical evidence documenting the presence of jumps, Giot *et al* (2010) revisited the volume (number of trades) - volatility relation. They divided realized variance (rv_t) into its continuous and jump components and showed that the relation holds only for the continuous part of rv_t.

To test which is the best volatility proxy, we closely follow the regression framework of Jones *et al* (1994), Ané and Geman (2000) and Chan and Fong (2006) and then regress different realized volatility measures on trading volumes

$$rv_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} rv_{it-j} + \gamma_i vol_{it} + \varepsilon_{it} \qquad [1.19.1]$$

$$rr_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} rr_{it-j} + \gamma_i vol_{it} + \varepsilon_{it} \qquad [1.19.2]$$

$$pv_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} pv_{it-j} + \gamma_i vol_{it} + \varepsilon_{it} \qquad [1.19.3]$$

$$bv_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} bv_{it-j} + \gamma_i vol_{it} + \varepsilon_{it} \qquad [1.19.4]$$

where M_t is a Monday dummy and 12 lags of realized volatility measures are used to account for serial correlation[7].

---

4. The volume of some stocks are trend stationary, in this case, we de-trend the series according to the ADF test.

Table 1.9 shows the results of volume regressions at sector average and overall average levels. Among four realized volatility measures, pv_t produces the highest $\bar{R}^2$ value of 67.8% in the volume regression on average. The lowest value for all realized volatility measures is reported by realized variance, which yields average $\bar{R}^2$ of 52.8% in the volume regression. However, regardless of the realized volatility measure used, our results are similar to those reported by previous studies which consider only rv_t (or its components) (Chan and Fong (2006), Martens and Luu (2003) and Giot *et al* (2010)) yet with higher $\bar{R}^2$. The percentage of stocks for which trading volume is statistically significant at 5% ranges from 96% to 98%.

[Table 1.9 here]

At sector level, ENG and MAR report the highest $\bar{R}^2$ across realized volatility measures. The lowest $\bar{R}^2$ across realized volatility measures is reported by TEL sector: the sector contains most jumps is most poorly explained by realized volatility measures and produces lowest $\bar{R}^2$ on average.

To conclude, the well-established volatility-volume relation holds when all four realized volatility measures are tested. We reach the same findings as previous studies and have higher $\bar{R}^2$ on average. The results show that pv_t is the most closely explained volatility measure in the volatility-volume relation, at both the sector and overall levels. The second most closely explained is realized bipower variation, followed by realized range and realized variance. We investigate this relation in greater detail in the next chapter which focuses on the MDH validity using high frequency data.

1.11 Distributions of Returns, Realized Volatility Measures and Standardized Returns

1.11.1 Distribution of Returns

The statistics from **Section 1.4.1** show that returns are not normally distributed. **Figure 1.6** fits the Gaussian distribution to sector average returns. The parameters of the distributions are estimated using the maximum likelihood (ML) method. The red dotted line is the unconditional log-density for the ML fit of the normal distribution while the black solid line is the unconditional log-density for the return series. The plots show that the return distributions are more peaked around zero and have fatter and more fluctuated tails than the standard normal. In contrast with standard normal distribution which has a fast decay rate, the return series' decaying rate is much slower.

BN-S (2002) show that the normal distribution using a log-log density representation has faster decay rates than log-linear. Densities with fast decay rates have so called 'sub-log-linear' tails and with slow decay rates have 'sup-log-linear' tails. The density plots in **Figure 1.7** show clearly that returns have 'sup-log-linear' tails.

3 out of 10 sectors appear to have longer left tails than right tails, namely ENG, FIN and MAR, suggesting more extreme values in negative returns. FIN has the longest left tail and it is also the sector which is most influenced by the financial crisis and yields largest degree of loss in the equity market. CD, IT and UTL have longer right tails, an indication that extreme positive returns are more than negative ones. The distributions of the returns of rest of sectors are more symmetric.

[Figure 1.7 here]

1.11.2 Distributions of Realized Volatility Measures

Another well-established stylized fact is that realized volatility measures are best approximated by Inverse Gaussian and Lognormal distributions.

Most of studies focus on the distributional properties of rv_t alone. The distribution of rv_t is lognormal or close to lognormal is assessed by using foreign exchange data (ABDL, 2001), US individual stock data (ABDE, 2001) as well as UK index data (Areal and Taylor, 2002). BN-S (2002) use the same data as ABDL (2001) and find that rv_t could also be approximated by the Inverse Gaussian distribution and the fits of IG and Lognormal are equally well. The same conclusion is found by Forsberg and Bollerslev (2002) from examining of 10 year ECU basket currencies/ US dollar and by Stentoft (2008) from examining US equity data.

In this section, we re-examine the distributional properties of realized volatility measures and try to find out the best fit of realized volatility measures distributions by fitting three different distributions: normal, lognormal and inverse Gaussian. Besides the sector average volatility measures, we also look at the volatility measures of most and least actively traded stocks in each sector. We use the Kolmogorv-Simirnov (KS) test which compares the empirical distribution function with the theoretical distribution function non-parametrically.

From **Table 1.10a**, all volatility measures across different sectors confirm with neither IG nor Lognormal distributions. However, the KS statistics confirm that all four realized volatility measures are closely approximated by IG and Lognormal distributions as the KS test statistics are closer to 0.0258 -- the critical value that the null cannot be rejected at 5%. Together with the test statistics, **Figure 1.8** shows that IG and Lognormal distributions of different realized volatility measures may be regarded as empirically indistinguishable -- a finding firstly proposed by BN-S (2002). The realized volatility measures are slightly better fitted by lognormal distribution than by the IG distribution across volatility measures, as the KS statistics for lognormal distribution is lower. Lognormal distribution fits bv_t best and IG fits pv_t best. rr_t is the worst fitted volatility by both IG and lognormal distributions as it yields highest KS statistics.

[Table 1.10a here]

[Figure 1.8 here]

Table 1.10b reports the KS statistics of volatility measures from the most and least actively traded stocks in every sector. The null hypothesis is not fully rejected for different realized volatility measures for three stocks. IG fits rv_t, pv_t and bv_t of MAR at the 5% level and fits rr_t of MAR and rv_t & bv_t of MSFT at the 1% level. Lognormal fits rv_t and bv_t of MAR, and rv_t and pv_t of S at the 1% level. The volatility measures are better fitted by the lognormal distribution than the IG distribution at individual stock level. For the 20 stocks that are investigated, both distributions fit bv_t well. However, when we further divide the individual stocks into the most and least actively traded groups, we find that the IG fits pv_t best and the Lognormal fits bv_t best among the most active stocks, whereas the IG fits bv_t best and the Lognormal fits pv_t best among the least active stocks.

[Table 1.10b here]

To sum up, the distribution of realized volatility measures across sectors can be equally described by the Inverse Gaussian and Lognormal distributions. The KS test accepts both as valid representation of the data. However, the fit of the lognormal distribution ranks better than IG. This holds at both sector average and individual stock levels. At the sector average, we also find that IG fits pv_t best whereas the lognormal fits bv_t best.

1.11.3 Distributions of Standardized Returns

The distributional properties of returns and realized volatility measures confirm the stylized facts that returns are not Gaussian and realized volatility measures are closely fitted by both Inverse Gaussian and Lognormal. Then we consider the distributional properties of returns standardized by realized volatility measures. A fundamental theorem of asset pricing implies that, in the absence of arbitrage effects, prices are semi-martingales under a given physical measure. Monroe (1978) asserts that any semi-martingale can be written as a time changed Brownian motion. Clark (1973) shows that subordinated returns are normal with trading volume acting as a subordinator. The Clark and Monroe assertions require continuity in the underlying Brownian motion process. In chapter 3 we will discuss whether continuity is a necessary assumption for recovering returns normality.

ABDL (2001), ABDE (2001), BN-S (2002), Areal and Taylor (2002), Fleming and Paye (2007, 2011), ABFN (2010) have addressed the issue of recovering returns normality using different volatility measures allowing for the effects of noise and jumps. The general finding from these studies indicates that return normality is achievable once we account for such as (potential) noise, leverage and jumps. As mentioned, in this chapter, our task is to investigate whether the various volatility measures behaves in a systematic way across sectors. We consider both raw and demeaned returns

$$sr_t = r_t / \sqrt{V_t} \qquad [1.20]$$

where sr_t is the standardized returns and V_t denotes different realized volatility measures.

Tables 1.11 and 1.12 respectively show the number of stocks that the hypothesis of normality of daily returns and demeaned returns standardized by realized volatility measures are rejected at the 1%, 5% and 10% level in overall sample in each sector.

[Table 1.11 and 1.12 here]

The results from JB statistics, skewness and kurtosis suggest that all four measures are able to recover the normality to some extent when they are used as the standardization factor. We also observe that subtracting the sample mean from the return series in the numerator (or the demean process) does not enhance the normality of the standardization process by much. This is contracted with the results reported by ABFN (2010) who find that the distribution of demeaned returns standardized by realized volatility is closer to normal. The JB statistics of standardized demeaned returns are approximately the same as that of standardized raw returns.

Among all volatility measures, the standardization of pv_t performs best as 79/100, 66/100 and 51/100 of $r_t/\sqrt{pv_t}$ cannot reject the Gaussian distribution at the 1%, 5%

and 10% level, respectively. The second best is bv_t, bringing 55/100 $r_t/\sqrt{bv_t}$ back to normality at the 10% level, and then followed by rr_t and rv_t. This conclusion holds for both raw and demeaned return series, with demeaned series providing slightly fewer rejections at different significance levels. We also observe that rv_t is less successful in working in the more actively traded sectors while bv_t is more successful in the less actively traded sectors. In our sample, when volatility measures are used to standardize returns, the two best performed sectors are the ENG and UTL, of which 24/40 standardized returns do not reject the normality null at the 10% level. Two least successfully performed sectors are the MAR and IT, of which 10/40 and 11/40 standardized returns do not reject the normality at the 10% level. Among all sectors, UTL is the least actively traded sector and IT is the most actively traded sectors. UTL is also one of the sectors which have a large number of jumps detected and IT detects lowest number of jumps, both at daily and intraday level.

Given the fact pv_t is the most successful standardization factor overall, it is not always the most efficient one for every sector. The performance of the realized measures of volatility varies from one to another. rv_t works best in CS sector, rr_t and bv_t works best in ENG sector while pv_t is most successful in HC sector. Besides ENG, a sector where all realized volatility measures work relatively well, and MAR, a sector that all realized volatility measures fail to standardized most of the return series, the standardization results using different volatility measures are rather mixed for the rest of 9 sectors.

The literature provides several justifications of why realized volatility measures standardized returns are still not normally distributed. Fleming and Paye (2011) propose microstructure noise which could distort the variance of the standardized returns and artificially inflate the kurtosis of the standardized returns. ABFN (2010) suggests that leverage effect and presence of jumps are important for the distributional properties of standardized returns. A more detailed discussion addressing recovering normality of returns is provided in the later chapter.

1.12 Sector Rankings

Table **1.13** ranks sectors by mean values of the volatility measures, market activity measure, intraday and daily jumps, persistence and the volatility-volume regression adjusted R-squared, and by the rejections of normality of standardized returns. The table shows that rankings of the various measurements considered tend to vary across sectors. This table aims to establish links among the various strands of the stylized facts addressed in this chapter.

First, only 4/10 sectors show that volatility measures rank the same across sectors. The existing literature provides no explanation as why this should be the case. Ideally we should have different rankings for every sector but given that four sectors agree on a similar ranking is of interest especially all four are quite distinct in the nature of activities and services they provide. It would be interesting to see whether the ranking shifts if the variables were to be observed at different time windows.

Second, high volatility and trading volume episodes are not aligned as we would expect. For example, the most actively traded sector IT ranks low by volatility and the most volatile sector MAR ranks low by trading volume. The volatility-volume relation as shown by the regressions also highlights this issue.

Third, the number of identified (significant) intraday and daily jumps does not match except for TEL and IT. The sectors which contain most (least) daily jumps also contain most (least) intraday jumps. The most actively traded sector has the smallest number of intraday and daily jumps. Moreover, our ranking results suggest that the number of jumps is not associated with the level of volatility measures: the most jumps-contained sector is neither the most actively traded nor the most volatile.

Returns normality could be achieved in various sectors and using all realized measures. The results obtained are independent of the level of jumps detected both at the daily and the intraday levels. The level of stock activity seems to be a dominant factor in the ability of the realized measures to recover normality. The IT sector has the lowest intraday and daily jumps but the highest percentage of normality rejections. The least actively traded sectors (CS and UTL) shows the lowest rejections of returns normality. One possible explanation is provided by Ait-Sahalia *et al* (2009) who find a positive relationship between microstructure noise and trading volume, hence high activity entitles high microstructure noise. In line with their findings, low activity entitles less microstructure noise and hence less discontinuities. A similar argument is provided by Hansen and Lunde (2005). Given that CS and UTL sectors are less actively traded they are expected to have more consistent realized volatility estimators, hence more capable of recovering returns normality.

Persistence and activity tend to move jointly as advocated by the mixture of distribution hypothesis. The ranking for persistence varies with the realized volatility measure tested. UTL and IT are the least (most) persistent sectors. UTL is the least active and least persistent whereas IT is the most active and persistent. The mixture of distribution assumes a common process driving volatility and market activity as characteristics shared by both activity and volatility measures. The ranking for persistence varies with the realized volatility measure tested. This can be attributed to two factors. The first is the presence of microstructure noise. For example, the realized range, by construction is known of its sensitivity to microstructure noise. Second, the sensitivity of certain measures to jumps.

The sectors that contain the most jumps show lowest $\bar{R}^2$ (CD and TEL). The presence of jumps in the volatility measures weakens the well-documented volatility-volume relation. This finding is supported by Giot *el al* (2010) who finds a negative jumps-volatility relation. The volatility-volume regression results are more significant in the most actively traded sectors and are sensitive to the volatility measure in use. For example, rr_t is mostly explained by volume in the MAR sector whereas rv_t, pv_t and bv_t are all mostly explained in the ENG sector.

The above findings which pertain to the realized measures of volatility might not all be novel, but provide a better picture about the interaction of activity and volatility measures across different market sectors. The different performance of sectors, as well as the stocks within these sectors provides a useful insight about the dynamics of the market and may help explain why a common warning system which treats all as being the same is bound to fail.

[Table 1.13 here]

1.13 Conclusion

We look at trade data of 100 stocks from 10 sectors traded in the US stock market for the period 2000-2010. We investigate the degree of variation across sectors in the stylized facts of realized volatility measures with respect to the optimal sampling frequency, correlations, jumps, leverage effect, volatility regimes, persistence, and volatility-volume relation. We also look at the distributional properties of returns, standardized returns and realized measures of volatility. Our findings can be summarised as follows.

We find that the 5-minute is the optimal sampling frequency for realized measures of volatility. This result does not hold for the power variation, particularly for stocks number of jump activity where the point of convergence overshoots the 5 minute point.

Rankings of realized volatility measures tend to vary across sectors. Deviations are more visible in active sectors. This result holds for all volatility measures with the exception of realized variance and bi-power variation. The rankings of these two measures are the same.

Power variation shows systematically best performance across sectors and outperforms other realized measures. It has the lowest standard deviation, well defined distributional properties, has the highest degree of persistence, most capable of recovering normality, and robust to jumps. It is the most accurate volatility estimator to the true volatility among the four volatility measures.

Realized volatility measures are highly and positively correlated with each other (over 85%). The jump component from realized variance is negatively correlated with volume and positively correlated with realized variance.

Realized measures regimes characteristics vary by regime type (high-low) and by sector. Realized power variation is detected with the longest period in the high regime while realized range has the shortest high regime. Regime patterns for realized variance and realized bipower mimic each other. Although most of the sectors are detected with both high-low volatility regimes, UTL, the least traded sector, is only identified with low regime for 3 out of 4 realized volatility measures.

The volatility-volume relation holds in all sectors as borne out by the various realized volatility measures. The relations holds best for realized power variation, followed by realized bi-power variation, realized range and realized variance. The presence of jumps tends to negatively impact the relation as sectors which contain more jumps tend show less association between realized measures and trading volume.

The distribution of the realized volatility measures can be equally described by both the Inverse Gaussian and Lognormal distributions. The realized power variation is best approximated by the Inverse Gaussian and the realized bi-power variation by the lognormal.

Returns standardized by the realized measures are normal. Realized power variation outperforms the other three measures in recovering returns normality. The result holds for both raw and demeaned return series. Returns normality in actively traded sectors is more difficult to achieve. We also find no relation between the degree of jump activity and recovering returns normality.

Tables

Table 1.1: 100 stock list (most to least active sectors/stocks according to trading volume)

Consumer Discretionary		Consumer Staples		Energy		Financials		Health Care	
5		9		7		2		3	
HD	The Home Depot	**WMT**	Wal-Mart	**XOM**	Exxon Mobil	**C**	Citigroup Inc.	**PFE**	Pfizer
DIS	Walt Disney	**PG**	Procter & Gamble	**HAL**	Halliburton Co.	**BAC**	Bank of America	**GILD**	Gilead Sciences
AMZN	Amazon.com Inc.	**KO**	Coca-Cola	**CVX**	Chevron Corp.	**WFC**	Wells Fargo	**MRK**	Merck
NWSA	News Corporation	**PEP**	PepsiCo Inc.	**CHK**	Chesapeake Energy	**JPM**	JPMorgan Chase	**AMGN**	Amgen Inc.
TWX	Time Warner Inc.	**COST**	Costco	**WMB**	Williams Cos.	**MS**	Morgan Stanley	**BSX**	Boston Scientific
GPS	Gap (The)	**AVP**	Avon Products	**OXY**	Occidental Petroleum	**AXP**	American Express	**JNJ**	Johnson & Johnson
MCD	McDonald's	**KMB**	Kimberly-Clark	**BHI**	Baker Huges	**GS**	Goldman Sachs Group	**UNH**	United Health Group
BBY	Best Buy Co. Inc.	**EL**	Estee Lauder Cos.	**DVN**	Devon Energy Corp	**BK**	Bank of NY Mellon	**ABT**	Abbott Laboratories
IPG	Interpublic Group	**UL**	Unilever	**SUN**	Sunoco Inc.	**ALL**	Allstate Corp.	**MDT**	Medtronic Inc.
MAR	Marriott Int'l.	**BFb**	Brown-Forman Corp.	**TE**	TECO Energy	**TRV**	Travellers	**HUM**	Humana Inc.
Industrials		**Information Technology**		**Materials**		**Telecommunications Services**		**Utilities**	
4		1		8		6		10	
GE	General Electric	**MSFT**	Microsoft	**AA**	Alcoa	**S**	Sprint Nextel Corp	**DUK**	Duke Energy
CAT	Caterpillar	**INTC**	Intel Corp.	**FCX**	Freeport-McMoran	**T**	AT&T	**EXC**	Exelon Corp.
LUV	Southwest Airlines	**CSCO**	Cisco Systems	**DOW**	Dow Chemical	**Q**	Qwest Communication Int	**SO**	The Southern Company
BA	Boeing	**ORCL**	Oracle Corp.	**NEM**	Newmont Mining	**VZ**	Verizon Communications	**PEG**	Public Serv. Enterprise Inc.
UTX	United Technologies	**DELL**	Dell Inc.	**X**	United States Steel Corp.	**VOD**	Vodafone Group Plc.	**AEP**	American Electric Power
HON	Honeywell Int'l Inc.	**AAPL**	Apple Inc.	**DD**	DuPont	**AMT**	American Tower Corp	**PCG**	PG&E Corp.
MMM	3M	**EMC**	EMC Corp.	**NUE**	Nucor Corp.	**FTR**	Frontier Communications	**CEG**	Constellation Energy Group
UPS	United Parcel Service	**HPQ**	Hewlett-Packard	**WY**	Weyerhaeuser Co.	**CTL**	CenturyTel Inc.	**ETR**	Entergy Corp.
CMI	Cummins Inc.	**IBM**	IBM	**IP**	International Paper	**TEF**	Telefonica S.A.	**PGN**	Progress Energy, Inc.
GD	General Dynamics	**XRX**	Xerox Corp.	**AKS**	AK Steel Holding Corp.	**BT**	BT Group Plc.	**OKE**	Oneok

Note: The table report the full sample of 100 stocks traded in the US equity market. The sectors are arranged alphabetically with the ranking according to the activeness of trading. Stocks are arranged from most actively traded to the least. For instance, IT sector is the most actively traded sector and MSFT is the most actively traded stock within the IT sector.

Table 1.2: TAQ and TICK comparative statistics (Intel Corporation (INTC))

	TICK				TAQ			
	RV_t	RR_t	PV_t	BV_t	RV_t	RR_t	PV_t	BV_t
Mean	5.359	5.874	9.858	5.186	5.893	19.173	10.294	5.544
Median	2.937	3.178	6.854	2.832	3.120	5.224	7.077	2.935
Maximum	89.431	104.425	82.261	74.089	155.640	17950.650	110.006	152.531
Minimum	0.167	0.237	0.772	0.153	0.194	0.275	0.881	0.167
Std. Dev.	6.557	7.127	8.393	6.427	8.731	342.054	9.179	8.021
Skewness	3.688	3.974	2.338	3.593	7.779	52.098	3.067	7.494
Kurtosis	26.166	32.838	11.067	23.555	110.711	2731.434	21.089	109.136
JB	6.815E+04	1.099E+05	1.002E+04	5.466E+04	1.365E+06	8.600E+08	4.206E+04	1.325E+06
ADF	-4.777	-4.022	-3.713	-4.891	-6.198	-52.527	-3.663	-5.236
P value	(0.000)	(0.001)	(0.004)	(0.000)	(0.000)	(0.000)	(0.005)	(0.000)

Note: This table reports the summary statistics of realized measures of volatility, namely realized variance, realized range, realized power variation, realized bipower variation, and calculated using tick to tick data from TICK and TAQ database, respectively. JB is the Jarque-Bera test statistics for normality. ADF denotes the augmented Dickey-Fuller statistics for the null of a unit root with 5% and 1% critical values of 2.862 and -3.433 respectively.

Table 1.3: Summary statistics of sector average daily returns (in %)

	Daily Return (R_t)									
	CD	**CS**	**ENG**	**FIN**	**HC**	**IND**	**IT**	**MAR**	**TEL**	**UTL**
Mean	0.032	0.050	-0.037	-0.029	-0.001	-0.001	0.015	-0.088	-0.027	0.011
S.D.	2.297	1.449	2.301	2.488	1.932	1.842	2.328	2.510	2.158	1.722
Skewness	0.166	0.132	-0.697	-0.192	0.030	-0.053	0.175	-0.186	-0.030	-0.785
Kurtosis	12.506	8.964	20.780	16.455	7.999	7.633	8.485	8.019	13.283	27.297
JB	19844.315	4351.420	74013.418	25542.559	3911.435	2737.765	4519.398	3121.055	17551.998	162754.940
ADF	-29.600	-21.213	-17.786	-14.728	-25.421	-22.511	-21.834	-18.936	-18.661	-29.177
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: Summary statistics of daily returns for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation. ADF denotes the augmented Dickey-Fuller statistics for the null of a unit root with 5% and 1% critical values of 2.862 and -3.433 respectively. The daily retunrs are pooled across stocks and then summarized.

Table 1.4a: Summary statistics of realized volatility measures (level)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					Realized Variance (rv_t)					
Mean	5.455	2.453	5.836	6.339	4.245	3.696	5.722	6.703	5.637	3.419
Median	3.334	1.486	3.493	2.537	2.624	2.336	3.124	4.389	3.024	1.752
S.D.	7.100	2.913	9.250	17.207	4.660	5.111	6.787	9.499	9.721	13.239
Skewness	7.440	5.183	8.681	11.436	4.918	8.352	3.648	8.620	8.961	31.670
Kurtosis	111.077	50.947	118.145	198.683	56.216	139.260	24.865	123.691	140.262	1267.273
ACF(36)	0.339	0.357	0.248	0.303	0.396	0.329	0.442	0.343	0.306	0.043
Q(36)	21271	26899	18632	16238	29603	25340	31406	27868	22637	1437
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					Realized Range (rr_t)					
Mean	4.866	2.069	5.061	6.115	3.761	3.197	6.046	5.687	4.748	2.836
Median	2.991	1.226	2.997	2.291	2.374	1.980	3.138	3.436	2.641	1.356
S.D.	6.253	3.217	8.925	17.777	4.533	5.051	7.449	9.419	8.123	12.086
Skewness	7.284	14.125	10.782	11.364	8.794	10.347	4.048	9.902	8.322	25.128
Kurtosis	112.688	343.540	197.363	194.919	174.717	198.423	31.502	165.886	118.315	746.710
ACF(36)	0.390	0.225	0.245	0.321	0.338	0.321	0.454	0.357	0.312	0.044
Q(36)	26766	11556	18134	17868	23956	22808	32480	28867	23597	1232
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					Realized Power Variation (pv_t)					
Mean	9.155	5.117	9.553	9.571	7.615	7.027	9.918	10.922	8.627	5.980
Median	6.928	3.798	7.231	5.814	5.723	5.428	6.772	8.405	6.217	4.299
S.D.	7.378	3.934	8.677	13.844	5.561	5.973	8.155	9.452	8.113	7.135
Skewness	3.470	3.472	5.516	6.320	3.129	4.564	2.425	5.499	3.940	10.348
Kurtosis	26.878	25.109	55.262	67.386	23.817	41.258	11.892	51.830	28.578	200.242
ACF(36)	0.507	0.468	0.385	0.485	0.503	0.443	0.542	0.454	0.495	0.230
Q(36)	40842	39755	33186	33234	41391	37477	43212	39195	43618	16063
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					Realized Bipower Variation (bv_t)					
Mean	4.994	2.230	5.407	5.977	3.877	3.445	5.457	6.231	5.071	3.143
Median	3.037	1.364	3.255	2.327	2.413	2.185	2.934	3.985	2.738	1.593
S.D.	6.148	2.655	8.713	16.063	4.303	4.877	6.616	9.403	9.304	13.728
Skewness	5.652	5.377	8.488	10.813	5.324	8.157	3.762	9.043	11.684	37.873
Kurtosis	68.374	54.831	111.520	176.455	65.569	125.802	26.269	134.358	244.450	1704.928
ACF(36)	0.394	0.348	0.249	0.319	0.384	0.318	0.432	0.335	0.270	0.037
Q(36)	28495	28333	19709	17661	28819	24876	30717	27660	18939	1075
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: Summary statistics of daily volatility measures for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation. ACF (36) is the autocorrelation at 36th lags and Q (36) is the Q statistics at 36th lag. The realized volatility measures are pooled across stocks and then summarized.

Table 1.4b: Summary statistics of volatility measures (logarithmic)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					Realized Variance ($\ln(rv_t)$)					
Mean	1.290	0.546	1.353	1.021	1.101	0.928	1.310	1.577	1.207	0.646
Median	1.204	0.396	1.251	0.931	0.965	0.848	1.139	1.479	1.107	0.561
S.D.	0.845	0.766	0.805	1.107	0.777	0.791	0.883	0.698	0.919	0.881
Skewness	0.537	0.751	0.767	0.782	0.612	0.676	0.544	1.042	0.698	0.907
Kurtosis	2.856	3.329	3.943	3.743	2.912	3.643	2.611	4.948	3.446	4.810
ACF(36)	0.677	0.596	0.560	0.720	0.625	0.601	0.688	0.504	0.700	0.544
Q(36)	56370	48439	45424	61836	49572	48552	57787	37553	60102	44375
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					Realized Range ($\ln(rr_t)$)					
Mean	1.189	0.365	1.200	1.021	1.101	0.928	1.310	1.577	1.207	0.646
Median	1.096	0.204	1.098	0.931	0.965	0.848	1.139	1.479	1.107	0.561
S.D.	0.827	0.752	0.791	1.107	0.777	0.791	0.883	0.698	0.919	0.881
Skewness	0.572	0.894	0.926	0.782	0.612	0.676	0.544	1.042	0.698	0.907
Kurtosis	2.935	3.947	4.544	3.743	2.912	3.643	2.611	4.948	3.446	4.810
ACF(36)	0.696	0.594	0.569	0.720	0.625	0.601	0.688	0.504	0.700	0.544
Q(36)	58889	48608	47209	61836	49572	48552	57787	37553	60102	44375
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					Realized Power Variation ($\ln(pv_t)$)					
Mean	1.997	1.446	2.050	1.833	1.847	1.746	2.049	2.213	1.897	1.527
Median	1.936	1.335	1.978	1.760	1.745	1.692	1.913	2.129	1.827	1.458
S.D.	0.628	0.569	0.589	0.830	0.572	0.594	0.670	0.533	0.669	0.649
Skewness	0.529	0.755	0.779	0.749	0.627	0.688	0.527	1.088	0.653	0.832
Kurtosis	2.841	3.386	4.114	3.610	3.024	3.722	2.596	5.081	3.271	4.344
ACF(36)	0.694	0.612	0.550	0.727	0.642	0.605	0.691	0.534	0.714	0.551
Q(36)	59540	51571	45475	63190	52608	49710	58666	41736	62835	46296
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
					Realized Bipower Variation ($\ln(bv_t)$)					
Mean	1.203	0.456	1.274	0.953	1.015	0.854	1.251	1.490	1.096	0.555
Median	1.111	0.310	1.180	0.845	0.881	0.782	1.076	1.383	1.007	0.465
S.D.	0.848	0.759	0.802	1.111	0.770	0.790	0.893	0.702	0.919	0.878
Skewness	0.529	0.758	0.804	0.796	0.623	0.708	0.542	1.131	0.713	0.948
Kurtosis	2.787	3.401	4.099	3.745	3.002	3.772	2.641	5.276	3.517	4.930
ACF(36)	0.687	0.589	0.551	0.722	0.622	0.589	0.686	0.496	0.692	0.532
Q(36)	57837	48462	44893	62630	49507	47553	57738	37522	59534	43733
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: Summary statistics of logarithmic daily volatility measures for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation. ACF (36) is the autocorrelation at 36th lags and Q (36) is the Q statistics at 36th lag. The realized volatility measures are pooled across stocks and then summarized.

Table 1.5: Summary statistics of realized variance and its components

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
						Realized Variance (rv_t)				
Mean	5.455	2.453	5.836	6.339	4.245	3.696	5.722	6.703	5.637	3.419
S.D.	7.100	2.913	9.25	17.207	4.66	5.111	6.787	9.499	9.721	13.239
Maximum	149.54	51.346	187.92	394.56	89.124	125.26	89.731	208.39	212.44	571.53
Skewness	7.44	5.183	8.681	11.436	4.918	8.352	3.648	8.62	8.961	31.67
Kurtosis	111.08	50.947	118.15	198.68	56.216	139.26	24.865	123.69	140.26	1267.3
						Continuous Variation (cv_t)				
Mean	5.367	2.363	5.757	6.272	4.159	3.622	5.663	6.616	5.525	3.328
S.D.	7.086	2.904	9.249	17.208	4.652	5.112	6.794	9.506	9.721	13.237
Maximum	147.86	51.279	187.88	394.43	89.091	125.17	89.736	208.41	212.62	571.64
Skewness	7.38	5.206	8.686	11.431	4.944	8.359	3.645	8.621	8.974	31.697
Kurtosis	109.25	51.368	118.14	198.53	56.673	139.26	24.822	123.62	140.7	1269.4
						Jumps (J_t)				
Mean	0.088	0.09	0.079	0.067	0.086	0.074	0.059	0.087	0.112	0.091
S.D.	0.124	0.094	0.091	0.094	0.096	0.083	0.077	0.089	0.107	0.1
Maximum	2.932	1.153	1.162	1.357	1.284	1.358	1.458	0.92	1.256	1.176
Skewness	8.528	3.002	2.556	4.936	2.941	3.074	4.315	1.833	2.526	2.197
Kurtosis	142.54	23.587	22.071	52.162	24.723	34.606	57.823	13.653	18.526	17.171

Note: This table reports the mean, standard deviation, skewness, kurtosis and maximum at sector average realized variance and its continuous and non-continuous components from 03/01/2000 to 31/12/2010. The realized variance and its components are pooled across stocks and then summarized.

Table 1.6: Jump detection methods and jump contribution to the return/variability

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	Overall
						L & M Jump Test (5%)					
Number of detected jumps	136	143	112	99	134	114	73	142	153	140	124
Number of days with at least 1 significant jump	127	132	105	95	125	107	69	131	146	129	117
Proportion of detected jumps	0.07%	0.07%	0.06%	0.05%	0.07%	0.06%	0.03%	0.07%	0.08%	0.07%	0.06%
Proportion of days with at least 1 significant jump	4.91%	5.30%	4.06%	3.53%	4.03%	4.06%	2.53%	5.05%	5.42%	5.25%	4.46%
						***MaxZ* Jump Test (5%)**					
Number of detected jumps	730	710	665	636	706	713	634	668	823	694	698
Proportion of detected jumps	27.98%	27.56%	25.64%	23.67%	27.14%	26.91%	24.39%	25.57%	31.65%	27.96%	26.85%
Average Jump Contribution	0.263	0.262	0.259	0.248	0.265	0.259	0.241	0.257	0.264	0.266	0.258
Maximum Jump Contribution	0.871	0.815	0.807	0.86	0.864	0.819	0.666	0.831	0.872	0.852	0.826
Std. Dev. of the Jump Contribution	0.11	0.106	0.102	0.104	0.112	0.104	0.088	0.101	0.106	0.109	0.104
Skewness of the Jump Contribution	1.933	1.685	1.719	2.075	1.761	1.698	1.528	1.742	1.836	1.745	1.772
Kurtosis of the Jump Contribution	8.359	6.849	7.339	9.153	7.109	6.971	5.866	7.338	8.205	7.204	7.439
Intraday Jumps Ranking	**5**	**2**	**8**	**9**	**6**	**7**	**10**	**3**	**1**	**4**	
Daily Jumps Ranking	2	4	8	5	9	3	10	7	1	6	

Note: The table reports sector average, overall average and top 30 average numbers of jumps detected at daily and intraday levels. The intraday jump detection is based on L&M jump test and the daily jump detection is based on the MaxZ Jump test. 'Intraday Jumps Ranking denotes the descending rankings of number of jumps detected by intraday jump test across sectors while 'Daily Jumps Raking' denotes the descending rankings of number of jumps by daily jump test cross sectors. All the jumps are reported at 5% significance level.

Table 1.7a: Correlation metrics I

	vol_t	rv_t	rr_t	pv_t	bv_t	vol_t	rv_t	rr_t	pv_t	bv_t	vol_t	rv_t	rr_t	pv_t	bv_t	vol_t	rv_t	rr_t	pv_t	bv_t
			CD					CS					ENG					FIN		
vol_t	1.000					1.000					1.000					1.000				
p value	-					-					-					-				
rv_t	0.360	1.000				0.385	1.000				0.456	1.000				0.548	1.000			
p value	(0.000)	-				(0.000)	-				(0.000)	-				(0.000)	-			
rr_t	0.404	0.938	1.000			0.416	0.861	1.000			0.485	0.967	1.000			0.556	0.967	1.000		
p value	(0.000)	(0.000)	-			(0.000)	(0.000)	-			(0.000)	(0.000)	-			(0.000)	(0.000)	-		
pv_t	0.403	0.930	0.920	1.000		0.398	0.943	0.839	1.000		0.491	0.969	0.935	1.000		0.612	0.963	0.936	1.000	
p value	(0.000)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-	
bv_t	0.379	0.896	0.907	0.951	1.000	0.402	0.958	0.860	0.956	1.000	0.466	0.983	0.964	0.964	1.000	0.551	0.988	0.959	0.961	1.000
p value	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-
			HC					IND					IT					MAR		
vol_t	1.000					1.000					1.000					1.000				
p value	-					-					-					-				
rv_t	0.345	1.000				0.486	1.000				0.341	1.000				0.427	1.000			
p value	(0.000)	-				(0.000)	-				(0.000)	-				(0.000)	-			
rr_t	0.371	0.927	1.000			0.475	0.919	1.000			0.345	0.937	1.000			0.472	0.963	1.000		
p value	(0.000)	(0.000)	-			(0.000)	(0.000)	-			(0.000)	(0.000)	-			(0.000)	(0.000)	-		
pv_t	0.355	0.954	0.906	1.000		0.518	0.959	0.882	1.000		0.356	0.974	0.924	1.000		0.474	0.965	0.940	1.000	
p value	(0.000)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-	
bv_t	0.352	0.971	0.927	0.953	1.000	0.491	0.974	0.908	0.953	1.000	0.338	0.982	0.933	0.967	1.000	0.434	0.982	0.961	0.960	1.000
p value	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-
			TEL					UTL					OVERALL							
vol_t	1.000					1.000					1.000									
p value	-					-					-									
rv_t	0.379	1.000				0.215	1.000				0.394	1.000								
p value	(0.000)	-				(0.001)	-				(0.000)	-								
rr_t	0.376	0.821	1.000			0.263	0.945	1.000			0.416	0.924	1.000							
p value	(0.000)	(0.000)	-			(0.000)	(0.000)	-			(0.000)	(0.000)	-							
pv_t	0.420	0.940	0.788	1.000		0.229	0.948	0.922	1.000		0.426	0.955	0.899	1.000						
p value	(0.000)	(0.000)	(0.000)	-		(0.002)	(0.000)	(0.000)	-		(0.000)	(0.000)	(0.000)	-						
bv_t	0.400	0.941	0.838	0.956	1.000	0.226	0.956	0.937	0.956	1.000	0.404	0.963	0.920	0.958	1.000					
p value	(0.000)	(0.000)	(0.000)	(0.000)	-	(0.001)	(0.000)	(0.000)	(0.000)	-	(0.000)	(0.000)	(0.000)	(0.000)	-					

Note: vol_t = Trading Volume, rv_t = Realized Variance, rr_t = Realized Range, pv_t = Realized Power Variation, bv_t = Realized Bipower Variation.

Table 1.7b: Correlation metrics II

	vol_t	rv_t	J_t	cv_t	vol_t	rv_t	J_t	cv_t	vol_t	rv_t	J_t	cv_t	vol_t	rv_t	J_t	cv_t
	CD				CS				ENG				FIN			
vol_t	**1.000**				**1.000**				**1.000**				**1.000**			
p value	-				-				-				-			
rv_t	**0.360**	**1.000**			**0.385**	**1.000**			**0.456**	**1.000**			**0.548**	**1.000**		
p value	(0.000)	-			(0.000)	-			(0.000)	-			(0.000)	-		
J_t	**-0.024**	**0.092**	**1.000**		**-0.020**	**0.090**	**1.000**		**-0.085**	**0.024**	**1.000**		**-0.038**	**0.019**	**1.000**	
p value	(-0.312)	(0.151)	-		(0.109)	(0.037)	-		(0.136)	(0.404)	-		(0.162)	(0.249)	-	
cv_t	**0.361**	**1.000**	**0.071**	**1.000**	**0.388**	**0.999**	**0.043**	**1.000**	**0.458**	**1.000**	**0.005**	**1.000**	**0.549**	**1.000**	**0.009**	**1.000**
p value	(0.000)	(0.000)	(0.357)	-	(0.000)	(0.000)	(0.137)	-	(0.000)	(0.000)	(0.141)	-	(0.000)	(0.000)	(0.432)	-
	HC				IND				IT				MAR			
vol_t	**1.000**				**1.000**				**1.000**				**1.000**			
p value	-				-				-				-			
rv_t	**0.345**	**1.000**			**0.486**	**1.000**			**0.341**	**1.000**			**0.427**	**1.000**		
p value	(0.000)	-			(0.000)	-			(0.000)	-			(0.000)	-		
J_t	**0.015**	**0.100**	**1.000**		**-0.040**	**0.041**	**1.000**		**-0.025**	**-0.010**	**1.000**		**-0.114**	**0.009**	**1.000**	
p value	(0.275)	(0.064)	-		(0.391)	(0.119)	-		(0.421)	(0.346)	-		(0.187)	(0.156)	-	
cv_t	**0.345**	**0.999**	**0.066**	**1.000**	**0.488**	**0.999**	**0.011**	**1.000**	**0.341**	**1.000**	**-0.029**	**1.000**	**0.428**	**1.000**	**-0.008**	**1.000**
p value	(0.000)	(0.000)	(0.076)	-	(0.000)	(0.000)	(0.441)	-	(0.000)	(0.000)	0.217	-	(0.000)	(0.000)	(0.175)	-
	TEL				UTL				OVERALL							
vol_t	**1.000**				**1.000**				**1.000**							
p value	(0.000)				(0.000)				(0.000)							
rv_t	**0.379**	**1.000**			**0.215**	**1.000**			**0.394**	**1.000**						
p value	(0.000)	-			(0.001)	-			(0.000)	-						
J_t	**-0.038**	**0.040**	**1.000**		**-0.036**	**0.055**	**1.000**		**-0.040**	**0.046**	**1.000**					
p value	(0.204)	(0.272)	-		(0.226)	(0.217)	-		(0.242)	(0.201)	-					
cv_t	**0.380**	**1.000**	**0.013**	**1.000**	**0.217**	**0.999**	**0.028**	**1.000**	**0.395**	**1.000**	**0.021**	**1.000**				
p value	(0.000)	(0.000)	(0.359)	-	(0.001)	(0.000)	0.177	-	(0.000)	(0.000)	(0.251)	-				

Note: vol_t= Trading Volume, rv_t= Realized Variance, J_t = Jumps, cv_t = Continuous Variance.

Table 1.8: Long memory test

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	Overall
vol_t	0.342	0.347	0.419	0.443	0.358	0.379	0.334	0.445	0.419	0.389	0.388
rv_t	0.398	0.404	0.402	0.402	0.407	0.421	0.439	0.420	0.396	0.350	0.404
cv_t	0.399	0.405	0.402	0.402	0.407	0.429	0.439	0.420	0.396	0.350	0.405
rr_t	0.422	0.388	0.408	0.411	0.415	0.420	0.442	0.431	0.406	0.314	0.406
pv_t	0.449	0.441	0.433	0.438	0.440	0.445	0.455	0.446	0.444	0.404	0.440
bv_t	0.428	0.418	0.404	0.403	0.411	0.421	0.437	0.420	0.401	0.372	0.411

Note: The table reports the mean Robinson's *d* statistics across sectors and overall. All the values are significant at 5% level.

Table 1.9: volatility-volume relation results

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	OVERALL
						Realized Variance (rv_t)					
$\overline{\gamma}_{it}$	0.652	0.884	0.862	0.754	0.366	0.546	0.148	0.502	0.987	6.524	1.222
$\overline{se}_{\gamma}$	0.248	0.298	0.313	0.276	0.066	0.106	0.031	0.107	0.395	3.755	0.559
% $\overline{R}^2$	45.209	47.722	63.934	58.489	48.804	56.552	58.540	61.338	41.284	45.708	52.758
% Significant	90	100	100	100	100	100	100	100	90	80	96
						Realized Range (rr_t)					
$\overline{\gamma}_{it}$	0.492	0.742	0.754	0.684	0.336	0.500	0.147	0.440	0.784	5.305	1.018
$\overline{se}_{\gamma}$	0.150	0.241	0.279	0.241	0.060	0.106	0.032	0.094	0.260	3.109	0.457
% $\overline{R}^2$	53.103	48.450	64.393	62.929	51.572	53.684	60.287	66.735	47.463	37.977	54.659
% Significant	90	100	100	100	100	100	100	100	90	80	96
						Realized Power Variation (pv_t)					
$\overline{\gamma}_{it}$	0.488	0.845	0.737	0.540	0.368	0.585	0.161	0.485	0.953	2.691	0.785
$\overline{se}_{\gamma}$	0.102	0.218	0.177	0.153	0.052	0.079	0.027	0.075	0.253	1.283	0.242
% $\overline{R}^2$	65.675	64.767	73.242	71.911	64.822	70.210	70.751	72.310	59.947	64.193	67.783
% Significant	100	100	100	100	100	100	100	100	90	90	98
						Realized Bipower Variation (bv_t)					
$\overline{\gamma}_{it}$	0.416	0.699	0.792	0.679	0.344	0.529	0.140	0.473	0.865	6.831	1.177
$\overline{se}_{\gamma}$	0.082	0.222	0.284	0.256	0.058	0.106	0.031	0.100	0.353	4.064	0.556
% $\overline{R}^2$	53.658	52.003	64.847	59.713	51.452	57.629	58.568	62.895	47.391	51.795	55.995
% Significant	100	100	100	100	100	100	100	100	90	90	98

Note: Volume is rescaled by dividing by 1,000,000. ρ_{ij} measures the persistence of volatility shock at lag *j*. M_t is the Monday dummy. $\overline{\gamma}_{it}$ is equally-weighted cross sectional mean coefficients for trading volume and $\overline{se}_{\gamma}$ is average Newey-West standard error. The $\overline{R}^2$ is the mean value of 100 stocks. The last row reports the percentage of $\hat{\gamma}_i$ coefficients which are statistically significantly from zero at 5% level.

Table1.10a: Distributional test of volatility (sector average)

	IG	Lognormal	Normal	IG	Lognormal	Normal	IG	Lognormal	Normal	IG	Lognormal	Normal	IG	Lognormal	Normal
	CD			CS			ENG			FIN			HC		
rv_t	0.068	0.072	0.252	0.087	0.081	0.239	0.066	0.055	0.292	0.071	0.048	0.362	0.075	0.074	0.220
rr_t	0.070	0.076	0.251	0.101	0.089	0.295	0.080	0.056	0.311	0.087	0.059	0.371	0.089	0.080	0.245
pv_t	0.074	0.075	0.175	0.087	0.081	0.181	0.062	0.051	0.209	0.056	0.049	0.276	0.080	0.077	0.175
bv_t	0.069	0.073	0.239	0.088	0.081	0.241	0.069	0.053	0.294	0.069	0.050	0.361	0.078	0.072	0.223
	IND			IT			MAR			TEL			UTL		
rv_t	0.054	0.042	0.264	0.083	0.088	0.227	0.094	0.065	0.279	0.056	0.049	0.300	0.083	0.050	0.406
rr_t	0.067	0.044	0.291	0.094	0.099	0.230	0.122	0.087	0.308	0.093	0.058	0.303	0.120	0.055	0.415
pv_t	0.050	0.046	0.189	0.086	0.086	0.173	0.087	0.072	0.213	0.056	0.055	0.203	0.060	0.049	0.255
bv_t	0.059	0.045	0.268	0.079	0.084	0.231	0.103	0.068	0.291	0.085	0.045	0.079	0.095	0.051	0.417

Note: KS denotes Kolmogorov-Smirnov test on the distributions of Inverse Gaussian, lognormal and normal, with the critical value of 0.0258 at 5% level. * denotes $p>0.05$ and * denotes $p>0.01$.

Table 1.10b: Distributional test of volatility (most and least active stocks)

KS Test	IG	Lognormal	Normal	IG	Lognormal	Normal	IG	Lognormal	Normal	IG	Lognormal	Normal
	Consumer Discretionary						Consumer Staples					
	HD			MAR			WMT			BFB		
rv_t	0.038	0.047	0.251	0.014**	0.030*	0.256	0.060	0.068	0.252	0.079	0.053	0.369
rr_t	0.053	0.060	0.273	0.027*	0.044	0.289	0.078	0.081	0.264	0.081	0.069	0.356
pv_t	0.046	0.049	0.178	0.023**	0.034	0.185	0.066	0.067	0.181	0.059	0.044	0.203
bv_t	0.038	0.047	0.257	0.019**	0.030*	0.264	0.056	0.066	0.252	0.058	0.041	0.322
	Energy						Financials					
	XOM			TE			C			TRV		
rv_t	0.069	0.046	0.318	0.065	0.043	0.325	0.053	0.042	0.398	0.065	0.047	0.345
rr_t	0.087	0.055	0.341	0.085	0.053	0.327	0.053	0.053	0.391	0.109	0.074	0.338
pv_t	0.064	0.046	0.222	0.061	0.044	0.226	0.050	0.050	0.316	0.064	0.049	0.246
bv_t	0.070	0.048	0.325	0.066	0.042	0.317	0.049	0.040	0.399	0.071	0.053	0.348
	Health Care						Industrials					
	PFE			HUM			GE			GD		
rv_t	0.045	0.041	0.245	0.050	0.049	0.248	0.040	0.044	0.326	0.065	0.054	0.270
rr_t	0.062	0.055	0.259	0.058	0.066	0.260	0.049	0.050	0.325	0.068	0.052	0.316
pv_t	0.046	0.041	0.165	0.062	0.064	0.186	0.040	0.0470	0.237	0.064	0.054	0.191
bv_t	0.041	0.038	0.242	0.048	0.062	0.251	0.042	0.046	0.327	0.069	0.056	0.286
	Information Technology						Materials					
	MSFT			XRX			AA			AKS		
rv_t	0.030*	0.048	0.241	0.039	0.049	0.309	0.078	0.061	0.314	0.053	0.041	0.284
rr_t	0.052	0.063	0.237	0.056	0.062	0.322	0.099	0.069	0.327	0.065	0.049	0.297
pv_t	0.042	0.049	0.173	0.045	0.045	0.217	0.077	0.061	0.226	0.052	0.042	0.195
bv_t	0.031*	0.050	0.245	0.036	0.043	0.320	0.081	0.058	0.321	0.054	0.044	0.296
	Telecommunications Services						Utilities					
	S			BT			DUK			OKE		
rv_t	0.035	0.030*	0.320	0.048	0.061	0.233	0.055	0.041	0.336	0.068	0.049	0.373
rr_t	0.045	0.059	0.332	0.054	0.053	0.264	0.096	0.058	0.354	0.124	0.075	0.437
pv_t	0.036	0.029*	0.229	0.055	0.058	0.175	0.050	0.041	0.234	0.064	0.049	0.241
bv_t	0.035	0.032	0.319	0.046	0.058	0.240	0.056	0.038	0.336	0.070	0.047	0.325

Note: KS denotes Kolmogorov-Smirnov test on the distributions of Inverse Gaussian, lognormal and normal, with the critical value of 0.0258 at 5% level. * denotes $p>0.05$ and * denotes $p>0.01$. The first and third columns from left are the most actively traded stocks in every sector and the second and fourth columns are the least actively traded stocks in every sector.

Table 1.11: Daily distributions of returns (overall)

	Raw Returns Significance			Demeaned Returns Significance		
	1%	5%	10%	1%	5%	10%
$r_t/\sqrt{rv_t}$	39	60	71	40	60	72
$r_t/\sqrt{rr_t}$	35	48	57	35	49	59
$r_t/\sqrt{pv_t}$	21	34	49	19	31	46
$r_t/\sqrt{bv_t}$	25	42	54	25	42	55

Note: The table reports the number of stocks (out of 100) for which the null hypothesis of normality of return is rejected based on the Jarque-Bera test when daily returns are standardized by volatility measures. r_t refers the daily return or demeaned returns, while rv_t, rr_t, pv_t, bv_t denote the realized variance, realized range, realized power variation and realized bipower variation respectively.

Table 1.12a: Daily distribution of raw returns (sectors)

	Daily Raw Return of 100 stocks											
	Significance			Significance			Significance			Significance		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Series	CD			CS			ENG			FIN		
$r_t/\sqrt{rv_t}$	6	7	7	2	2	4	1	5	7	5	8	9
$r_t/\sqrt{rr_t}$	7	8	8	4	4	6	1	1	3	1	3	5
$r_t/\sqrt{pv_t}$	2	3	5	2	4	9	0	2	2	5	5	5
$r_t/\sqrt{bv_t}$	5	6	8	2	3	3	0	1	4	4	5	7
Series	HC			IND			IT			MAR		
$r_t/\sqrt{rv_t}$	4	6	8	3	6	6	7	8	9	5	8	9
$r_t/\sqrt{rr_t}$	2	4	4	4	5	5	4	5	7	5	8	8
$r_t/\sqrt{pv_t}$	1	1	3	2	3	4	1	2	4	2	4	5
$r_t/\sqrt{bv_t}$	2	3	3	2	4	5	4	9	9	3	6	8
Series	TEL			UTL								
$r_t/\sqrt{rv_t}$	4	7	9	2	3	3						
$r_t/\sqrt{rr_t}$	4	5	5	3	5	6						
$r_t/\sqrt{pv_t}$	3	6	7	3	4	5						
$r_t/\sqrt{bv_t}$	2	3	5	1	2	2						

Note: The table reports the number of stocks in each sector (out of 10 in each sector) for which the null hypothesis of normality of return is rejected based on the Jarque-Bera test when daily returns are standardized by volatility measures. r_t refers the daily return, while rv_t, rr_t, pv_t, bv_t denote the realized variance, realized range, realized power variation and realized bipower variation respectively.

Table 1.12b: Daily distributions of demeaned returns (sectors)

	Daily Demeaned Return of 100 stocks											
	Significance			Significance			Significance			Significance		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Series	CD			CS			ENG			FIN		
$r_t/\sqrt{rv_t}$	6	7	7	2	2	5	2	5	7	5	8	9
$r_t/\sqrt{rr_t}$	7	8	8	4	5	6	1	1	4	1	2	5
$r_t/\sqrt{pv_t}$	1	4	6	2	4	8	0	2	2	5	5	5
$r_t/\sqrt{bv_t}$	5	6	7	3	3	3	0	1	5	3	5	8
Series	HC			IND			IT			MAR		
$r_t/\sqrt{rv_t}$	4	6	8	3	6	6	7	8	9	5	8	9
$r_t/\sqrt{rr_t}$	2	4	5	4	5	5	4	5	7	5	8	8
$r_t/\sqrt{pv_t}$	1	1	3	2	2	3	1	2	3	2	3	5
$r_t/\sqrt{bv_t}$	2	3	3	2	4	5	4	9	9	3	6	8
Series	TEL			UTL								
$r_t/\sqrt{rv_t}$	4	7	9	2	3	3						
$r_t/\sqrt{rr_t}$	4	5	5	3	6	6						
$r_t/\sqrt{pv_t}$	2	4	6	3	4	5						
$r_t/\sqrt{bv_t}$	2	3	5	1	2	2						

Note: The table reports the number of stocks in each sector (out of 10 in each sector) for which the null hypothesis of normality of demeaned return is rejected based on the Jarque-Bera test when daily returns are standardized by volatility measures. r_t refers the demeaned daily return, while rv_t, rr_t, pv_t, bv_t denote the realized variance, realized range, realized power variation and realized bipower variation respectively.

Table 1.13: Table of rankings

Sector	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
Realized Volatility Measures										
rv_t	6	10	3	2	7	8	4	1	5	9
rr_t	5	10	4	1	7	8	2	3	6	9
pv_t	5	10	4	3	7	8	2	1	6	9
bv_t	6	10	3	2	7	8	4	1	5	9
Market Activity Measure										
vol_t	5	9	7	2	3	4	1	8	6	1
Jumps										
$J_{t,j}$	5	2	8	9	6	7	10	3	1	4
J_t	2	4	8	5	9	3	10	7	1	6
Standardized Returns										
$r_t / \sqrt{rv_t}$	6	1	3	8	4	4	8	8	6	2
$r_t / \sqrt{rr_t}$	9	3	1	2	3	5	5	9	5	5
$r_t / \sqrt{pv_t}$	4	6	2	9	1	4	2	6	10	6
$r_t / \sqrt{bv_t}$	8	3	1	7	3	6	9	8	3	2
Robison's Long Memory										
"d" (rv_t)	8	5	6	7	4	2	1	3	9	10
"d" (rr_t)	3	9	7	6	5	4	2	1	8	10
"d" (pv_t)	2	6	9	8	7	4	1	3	5	10
"d" (bv_t)	2	5	7	8	6	3	1	4	9	10
Volatility-Volume Relation										
$\bar{R}^2(rv_t)$	9	7	1	4	6	5	3	2	10	8
$\bar{R}^2(rr_t)$	7	8	2	3	6	5	4	1	9	10
$\bar{R}^2(pv_t)$	6	8	1	3	7	4	5	2	10	9
$\bar{R}^2(bv_t)$	6	7	1	3	9	5	4	2	10	8

Note: This table reports the rankings of sectors according to realized volatility measures, trading volume, intraday and daily jump tests, number of rejections of normality on standardized returns, long memory test and the OLS results. 1 denotes the highest value (lowest rejection) and 10 denotes the lowest (highest rejection). $rv_t, rr_t, pv_t, bv_t, vol_t, J_t, J_{t,j}$ denote realized variance, realized range, realized power variation, realized bipower variation, trading volume, daily jumps and intraday jumps, respectively.

Figures

Figure 1.1a: Volatility signature plot of GS (the stock with the smallest number of jumps)

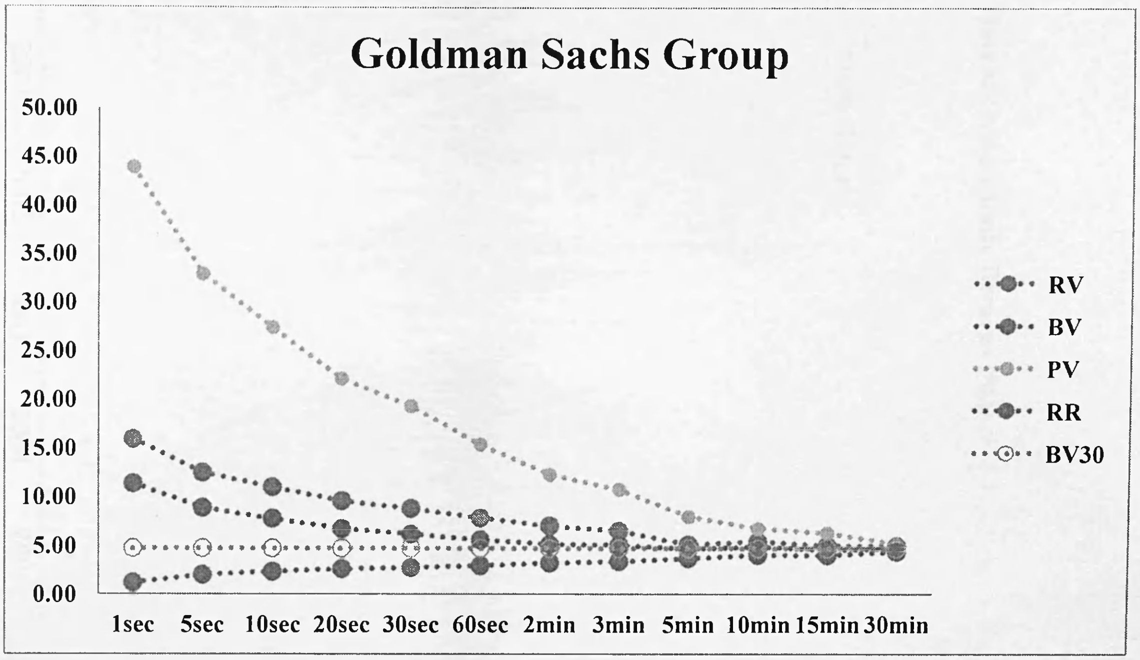


Figure 1.1b: Volatility signature plot of FTR (the stock with the largest number of jumps)

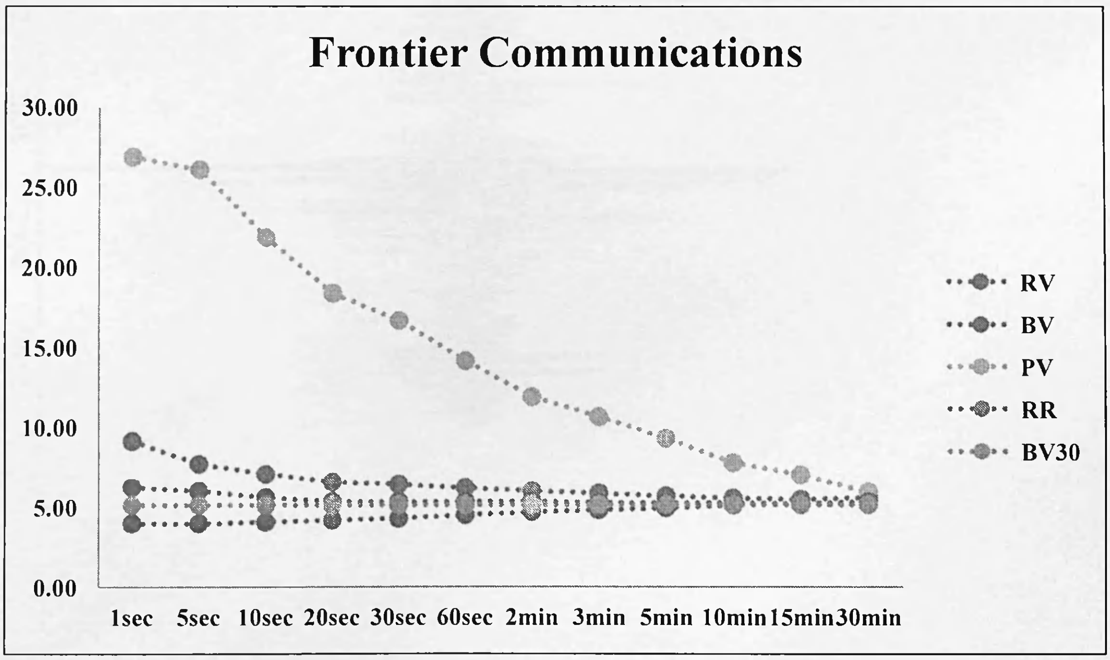
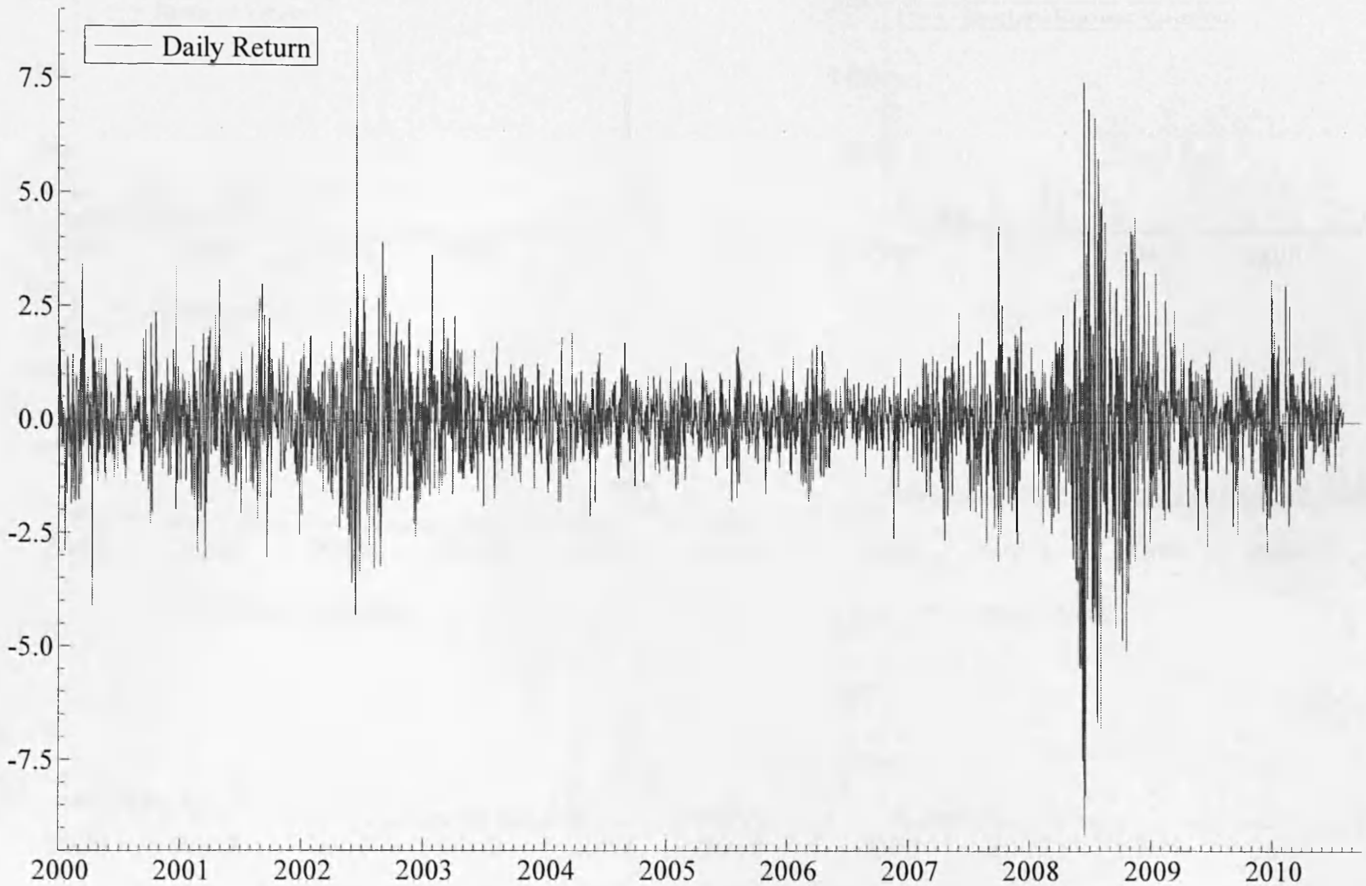
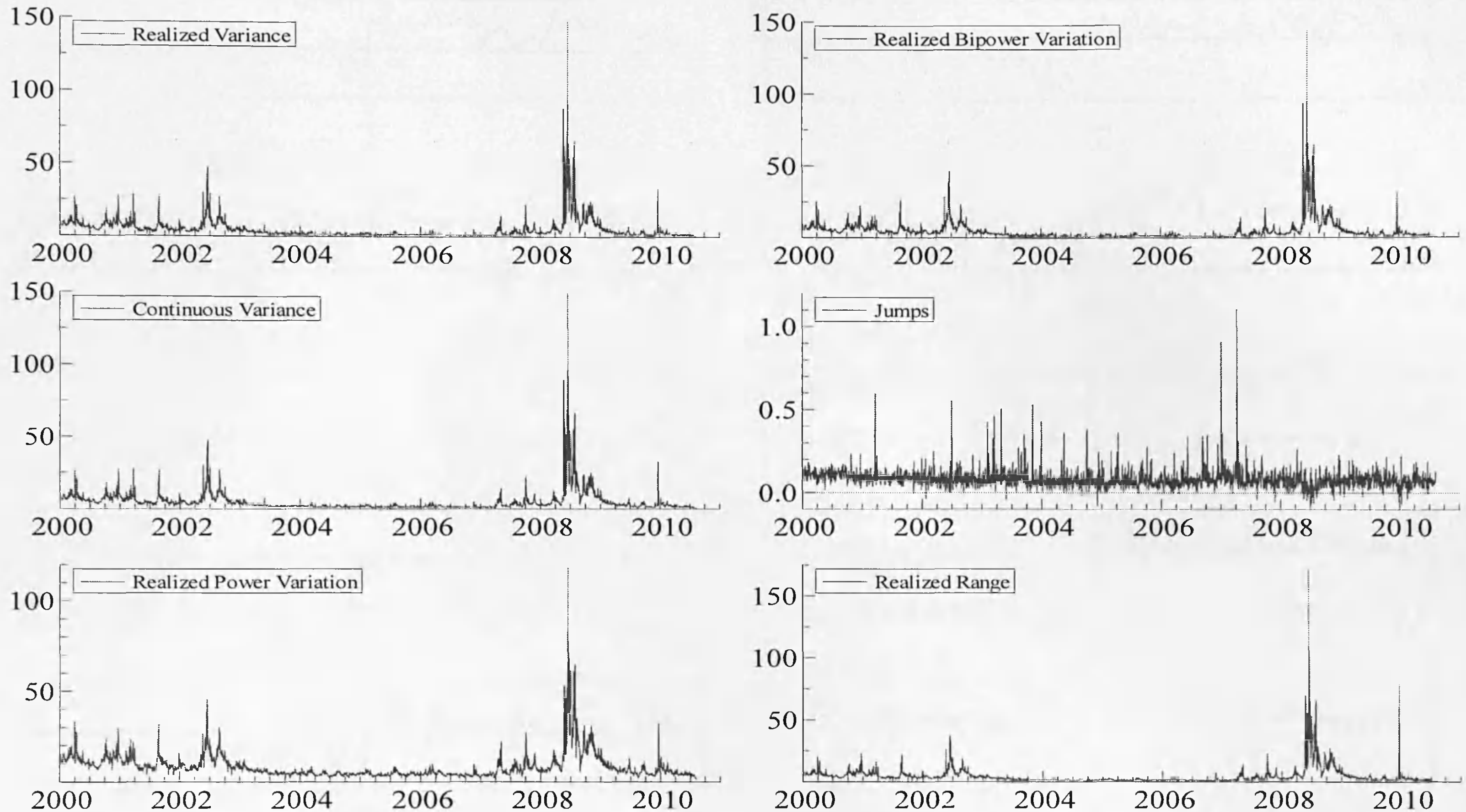


Figure 1.2: Plots of Overall Daily Returns (%)



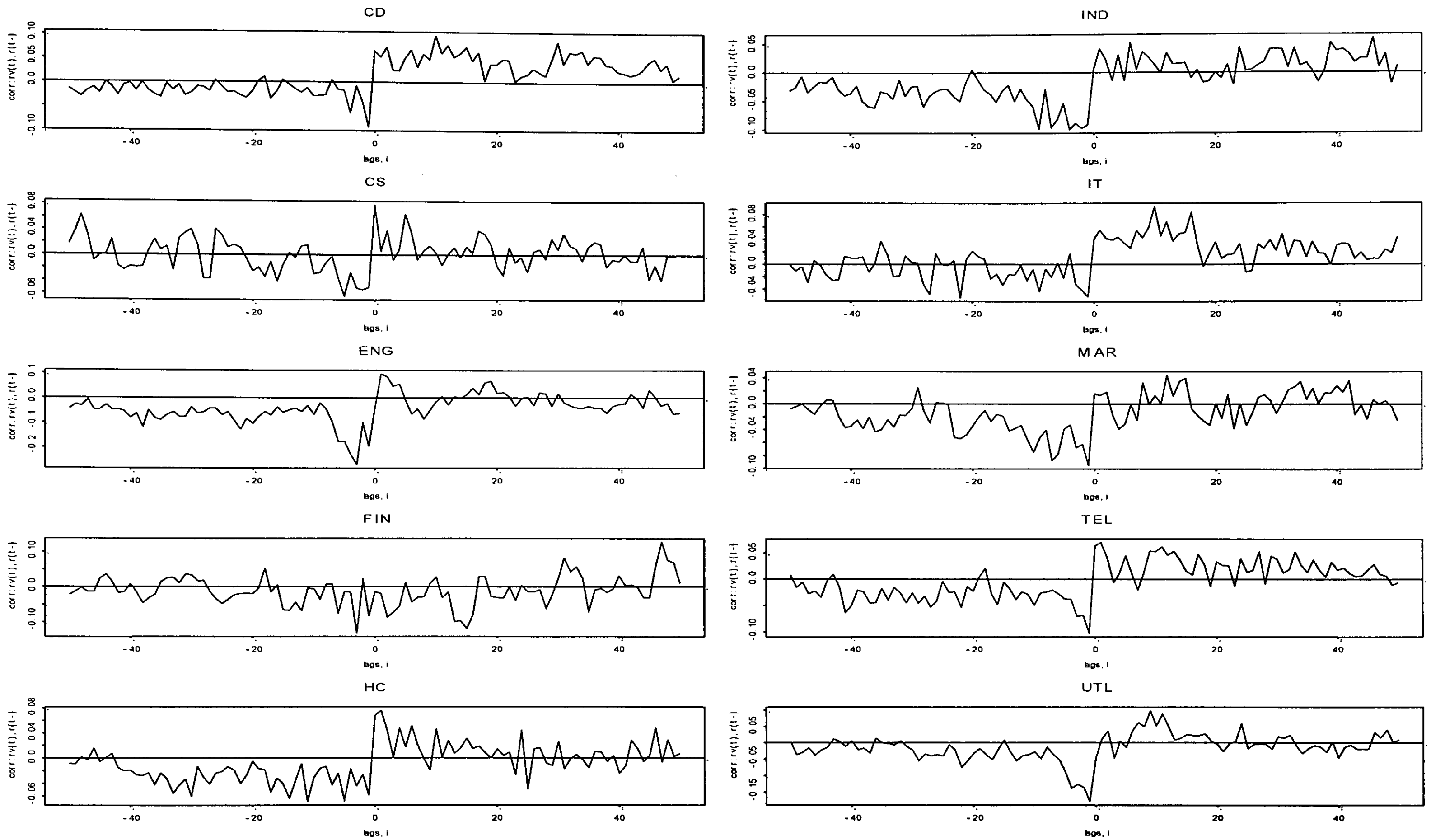
Note: This figure reports the overall sample average daily returns from 2000 to 2010 and the value of returns is reported in percentage.

Figure 1.3: Plots of overall average realized volatility measures, continuous variance and jumps



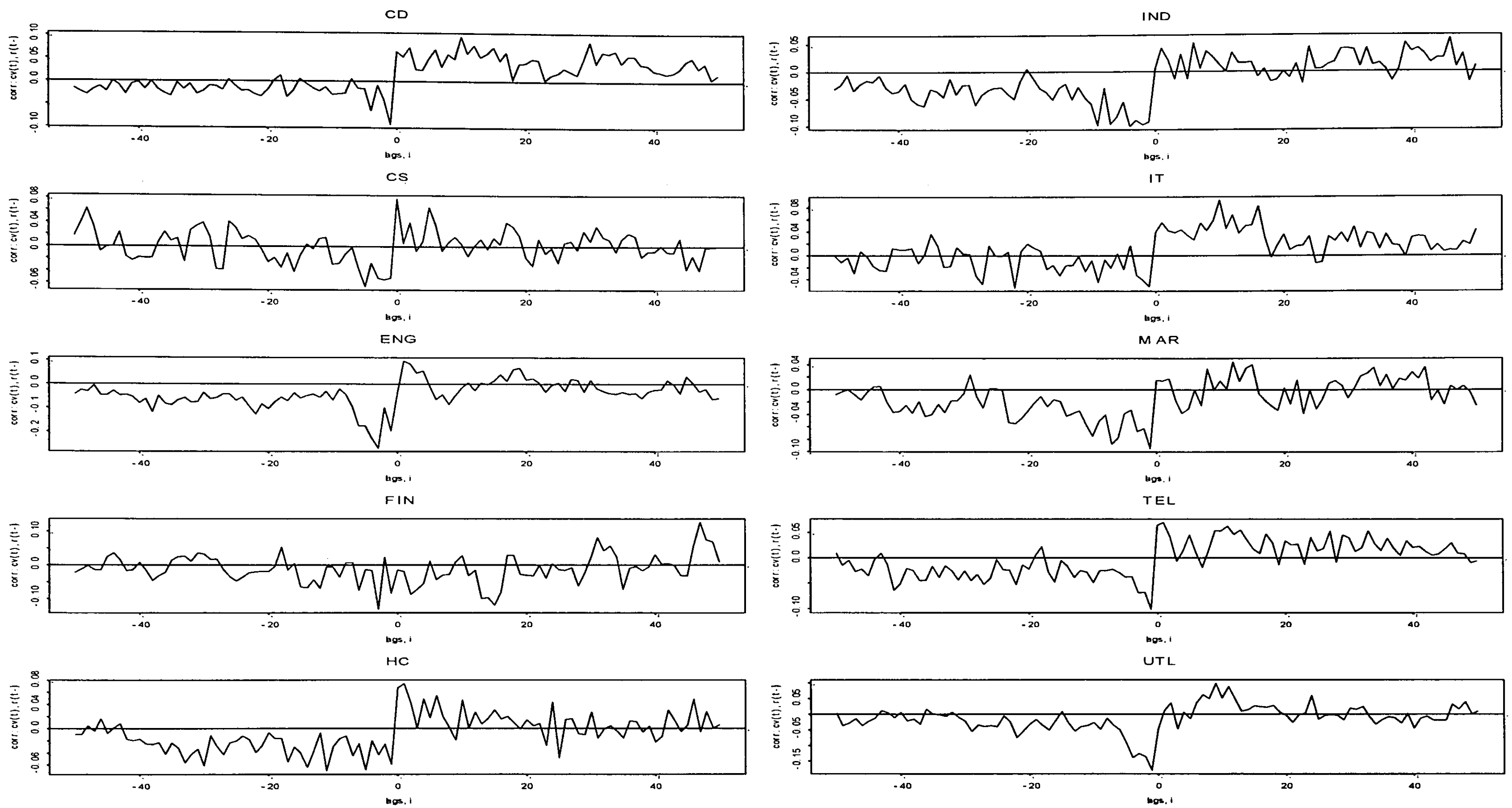
Note: This figures represent the overall sample average daily realized volatility measures, daily jumps and daily continuous variance from 2000 to 2010

Figure 1.4a: Leverage and feedback effect



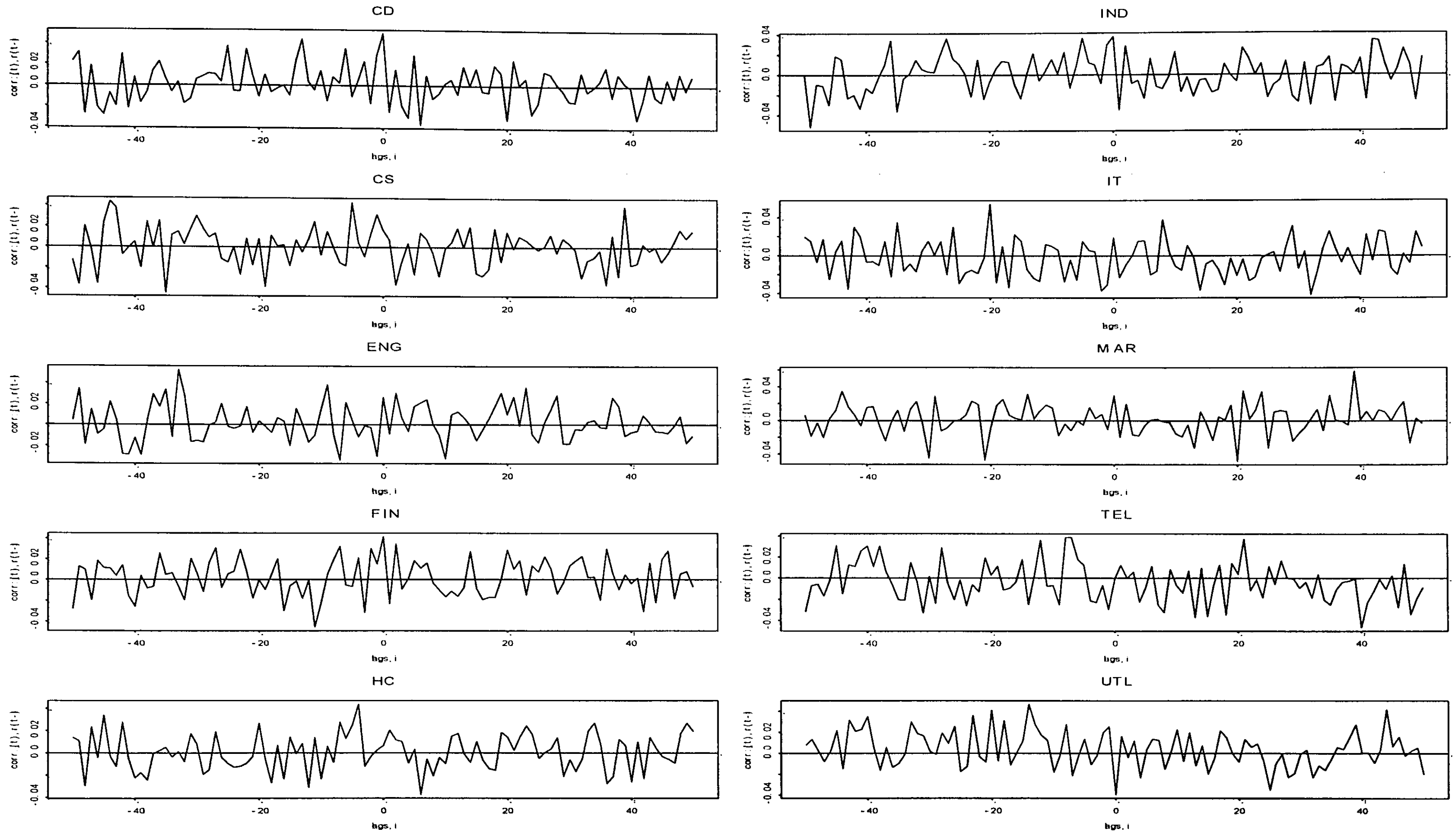
Note: The plots represent the sector median cross-correlations between current realized variances and one day lagged daily returns.

Figure 1.4b: Leverage and feedback effect



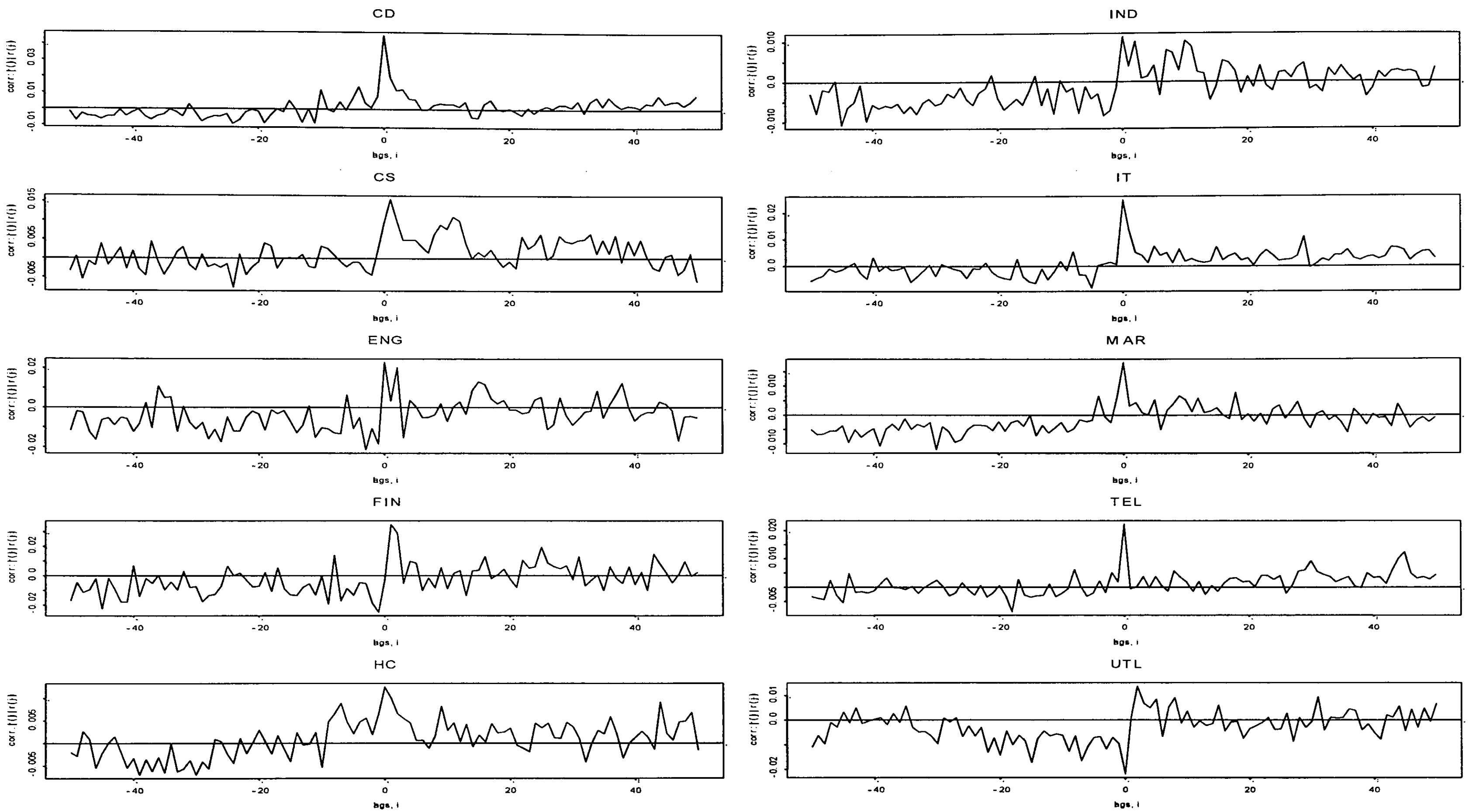
Note: The plots represent the sector-median cross-correlations between current continuous variances and one day lagged daily returns.

Figure 1.4c: Leverage and feedback effect



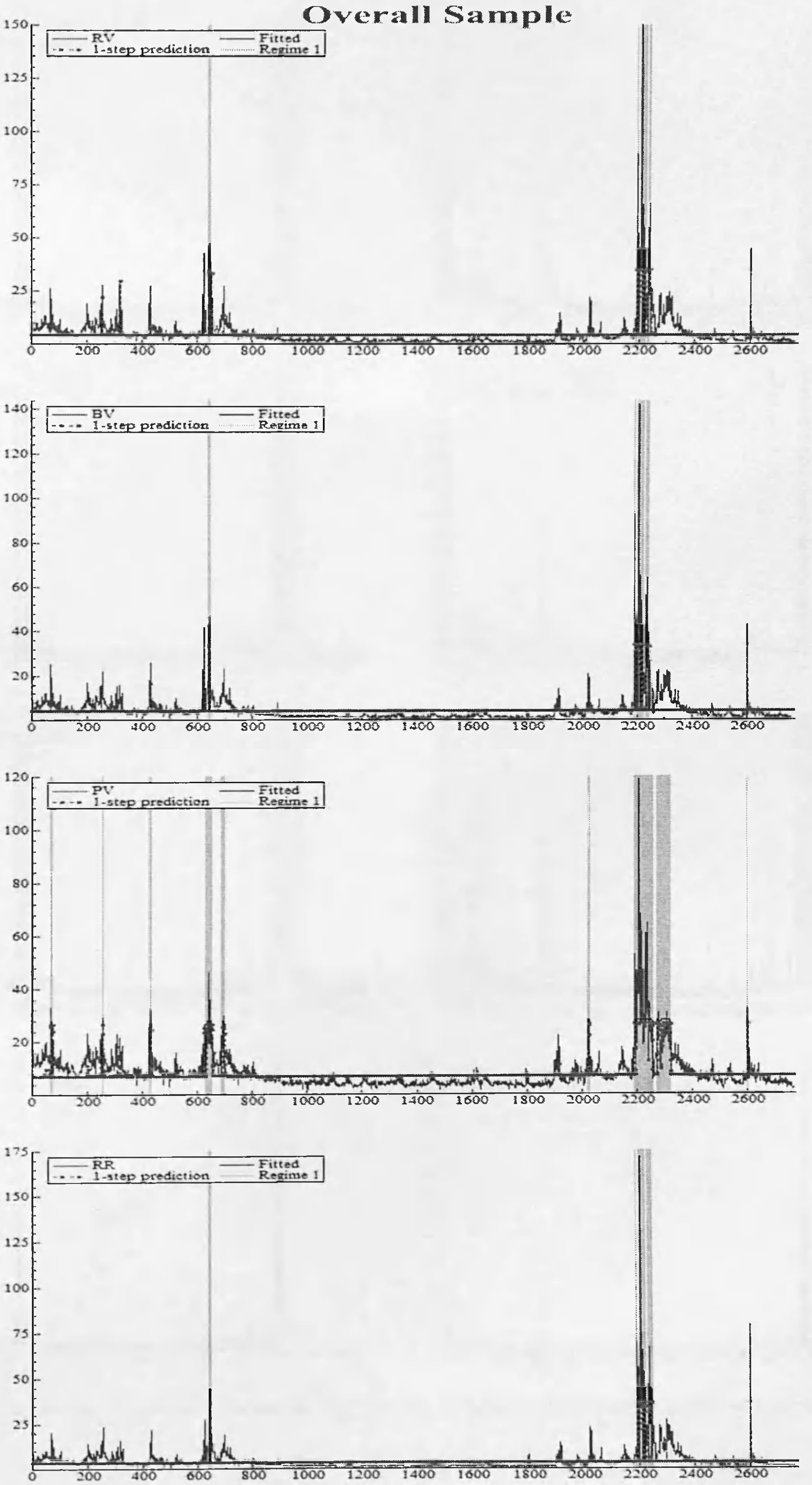
Note: The plots represent the sector-median cross-correlations between current jumps and one day lagged daily returns.

Figure 1.4d: The leverage and feedback effect



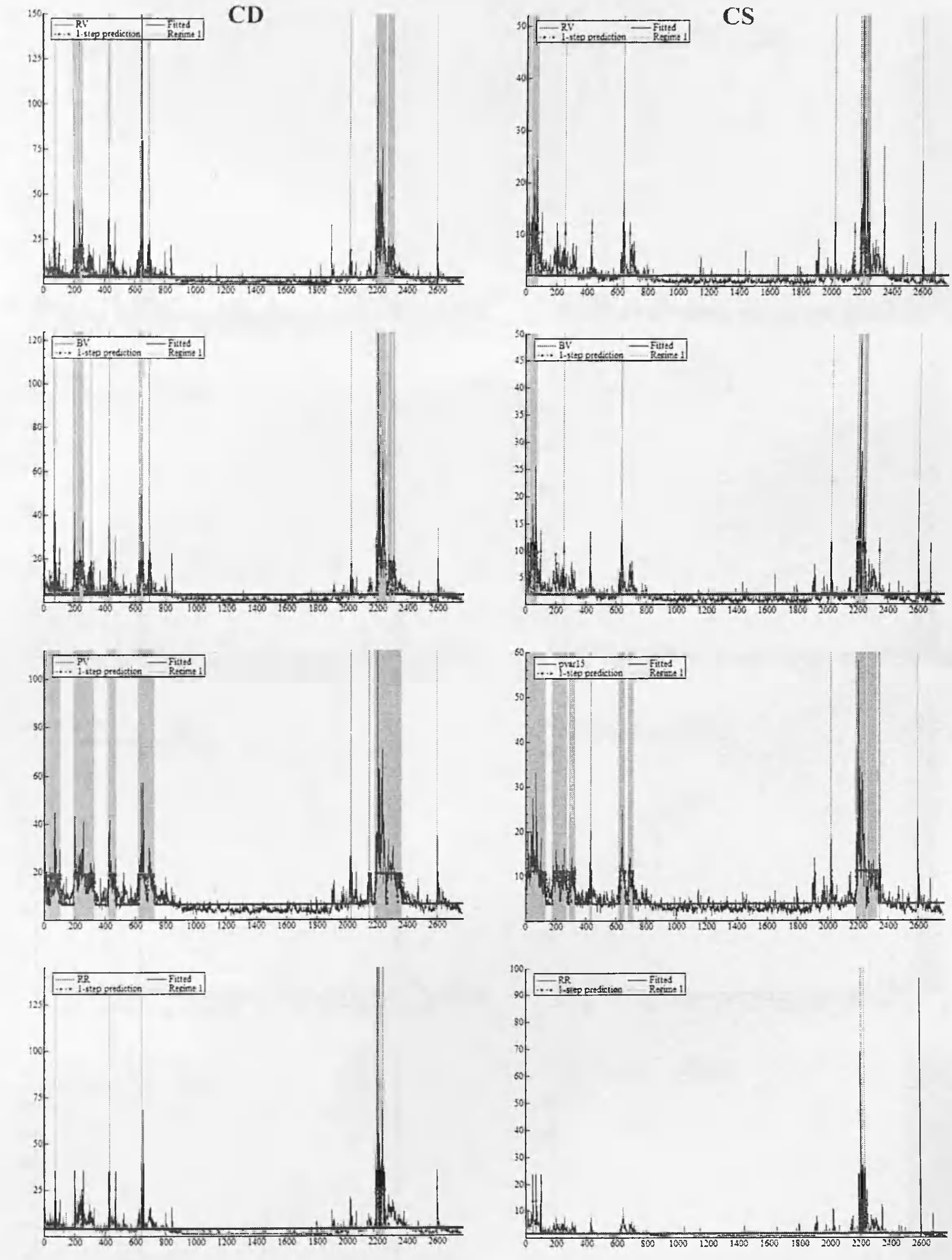
Note: The plots represent the sector-median cross-correlations between current intraday (5min) absolute returns and one period lagged intraday returns.

Figure 1.5a: Markov regime switching model on realized volatility measures (overall sample)



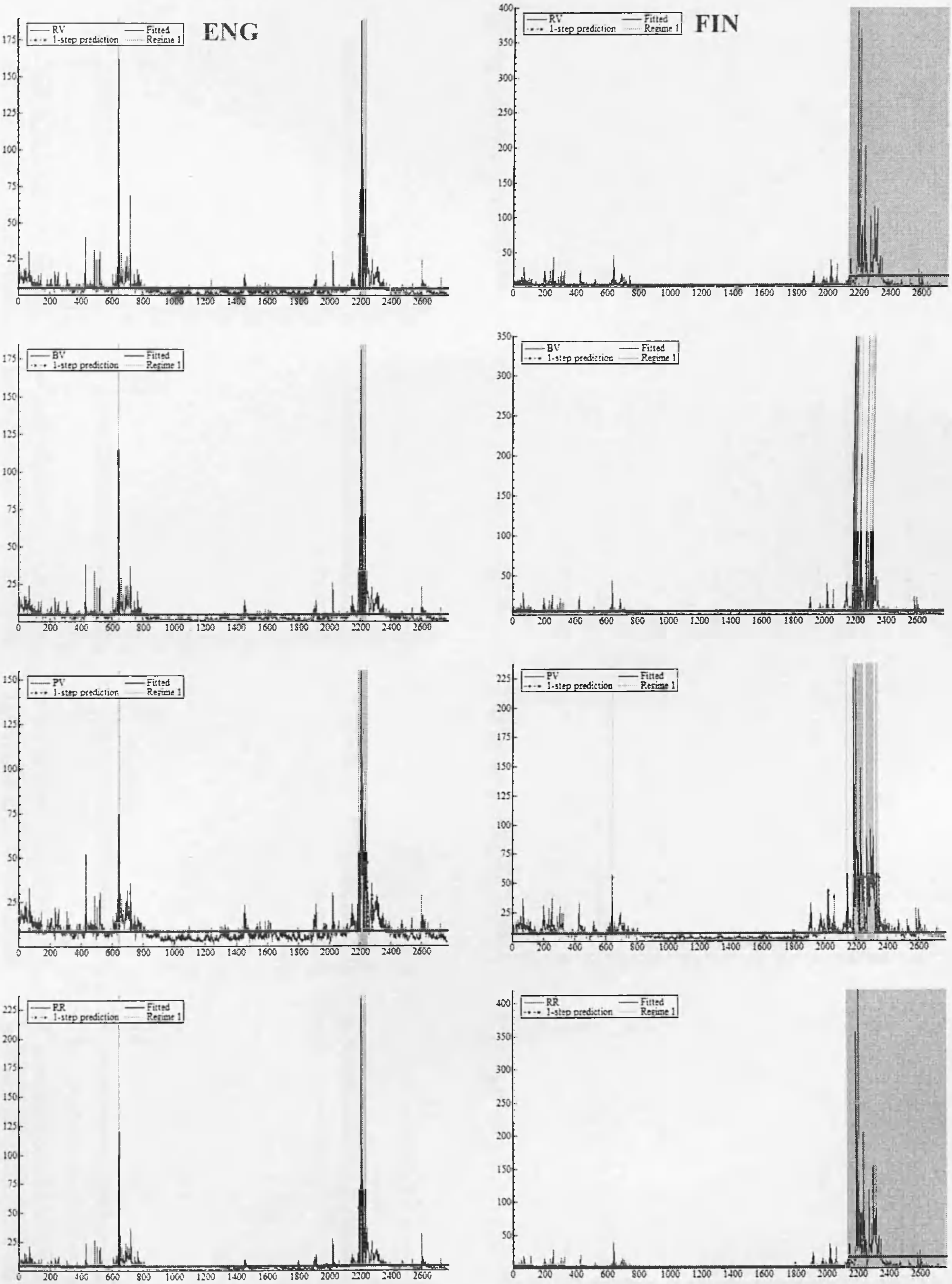
Note: From up to down: realized variance, realized bipower variation, realized power variation and realized range.

Figure 1.5b: Markov regime switching model on realized volatility measures (Sector CD and CS)



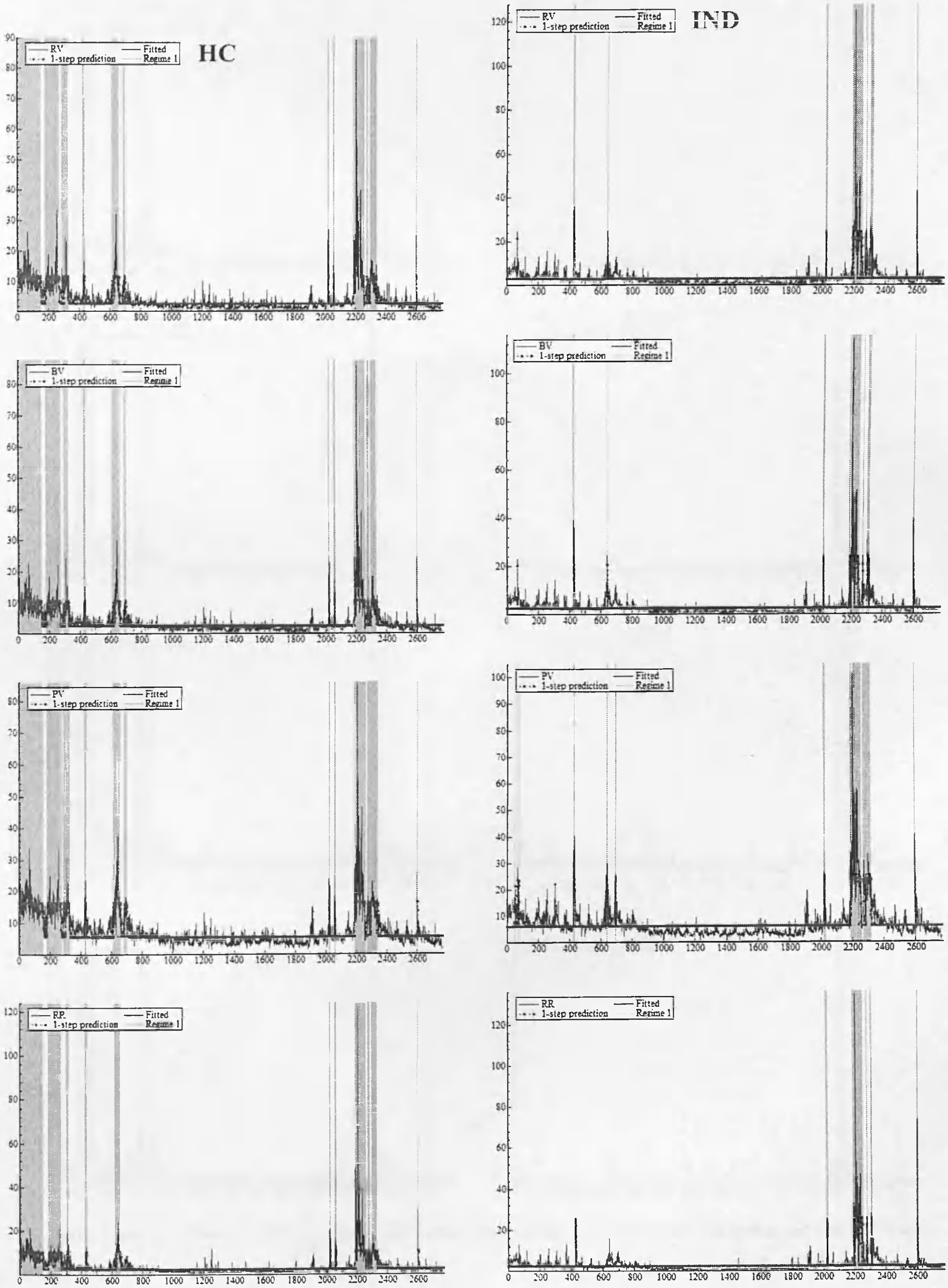
Note: From up to down: realized variance, realized bipower variation, realized power variation and realized range.

Figure 1.5c: Markov regime switching model on realized volatility measures (Sector ENG and FIN)



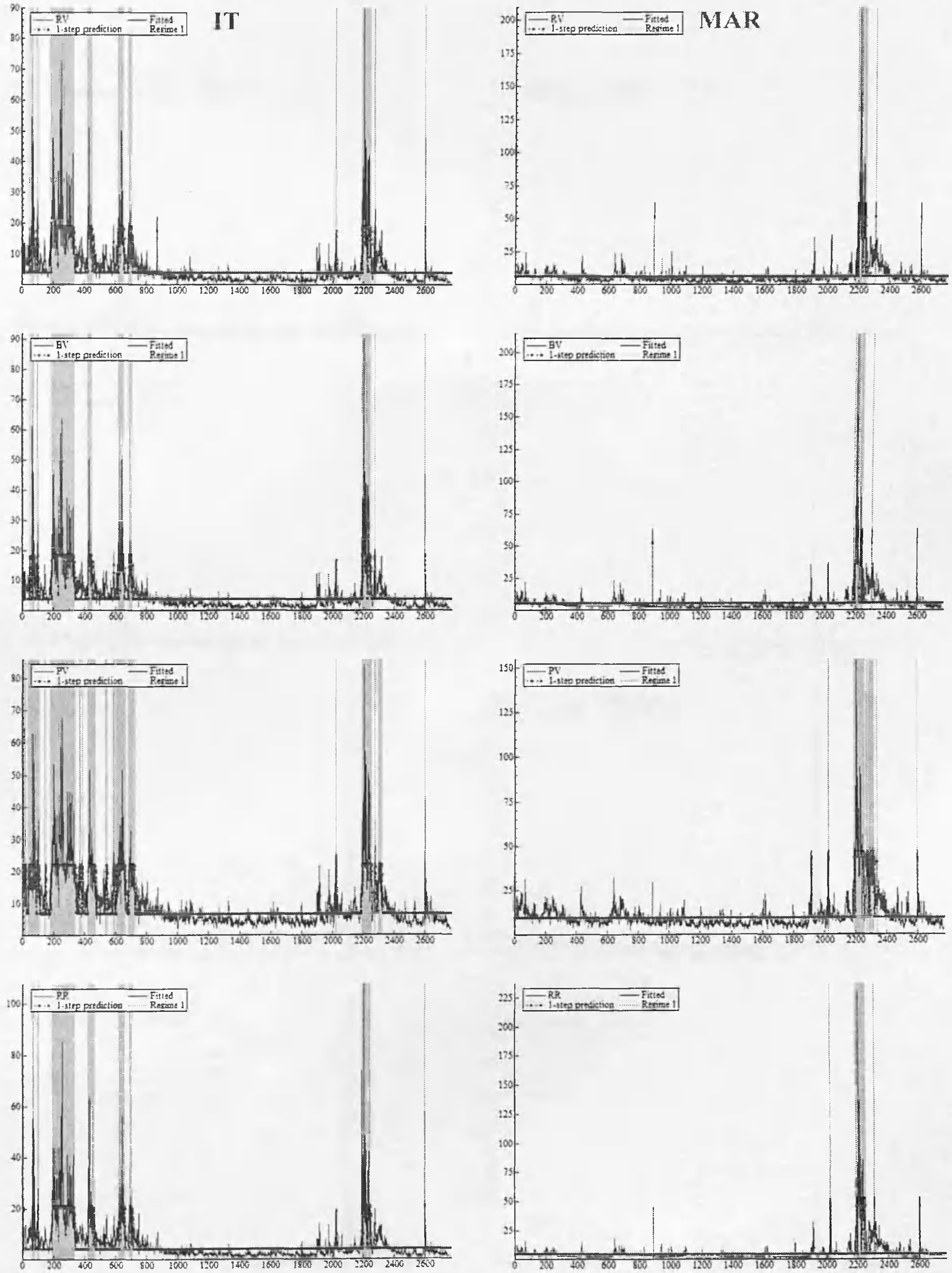
Note: From up to down: realized variance, realized bipower variation, realized power variation and realized range.

Figure 1.5d: Markov regime switching model on realized volatility measures (Sector HC and IND)



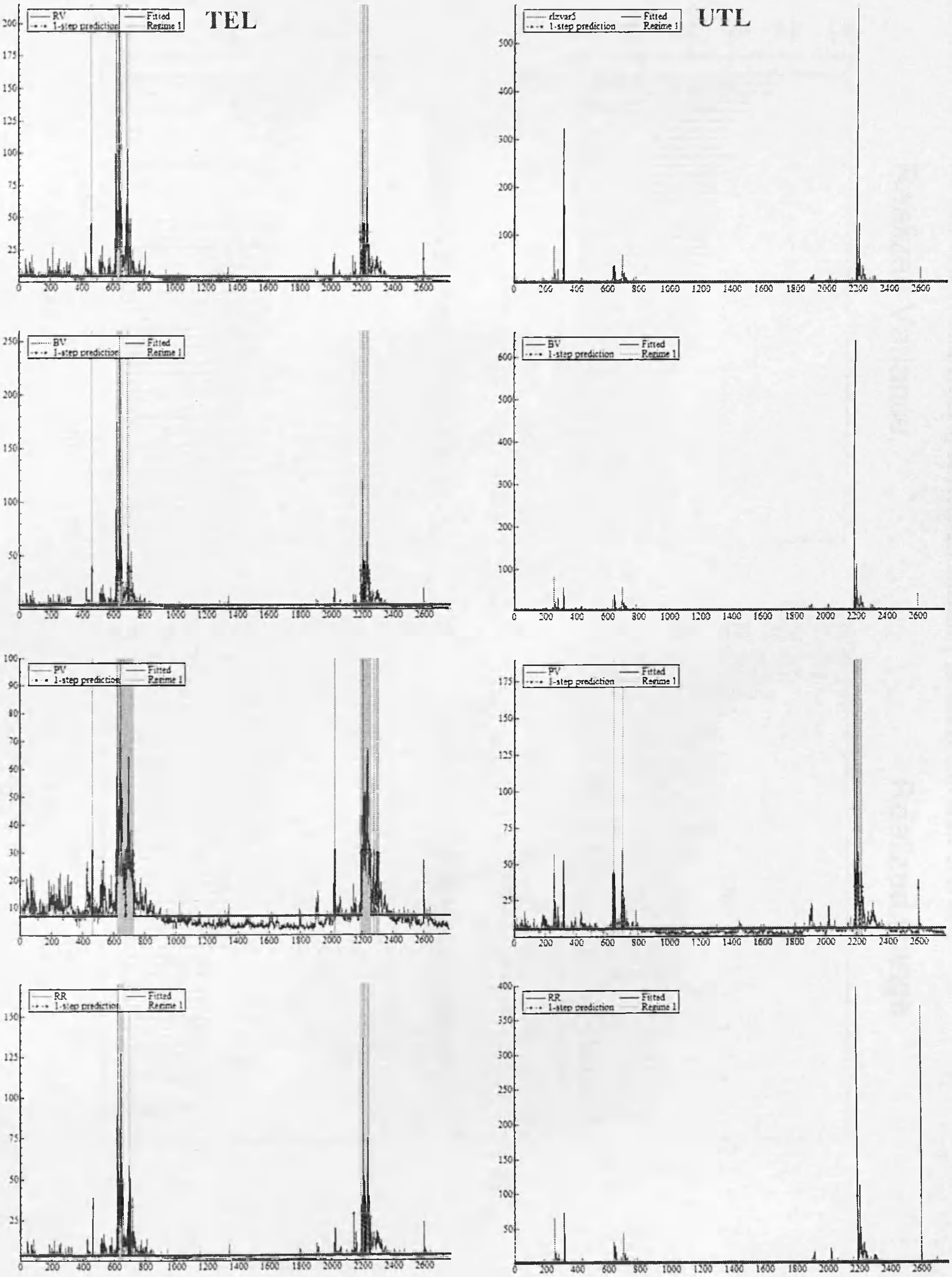
Note: From up to down: realized variance, realized bipower variation, realized power variation and realized range.

Figure 1.5e: Markov regime switching model on realized volatility measures (Sector IT and MAR)



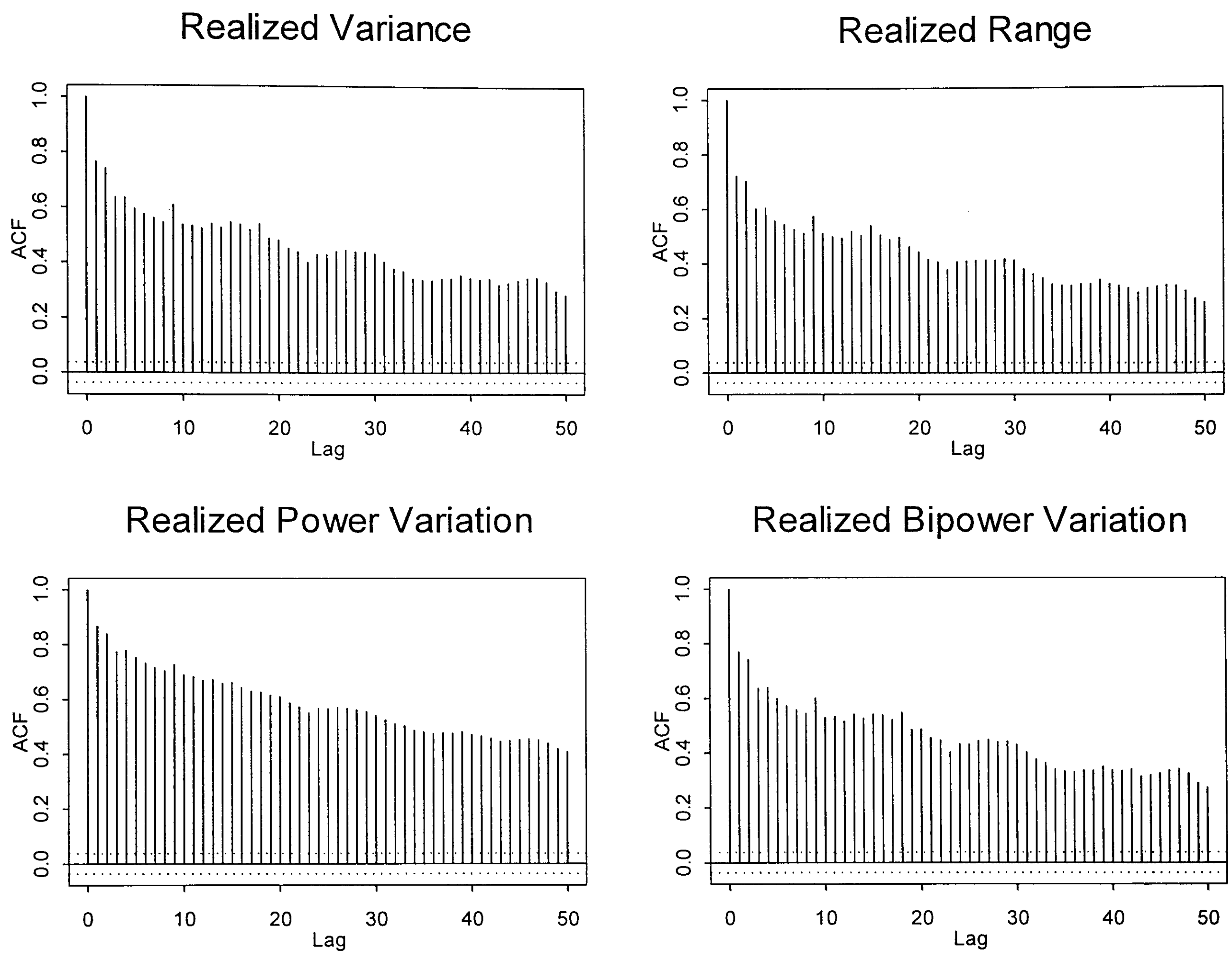
Note: From up to down: realized variance, realized bipower variation, realized power variation and realized range.

Figure 1.5f: Markov regime switching model on realized volatility measures (Sector TEL and UTL)



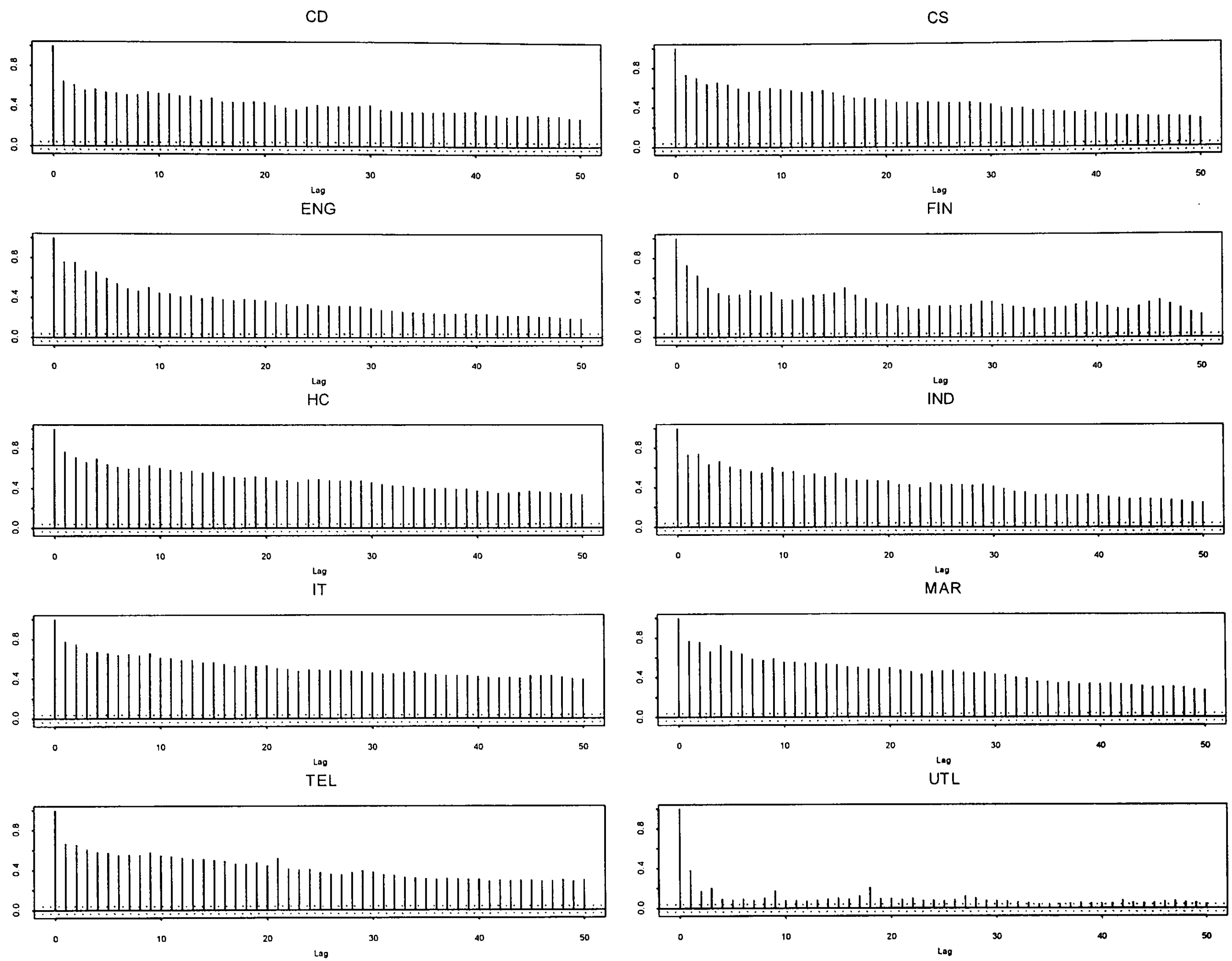
Note: From up to down: realized variance, realized bipower variation, realized power variation and realized range.

Figure 1.6a: Autocorrelation functions (ACFs) of realized volatility measures (overall average)



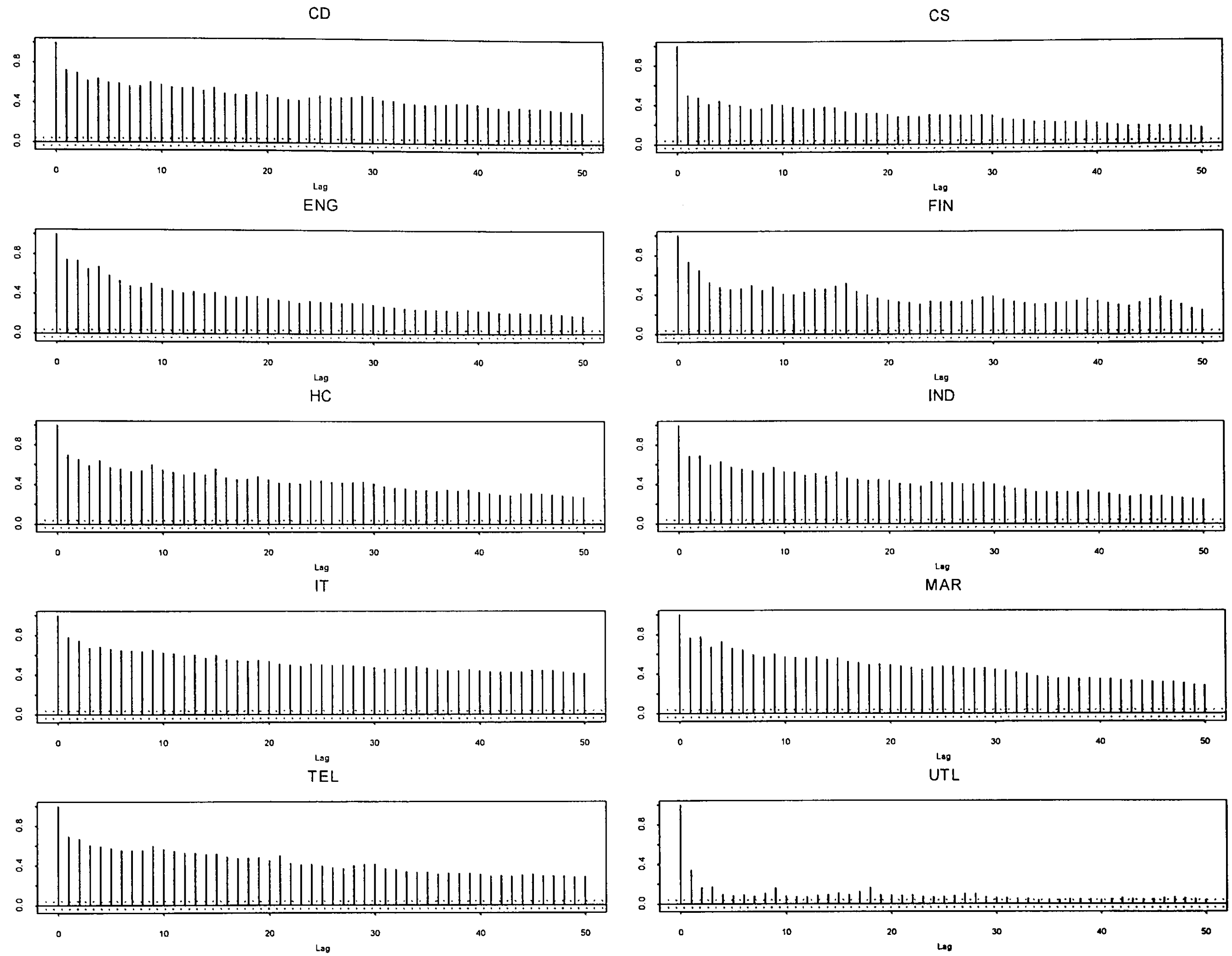
Note: The figure represent the autocorrelation functions of the overall sample average realized volatility measures up to 50 lags.

Figure 1.6b: Autocorrelation functions (ACFs) of realized variance (sector average)



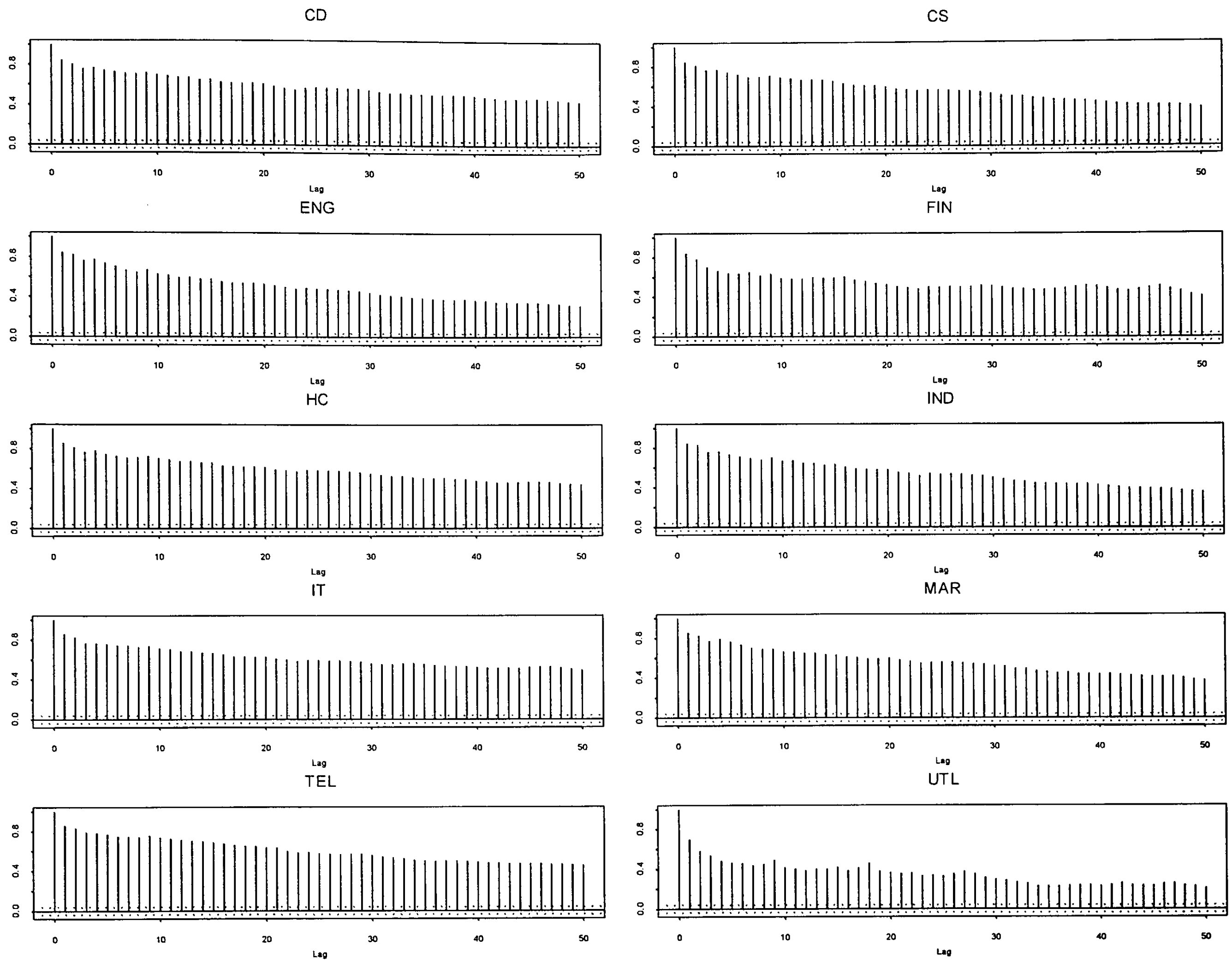
Note: The figure represent the autocorrelation functions of the sector average realized variance up to 50 lags.

Figure 1.6c: Autocorrelation functions (ACFs) of realized range (sector average)



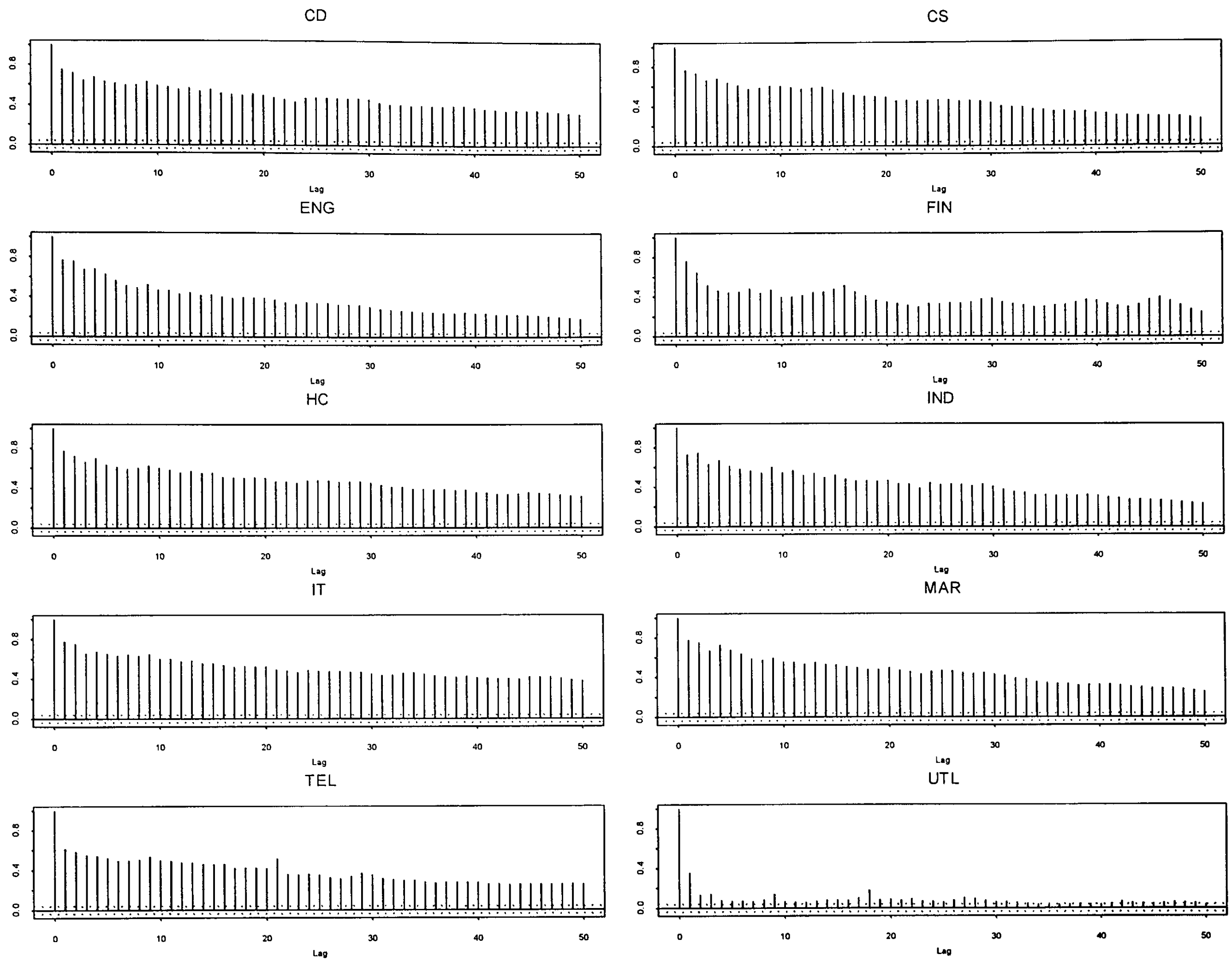
Note: The figure represent the autocorrelation functions of the sector average realized range up to 50 lags.

Figure 1.6d: Autocorrelation functions (ACFs) of realized power variation (sector average)



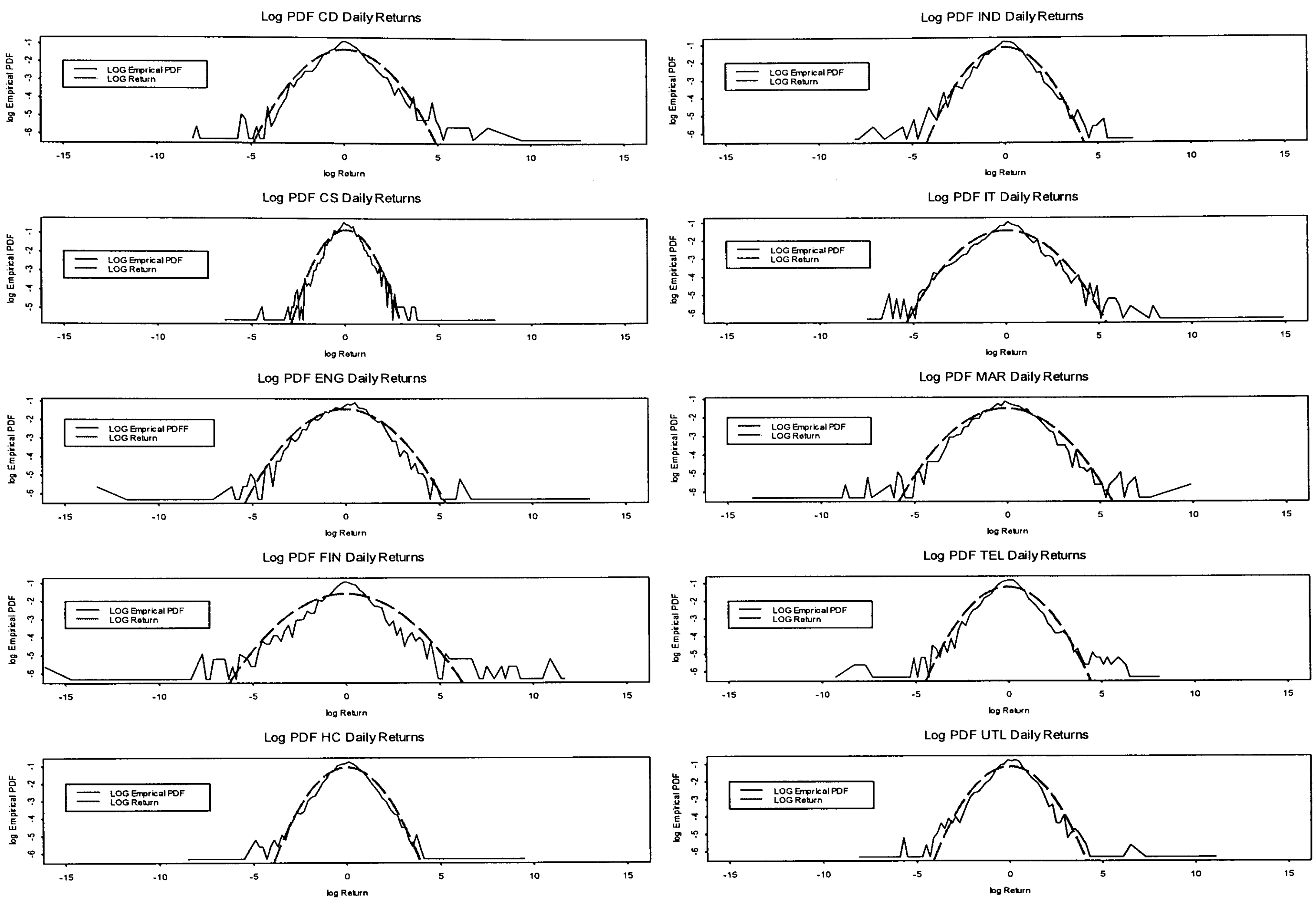
Note: The figure represent the autocorrelation functions of the sector average realized power variation up to 50 lags.

Figure 1.6e: Autocorrelation functions (ACFs) of realized bipower variation (sector average)



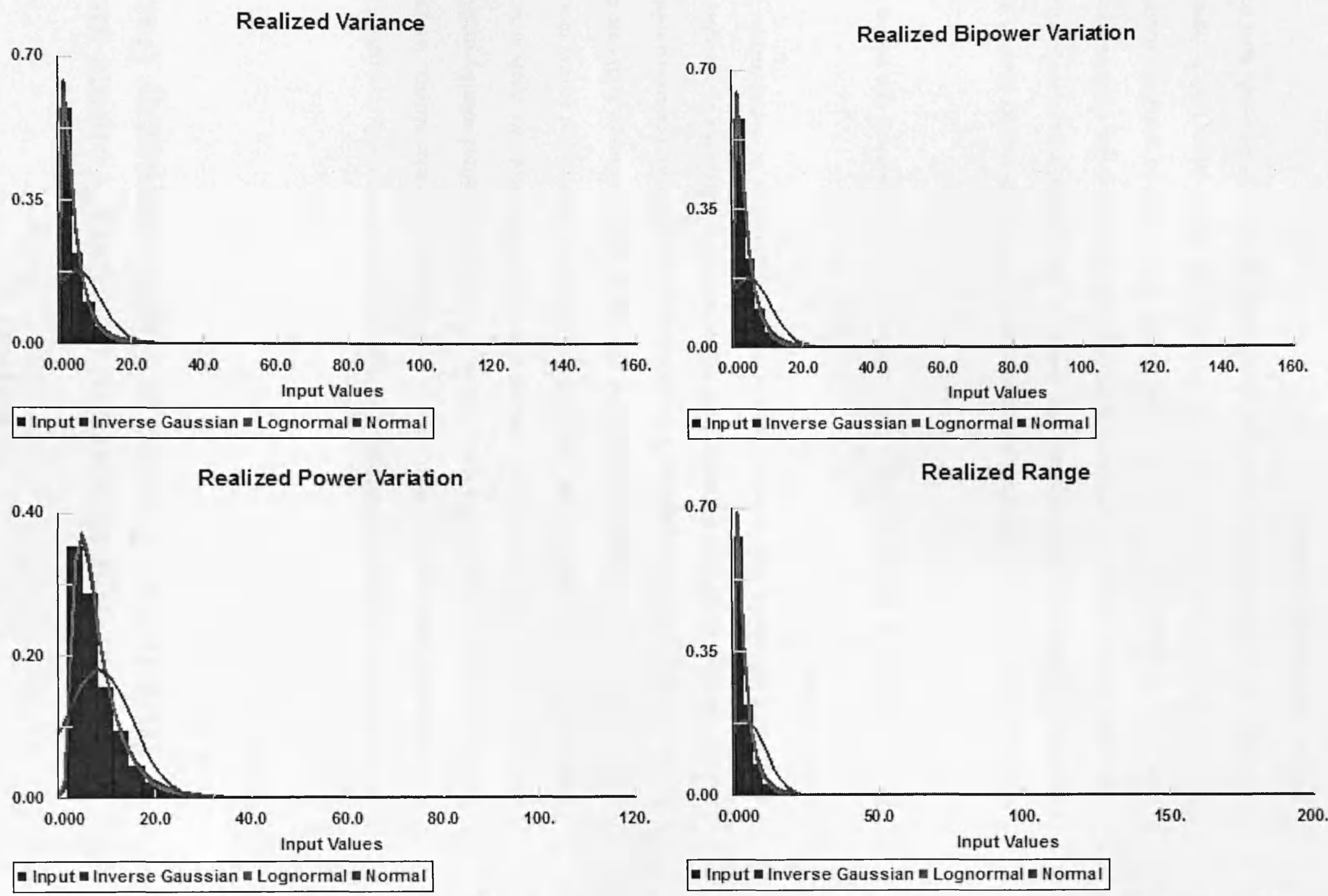
Note: The figure represent the autocorrelation functions of the sector average realized bipower variation up to 50 lags.

Figure 1.7: Log-PDF plots of returns (sector average)



Note: The figures presents Log-PDF plots of sample distribution (black line) for the sector average returns against standard Normal distribution (red line).

Figure 1.8: Distributional plots for realized measures of volatility (overall sample average)



Note: The figures represent the distributional plots for overall sample average realized measures of volatility. Green denotes lognormal distribution, red denotes Inverse Gaussian distribution and purple denotes Normal distribution.

Chapter 2

On the Significance of Trading Volume and Number of Trades in High Frequency Data

Abstract

The value relevance of trading volume and the number of trades has been widely investigated but with no definitive conclusion emerging as to which is a better measure of market activity. We generalize from previous studies in a number of ways. First we consider various realized volatility measures and investigate whether the relation varies with the volatility measure used. We apply a variety of econometric techniques which include looking at: 1) long memory properties of activity measures as compared to measures of volatility; 2) correlation structure; 3) whether structural breaks are common between volatility and activity measures; 4) the explanatory power of activity measures on volatility; 5) Granger causality; 6) GARCH augmented with market activity measures; and 7) estimation of the moments of information flow, matching them with empirical moments of trading volume and the number of trades.

Our findings show the number of trades to be a better measure of market activity than trading volume. The result holds across sectors.

Giving established the superiority of the number of trades, we segment the number of trades into three different categories: trades that takes place at the same price; at a higher price and at lower price; where higher and lower is relative to the price in the previous period. The aim is to find which trades measures carry most significant information, in the sense that it is more capable of explaining volatility. Our results suggest that trades which happen at either lower or higher prices tend to have greater explanatory power than those taking place at an unchanged price. Again our results hold across the different sectors being investigated.

2.1 Introduction

The relation between volatility and variables measuring market activity has been studied extensively. Early treatment of the topic dates from Osborne (1959), who models the stock price change as a diffusion process to the market volatility. Various studies have used different volatility measures and market activity measures, with different types various frequencies of data. All of these together have settled in a well-documented contemporaneous positive relationship between the two measures. Giot *et al* (2010) summarize the volatility-volume relation under three categories. Competitive model and strategic model are both built on adverse selection and asymmetric information. The third model, Mixture Distribution of Hypothesis (MDH) model, assumes that volatility and trading volume are positively correlated, given that both are driven by the same underlying latent market information flow (Clark, 1973). In this chapter, we examine the volatility-volume relation within the framework of MDH.

The pioneering work under MDH by Clark (1973) shows that a mixture of normal distributions can be used to model the distribution of volatility. He finds that squared daily price changes are positively related to volume, and that stock returns generally follow the normal distribution when volume is used as a subordinator. Tauchen and Pitts (1983) adopt the standard MDH model and argue that volatility and volume are both normality distributed and subordinated to the same latent market information arrival process. A later modification by Andersen (1996) allows the model to distinguish uninformed and informed volume. Only the informed volume, which is due to asymmetric information, incorporates the same information arrival flow with market volatility. Liesenfeld (2001) introduces a new latent information flow which primarily affects volatility, rather than trading volume.

In addition to the above five papers which focus more on the adequacy of the model framework according to data fitness, many other studies test the volatility – volume relation empirically. Epps and Epps (1976) find a positive relation between the sample variance of the change in price and volume during a fixed time interval. Westerfield (1977) gives further confirmation to Clark's subordination process and checks the validity of the model by investigating a large number of sample stocks. Harris (1987) extends MDH and finds the same results as Clark (1973), using data of different transaction levels within a day. Garcia *et al* (1986) model volume and volatility under the VAR framework. Gallant, Rossi and Tauchen (1992) suggest positive correlation between conditional volatility and volume using 60 years of S&P 500 composite index at daily frequency. Parametric methods investigating the volatility-volume relation also confirm the relation. For example, Lamoureux and Lastrapes (1990) show that daily trading volume has a good explanatory power on the stock volatility. When the GARCH (1, 1) variance equation is augmented with contemporaneous volume, ARCH effects tend to be negligible.

Many studies focus, not only upon the relation between volume and volatility but also upon the explanatory power of the number of trades, which is another widely used market activity measure. Harris (1987) suggests the number of trades may be a good estimator in the information arriving process, which therefore could be used in the mixture model. Jones, Kaul and Lipson (1994) conclude that stock volume adds no additional information in explaining volatility, beyond the number of trades. Chan and Fong (2000) and Ané and Geman (2000) reach similar conclusions.

Apart from the different market activity measures used in the literature, volatility, a variable that is latent as the market information arrival flow, had also attracted attention. The volatility measures that are used in the papers mentioned above are primarily (absolute) price changes, squared returns, and regression residuals (to overcome the persistence). These measures are calculated from low frequency data (daily, week, or monthly) and often appear to be noisy and inconsistent.

The availability of high frequency data enables the construction of more accurate volatility measures. ABDL (2001) compute realized variance using sums of squared intraday returns. It is now well documented that realized variance, as well as other realized measures of volatility, is less noisy than the traditional low frequency counterparts, thus providing better model estimates both in-sample and out-of- sample than the traditional low frequency volatility measures.

With the use of high frequency data and realized measures of volatility, the volatility-volume relation has been re-examined. The positive relation is further strengthened by the fact that realized measures of volatility become less noisy and more closely convergent upon the true latent volatility when the data is sampled infinitely. Marten and Luu (2003) test the MDH using high frequency data and realized variance. Their findings show that realized measures of volatility challenge almost every conclusion based on daily squared returns under MDH framework. The use of accurate volatility measures gives support to MDH. Similar and improved findings are reported by Chan and Fong (2006) who use high frequency data of 30 US stocks to test the OLS regression first proposed by Jones *et al* (1994). Giot *el al* (2010) further test the volatility-volume relation by decomposing realized variance into a continuous part and a jump part based on Barndorff-Nielsen and Shephard (2004, 2006) bi-power variation and the Huang and Tauchen (2005) method. Their findings suggest that the volatility-volume relation holds only for the continuous part in the realized variance. Both Chan and Fong (2006) and Giot *el al* (2010) find the number of trades rather than volume to be the stronger driving force in the volatility-volume relation.

Even so, notwithstanding all the empirical studies, there remain gaps to be filled. For instance, although the extant literature examines the most widely used high frequency volatility estimator, realized variance, many other high frequency measures have been neglected. In addition to the trading volume, the number of trades and average trades size, this is calculated as volume divided by number of trades, is both extensively analyzed. However, we haven't found any research which further decomposes the number of trades. Trades that are happened when there is a price change are likely to contain different market information to those trades that happened when prices do not change. Since most empirical studies claim that the number of trades is the dominant factor in the volatility-volume relation, it is interesting to ask if there is any dominant factor in the number of trades itself.

In this context, the objective of this chapter is to address the volatility-volume relation using different realized measures of volatility and different market activity measures. Various models are adopted to address this issue, both graphically and econometrically. We aim to investigate if any single market activity measure emerges as the primary driving factor to any realized volatility measures.

In short, this chapter contributes to the literature in a number of distinct ways. First, previously indicated, our sample covers a period of 11 year and consists of 100 stocks from 10 sectors traded in the US equity market since January, 3rd, 2000. To the best of our knowledge, our sample is the most extensive data set used to investigate the volatility-volume relation. Moreover, we obtain our data from the TICK database rather than the commonly used TAQ database. The TICK database adjusts the dividend payments and stock splits and so provides greater accuracy. Second, we use four different realized measures. Previous literature either uses low frequency volatility measures or uses realized variance. Here we use realized variance, realized range, realized power variation and bipower variation. Third, we decompose the extensively studied the number of trades into three components, the up trades, the down trades and the same trades, and attempt to determine whether the well-documented stronger relation between number of trades and volatility (as against between volume and volatility) still holds when the number of trades is further decomposed. Fourth, we not only investigate the volatility-volume relation on a 100-stock collective aspect, but also look at it from a sector specific aspect. Our database includes both very actively traded stocks from the IT and Financial Services sectors and less actively traded stocks from the Utility sector. With stocks from varying degrees of trading activity, we achieve a better insight into the relationship. In particular, we observe that some sectors with lower trading volumes have a relatively higher number of trades.

The rest of chapter unfolds as follows. Section 2 is provides literature review. Section 3 presents the theoretical framework. Section 4 sets out the. Section 5 presents a data description. Section 5 gives the empirical results. Section 6, the final section, summarizes our conclusions.

2.2 Literature Review

The volatility-volume relationship has been investigated extensively over decades. Karpoff (1987) summarizes the importance of volume and volatility relation in four parts. First, the relation affords insight into the financial market structure. Second, the relation of volume and volatility may help researchers to draw inferences in regard to their joint determination. Third, the volatility-volume relation could be applied to the empirical distribution analysis. For instance, the mixture of distribution hypothesis (MDH) is generally supported by tests upon volume and price changes. Finally, the volatility-volume relation may permit inference to be made about further implications for derivatives market analysis.

Unlike volume, or other observable market activity measures, volatility is usually treated as latent. There are various measurements for volatility. To cite a few: price changes or absolute price changes (Ying (1966) Clark (1973); squared returns (Andersen, 1996); absolute returns (Andersen, 1996); GARCH (Engle, 1982); and realized volatility (Andersen *et al*, 2001). All of these volatility measures have all been used to test the volatility-volume relation empirically. Even with various definitions of volatility, the conclusions reached under the MDH are generally compatible, although some tests do present contradictory findings. In the following section, we review the most frequently cited empirical studies which discuss the relation of volatility and volume (or number of trades) within the MDH framework. We start from early studies that primarily use low frequency data and then discuss papers that consider realized measures of volatility. A detailed literature review relating different realized volatility measures is provided in Chapter 1.

The Ying (1966) test on the absolute price change and volume is one of the earliest studies of the volatility-volume relation. Standard & Poor's 500 composite stocks closing price index and volume of stock sales on the NYSE are used from January 1957 to December 1962, at daily frequency. The sample is divided into 5 classes according to the mean and standard deviation. Ying's findings all suggest a positive correlation between volume and volatility.

Clark (1973) also finds a positive relation between stock volatility and volume for daily data from the US cotton futures market, from 1945 to 1958. He shows that price change (as a measure of volatility) is not normally distributed. However, it could exhibit a generally normal distribution through the subordination process. This paper introduces the MDH, which would be viewed as a core model to explain the volume and volatility relationship. According to MDH, stock return is directed by a stochastic information arrival process which affects stock prices and returns in turn. Clark assumes volume is exogenous. Under this assumption, he concludes that volume has a good explanatory power on the stock volatility.

Epps and Epps (1976) also test the distribution of price variability. They build up a model of the price-formation process in which the sample stock volatility, defined as conditional variance of the price changes in the paper, is a function of volume. There exists a positive relation between them. Transaction data of 20 stocks from the NYSE, during January 1971, are used. The OLS results for single stocks suffer varying degrees of biases. Moreover, not all of the coefficients on volume are significant. To address these problems, Epps and Epps compute the maximum likelihood of the coefficients on volume under the assumption that the error term is normal. After this adjustment, the overall significance level of the estimation is considerably improved.

Tauchen and Pitts (1983) conducts another important study on the relation of volume and volatility. It extends earlier work in two ways. First, their study is derived from the joint probability of volume and volatility over any interval within the trading. Second, the authors determine how the joint distribution changes as the number of trades increases. Using 90-day US T-bills from January 1976 to June 1979, the authors build up a more general model of the volatility and volume, known as the bivariate MDH later, which could explain both the previous studies and the data in their paper. The Tauchen and Pitts model is mainly relevant to intra-day trading. In their preliminary test, the authors find when they add a trend about the number of trades in the market (as the number of trades cannot be provided directly by then), the volatility and volume regression will raise the coefficients on volume. This finding is consistent with their prediction that the mean volume should increase and the volatility should decrease as the size of the market increases. The paper reaches three conclusions. First, the mixture distribution of volatility and volume incorporates all the relevant information flow on volume and volatility. The correlation of trading volume and volatility increases with the variance of the rate of information flow. Like previous studies, their data shows a positive relationship between volume and volatility. Furthermore, their paper argues that the test results of volatility-volume relation could be misleading where volume is not fixed during the sample period.

Harris (1987) further extends the findings of Clark (1973) and Epps and Epps (1976) about MDH in two aspects: the number of trades should be proportional to the number of information processes, and the number of trades should be correlated to the volatility and volume. If there is autocorrelation in the number of trades, it should be stronger than this of volume. Other predictions are concerned with the transaction intervals volatility and volume. Under the assumption that transactions occur at a uniform rate in the event time, Harris predicts that 1) as transaction intervals become longer, volatility and volume should become more closely normally distributed; 2) transactions interval volatility should not be correlated with transaction interval volume; 3) Transaction interval volatility and volume should have no autocorrelation problem. 50 stocks from NYSE are selected to test the predictions. For each stock, Harris computes the volatility and volume over daily intervals and fixed intervals of 1, 50 and 100 transactions. His test results confirm most of the predictions. In particular, Harris concludes that the squared daily price changes and volume are positively correlated and that actual daily price changes and volume are weakly positively correlated with a bivariate normal mixture of distributions model. Another is that the test results suggest that the number of trades can produce a good estimate of market information flow

Gallant, Rossi and Tauchen (1992) undertake a comprehensive test using data of daily closing prices of S&P composite stock index and the daily volume of stocks from the NYSE from 1928 to 1985. In consideration of the long sample period and possible bias arising from calendar effects and long-term trends, the authors make various adjustments on the raw data. The adjustments include using day-of-week dummies, dummy variables for non-trading days, for special months (e.g., January effects) and for special years (war time), and also time trend variables. Following adjustment, both volume and volatility show more stable properties. The strategy on conditional density estimation is primarily nonparametric where the semi-nonparametric (SNP) estimation is the main technique. More specifically, the contemporaneous conditional price-volume relationship, the conditional moment structure of stock price change, dynamic price-volume relationship, and risk premium and conditional price volatility are examined sequentially. There are four findings: first, there is a positive relation between volume and conditional volatility; second, large price movements are usually associated with large trading volume; third, the leverage effect is alleviated when volatility is conditioned on lagged volume; and finally, the positive relation of risk and return exists after conditioning on lagged volume.

A similar piece of research concerning volatility-volume relation but using parametric method is that of Lamoureux and Lastrapes (1990). They use daily data of 20 actively traded stocks from 1981 to 1985 to test GARCH model and investigate ARCH effects. When GARCH (1, 1) is tested, all 20 stocks show that ARCH effects exist, supporting the hypothesis that ARCH reflects an uneven but persistent information flow to the market. When GARCH (1, 1) variance equation augmented with volume is tested, the coefficient on volume is significantly positive for each single stock. This suggests a positive correlation between volume and volatility. Alternatively, the lagged squared residuals contribute very little explanatory power when volume is added to the model. The ARCH effects tend to disappear. From their test results, the authors conclude that daily trading volume has significant explanatory power on stock volatility, and the positive coefficient further confirms a positive correlation.

Other empirical works which build on Lamoureux and Lastrapes (1990) include Najand and Yung (1991) Chen *et al* (2001), Hussain (2011) and Louhichi (2011). All these works use GARCH family models. Different market activity measures (contemporaneous and lagged) from different markets and different countries have been examined, including trading volume, number of trades, size of trades, and bid-ask spread. Of course, there are variations between the different studies. For instance, one of the most important findings of Lamoureux and Lastrapes (1990) paper is that, when contemporaneous volume is augmented in the GARCH model, the augmented variable becomes positively significant and the GARCH effect is greatly reduced. Only Louhichi (2011) reaches the same conclusion using stocks from Euronext Paris and CAC40 Index. Hussain (2011) augments both contemporaneous and lagged market activity measures with EGARCH model using high frequency DAX 30 data. He finds that statistically significant augmented variables will not reduce the persistency of GARCH models. The paper further finds that the relationship between volatility and unexpected volume is positive and that the relationship between volatility and expected volume is negative. With and research based on low frequency data, Chen *et al* (2001) reach the same conclusion of Hussein (2011).

Studies that investigate the role of the number of trades in the volatility-volume relationship have received increasing attentions. Jones, Kaul and Lipson (1994) use daily data of 853 stocks to investigate the explanatory power of the number of trades. They find that it is the number of trades, not volume, that determine volatility. The data are calculated from the average of closing bid and ask quotes and are collected from NASDAQ-NMS between 1986 -1991. Jones, Kaul and Lipson divide stocks into 5 portfolios based on market capitalization. At first, the paper reports the correlation between volume, number of trades and average trading size, which is defined as "the total number of share divided by number of trades". The correlation between number of trades and average trading size is much lower than both correlation between volume and number of trades, and between volume and the average trading size. This implies that volume and number of trades contain different information. The volatility is calculated according to Schwert (1990), but more dummies are used. On the next step, The OLS, which regresses volatility on average trading size and the number of trades, both separately and together, is applied to each portfolio. There are two main findings: 1) volume has very little explanatory power when volatility is both conditioned on the number of trades and trading size. It contains no information beyond the number of trades. 2) both the number of trades and the size of transactions are endogenously determined. Each of these findings remains the same when the sample is divided into two sub-periods, or use different measure of volatility and volume, or alternative regression specifications.

Another frequently cited work giving support to the number of trades is that of Ané and Geman (2000). Their focus is upon the normality of the stock returns and they extend the work of Clark (1973. The paper shows that the normal distribution of returns can be obtained through a stochastic time change. Traditionally, empirical studies have used trading volume, but Ané, and Geman (2000) use both volume and number of trades for the stochastic time changes. The paper uses high frequency data of two stocks, at frequencies of one, five, ten and fifteen minutes respectively. The density estimation mainly uses Kernel methods, which are nonparametric. The distributions of stock returns, volume and number of trades are constructed. When returns are conditional on the number of trades, they tend to be normal. Therefore, it is concluded that the number of trades is a better mixing variable to recover the normality of the stock returns than traditionally used volumes.

The models used in the paper are given as

$$\hat{\sigma}_t = a + \beta \Delta vol_t + \sum_{J=1}^{12} \rho_j \left|\hat{\sigma}_{t-j}\right| + \eta_t^1 \qquad [2.1.1]$$

$$\hat{\sigma}_t = a + \gamma \Delta nt_t + \sum_{J=1}^{12} \rho_j \left|\hat{\sigma}_{t-j}\right| + \eta_t^2 \qquad [2.1.2]$$

$$\hat{\sigma}_t = a + \beta \Delta vol_t + \gamma \Delta nt_t + \sum_{J=1}^{12} \rho_j \left|\hat{\sigma}_{t-j}\right| + \eta_t^3 \qquad [2.1.3]$$

where Δvol_t and Δnt_t are the first difference of trading volume and number of trades.

Luu and Martens (2003) tests the MDH using both realized volatility and squared returns. The data used in this paper are constructed from 10-year S&P 500 index-futures transaction prices. After conducting a series of tests (for instance the GARCH extended with volume, bivariate VAR model for volume and volatility and long memory test in volume and volatility), they find their results vary significantly when different volatility measures are used. When lagged realized volatility is added to the variance equation of GARCH, the lagged volume over open interest is no longer significant. This is because all useful information is already captured when realized volatility is included in the equation and therefore the lagged volume becomes unnecessary. The authors further conclude that the insignificance of lagged volume does not necessarily reject the MDH. Although lagged volume is not significant in the variance equation, the lagged volume over open interest is. This is consistent with the Andersen (1996) which suggests that volume should be split into two parts: namely liquidity trading and speculative trading. Another finding is that, when VAR of daily squared returns and volatility are performed, the uni-directional causality is only found from volatility to volume, but not in reverse. This result fails to support the MDH. When daily squared return is replaced by realized volatility, however, the bi-directional causality is observed, giving the support to MDH. The paper also tests the long memory of squared returns, realized volatility as well as volume and in all cases the statistics confirm that the null of long memory cannot be rejected. The authors conclude that realized volatility is a more accurate volatility measure to support the MDH comparing to the daily squared returns, which is believed to be noisier and less efficient.

Chang and Fong (2006) study the volatility-volume relation using high frequency data. This work may be viewed as an extension of the papers by Jones, Kaul and Lipson (1994) and Ané and Geman (2000), but it also investigates the number of trades-volatility relation and order imbalance (absolute and with signs)-volatility relation. The authors compare the results of regressions of absolute residuals on trade frequency measures with the results of regressions of realized volatility on trade frequency measures. The latter performs much better than the daily absolute residual, as it is much less noisy. The regression results also suggest number of trades is the most capable measure in explaining realized volatility. When realized volatility is regressed on the number of trades, the adjusted R-squared is 42%. The finding is consistent across time period, firm size (defined by market capitalization) and aggregation frequency. Although they are all statistically significant, the trade size and order imbalance add little explanatory power.

Izzeldin (2007) performs a comparative study between trading volume and number of trades to see either is a good proxy for market activity. Like Ané and Geman (2000), this paper also uses the high frequency data of Inter and Cisco in 1997, but at different time frequency of 10, 30, and 60 minutes and daily. The Generalized Method of Moments (GMM) J test is first applied to test the moment restrictions implied by the MDH. The test results show that both volume and number of trades could be used as the mixing variable in the MDH, but number of trades provides stronger support. The other tests, including augmented GARCH and subordination process to achieve returns normality. The results are all in favour of the number of trades.

Giot, Laurent and Petitjean (2010) look at the relation among trading activity, realized volatility and jumps, inspired by the work of Chang and Fong (2006). In order to better study the positive relation between volume and volatility, the authors not only decompose the realized volatility into continuous part and jump part, but also compare the explanatory power of different trade frequency measures, namely trading volume, number of trades and absolute order imbalance. They use high frequency equity data from 100 largest stocks traded in NYSE from 1995 to 1999, which includes 1199 trading days. After conducting a series of regressions such as OLS, GMM, Robust

Regression, Median Regression and TOBIT with and without GARCH effect, the authors find that the positive volatility-volume relation holds for the continuous component of volatility only. Trade variables are positively related to the continuous part, while the relation between jump components and volatility appears to be negative. They also find that number of trades is the dominant factor in the relation of volume and volatility, while trading volume and order imbalance fail to increase explanatory powers in the relation. The conclusion holds for both continuous volatility and discontinuous jumps.

2.3 MDH Framework

The Mixture of Distribution Hypothesis states that both daily price changes and trading volume are governed by the same latent information process, therefore, prices changes and trading volume are positively correlated. Both price volatility and volume should be a mixture of the conditional distribution, given the condition is the market information flow. Returns conditional on the mixing variables are normally distributed.

MDH is firstly proposed by Clark (1973) who considers a univariate framework.

$$returns = r_t \mid i_t \sim N(\mu_{r_t}, \sigma_{r_t}^2) \qquad [2.2.1]$$

$$volume = \log a_t \sim N(\mu_{a_t}, \sigma_{a_t}^2) \qquad [2.2.2]$$

$$\mathrm{cov}(r_t, a_t \mid i_t) = 0 \qquad [2.2.3]$$

where r_t is returns over a specific time interval t, which is set empirically at the daily level. a_t is trading volume and i_t is the latent market information flow. Finally, σ_{r_t} σ_{a_t} μ_{v_t} are all positive. Clark models volume as lognormal which allows the presence of excess kurtosis in the unconditional distribution of return series.

The bivariate MDH model is first introduced by Tauchen and Pitts (1983). Tauchen and Pitts's model assumes volume is endogenous. Hence, returns and volume follow a bivariate normal distribution conditional on the daily information flow,

$$r_t \mid i_t \sim N(0, \sigma_{r_t}^2 i_t) \qquad [2.3.1]$$

$$a_t \mid i_t \sim N(\mu_{a_t} i_t, \sigma_{a_t}^2 i_t) \qquad [2.3.2]$$

$$\text{cov}(r_t, a_t \mid i_t) = 0, \text{cov}(r_t^2, a_t) \neq 0 \qquad [2.3.3]$$

The bivariate MDH model allows volatility to be dependent on the time series behaviour of the mixing variable i_t which also drives the trading volume.

The models by Clark (1973) and Tauchen and Pitts (1983) both assume that the latent market information flow is serially independent. An assumption that is sometimes unrealistic, given the observed serial dependence of both volume and returns is assumed to be driven by the latent information. Andersen (1996) further modifies the MDH model by allowing the split of volume into liquid trading and informed trading and further allowing the liquid trading unrelated to the information flow. It also assumes that i_t is serially dependent. In the presence of serial correlation of information flow, both volume and volatility should be positively related to the previous volume and volatility, respectively. The Andersen (1996) model is given by

$$r_t \mid i_t \sim N(\mu_{r_t}, \sigma_{r_t}^2 i_t) \qquad [2.4.1]$$

$$a_t / c \mid i_t \sim P(m_0, m_1 i_t) \qquad [2.4.2]$$

$$\operatorname{cov}(r_t, v_t \mid i_t) = 0 \qquad [2.4.3]$$

m_0 reflects the liquidity trading part while the informed trading is proportional to the information flow. m_0 itself is independent of the information flow. The remaining part of trading volume that is due to new information is represented by $m_1 i_t$. Andersen argues that, if a large number of information arrivals leads to $(m_0, m_1 i_t)$ becoming sufficiently large, the Poisson process in [2.4.2] might be approximated by the normal distribution as

$$a_t / c \mid i_t \sim N([\mu_{a_t} i_t, \sigma_{a_t}^2 i_t) \qquad [2.4.4]$$

where c is an unknown scaling factor when detrended volume is used. Another difference between Andersen and Tauchen and Pitts's bivariate models is that Andersen (1996) allows for a non-zero mean of the returns series.

Liesenfeld (2001) extends the modified MDH model of Andersen (1996) by assuming information flow could be written as a lognormal stochastic volatility model,

$$\ln(i_t) = \lambda_t = \gamma + \delta\lambda_{t-1} + v\varepsilon_t, \varepsilon_t \sim i.i.d.N(0,1) \qquad [2.5]$$

This normal AR (1) process of i_t looses the assumption made by Clark (1973) and Tauchen (1983) which the mixing variable is not serially correlated and allows lognormal distribution at the same time.

The Liesenfeld bivariate MDH model is given as

$$r_t \mid \lambda_t \sim N(\mu_{r_t}, e^{\lambda_t}) \qquad [2.6.1]$$

$$a_t \mid \lambda_t \sim N(\mu_{a_t} e^{\lambda_t}, \sigma_{a_t}^2 e^{\lambda_t}) \qquad [2.6.2]$$

$$\lambda_t \mid \lambda_{t-1} \sim N(\gamma + \delta\lambda_{t-1}, \nu^2) \qquad [2.6.3]$$

The stochastic volatility specification allows the latent information flow affect volatility primarily, in turn puts more weight on the price volatility. However, the stochastic volatility specification also makes the model estimation more complicated as stochastical volatility is not known in closed form as previous univariate and bivariate MDH models are.

2.4 Data and Variable Description

We construct our dataset using 100 stocks originally traded in the US equity market from January, 2000 to December, 2010: 2767 trading days in total. All stocks are obtained from the high frequency TICK database which adjusts stock splits and dividend payments. As discussed in Chapter 1, our data are aggregated at the 5-minute level and all realized measures of volatility are calculated from the 5-minute aggregated returns. We further segment into 10 sectors according to the sector segmentation criterion of the S&P 500. The method of construction of realized measures of volatility was discussed in detail in a previous chapter.

Since TICK allows stock splits, trading volumes (vol_t) are also adjusted indirectly. Furthermore, in addition to the most frequently discussed trading volume and number of trades, we decompose the number of trades into 'up trades' (ut_t), 'down trades' (dt_t) and 'same trades' (st_t). ut_t is recorded when a transaction is made at a higher price than its previous trade. dt_t is recorded when a transaction occurs at a price below the previous trade. st_t is a transaction in which a security is traded at exactly the same price as its previous one. The sum of ut_t, dt_t and st_t is nt_t. To our knowledge, this is the first empirical study which uses up/down/same trades to investigate the volatility-volume (trades) relation.

The descriptive statistics of volume and trades measures is reported in **Table 2.1**. As trading volume is, *per se*, a bigger measure than trades, we rescale the volume by 1/1,000,000 and trades measures by 1/10,000 for easier comparison.

All market activity measures are all rightly skewed, highly persistent and have fat tails. The trade measures are more serially correlated than vol_t, according to the Q statistics. st_t is the most persistent measure of all. ut_t and dt_t exhibit similar properties to each other. We find that nt_t is mainly composed of st_t (over 70% across sectors). This implies that most trades happen at level prices. The trade measures are not stationary in many stocks. This could be explained by the relatively long sample period time and the rapid development of the US equity market. vol_t exhibits higher stationarity than trade measures. There is a general upward trend of all the market activity measures (**Figure 2.1**). The IT sector is the most actively traded sector with average trading volume of 30.6 million and average number of trades of 5.7 million. The ULT sector is the least actively traded sector with 2.1 million average trading volume and 0.7 million number of trades. Ranking patterns are different for the number of trades: ENG sector ranks 7th according to trading volume but is the 4th most active sector if ranked by number of trades. For consistency with Chapter 1, we will still use trading volume as the benchmark for sector activity in the chapter.

[Table 2.1 here]

[Figure 2.1 here]

Robinson's d long memory test results are provided in **Table 2.2**, together with the overall average autocorrelation functions in **Figure 2.2**. The statistics confirm that all the market volatility measures exhibit long memory. st_t is the most persistent and vol_t the least. The d statistics of ut_t and dt_t are similar across sectors and generally less persistent than NT and st_t. MAR has the most persistent vol_t, nt_t and st_t while ENG reports the most persistent ut_t and dt_t. IT has the lowest d statistics of vol_t and it is also the most activity traded sector.

[Table 2.2 here]

[Figure 2.2 here]

Table 2.3 reports the cross-sectional mean correlations between market activity measures and realized measures of volatility. Our results are very similar to those reported by Chan and Fong (2006) and Giot *et al* (2010). Overall, we find that: trading volume is sigfinicantly and positively correlated with trade measures (over 70%); trade measures are highly correlated with each other (over 90%); all market activity measures are highly correlated with realized measures of volatility; realized measures of volatility are more highly correlated with vol_t than with nt_t ($corr(rv_t,vol_t)$ =

39.4%, $corr(rv_t, nt_t)$ = 31.2%); realized measures of volatility are more highly correlated with ut_t and dt_t then with vol_t ($corr(rv_t, ut_t)$ = 41.4%, $corr(rv_t, dt_t)$ = 41.3%); the correlation between realized measures of volatility and st_t is the weakest ($corr(rv_t, st_t)$=26.3%). This finding contradicts with those of Chan and Fong (2006) and Giot *et al* (2000) which both find higher correlations between volatility and number of trades than between volatility and volume.

We then look at the sector average correlations, where we find that: realized measures of volatility correlate more highly with vol_t than with nt_t in 8 out 10 sectors. Realized measures of volatility are more highly correlated with ut_t and dt_t than with vol_t in 7 out of 10 sectors. The exceptions are CD, HC and TEL. The remaining of conclusions are the same as for the overall averages.

To conclude, the positive correlations between realized measures of volatility and market activity measures suggest that volatility and market activity measures are driven by common latent factors. It may the number of traded that is happen when prices change be the main driving force behind the volatility-volume relation.

[Table 2.3 here]

2.5 Methodology

In the previous section, we summarize the main theoretical models which focus on the estimation of model parameters and latent variable(s). However, our main interest is to compare the theoretical implications of the MDH model with the empirical dynamics.

In this section, we outline a series of econometric models to test the validity of the volatility-volume relation under MDH. We first identify whether common structural breaks exists between realized volatility measures and market activity measures. Then, we regress different realized measures of volatility on market activity measures. The model was first proposed by Jones *et al* (1994) and later modified by Chan and Fong (2006) who confirm the validity of the model using high frequency data. GARCH augmented with lagged market activity measures is discussed next. This model was first discussed by Lamoureux and Lastrapes (1990) who augmented contemporaneous volume with GARCH. In our model, we follow Najand *et al* (1991) and treat market activity measures at a one period lag. Bivariate VAR is then reported as different researches reach the different conclusions so that the bi-directional Granger causality (which supports MDH) is not universally found. The fifth model is the estimation of information flow using Generalized Method of Moment (GMM). We adopt and develop the model from such previous works as Richard and Smith (1994), Andersen (1996) and Izzeldin and Murphy (2010). We estimate the theoretical information moments and compare them with the observed empirical moments. The last model is the standardization procedure to recover returns normality by using market activity measures as the standardization factors.

2.5.1 Single Structural Break

Although the Augmented Dickey-Fuller (1979) test is the most commonly used to detect the unit root, one of its disadvantages is that it assumes no structural break. This assumption is unlikely to hold when the sample time period is long. Hence, Zivot and Andrews (1992) propose a single structural break test which assumes the presence of the break is as an exogenous phenomenon.

The original Zivot and Andrews (ZA afterwards) test has three types of models. In this chapter, we consider the first model, which permits a one-time change in the level (mean) of the series.

$$\Delta y_t = c + a y_{t-1} + \gamma DU_t + \sum_{j=1}^{k} d_j \Delta y_{t-j} + \varepsilon_t \qquad [2.7]$$

Where DU_t is an indicator dummy variable for a mean shift happening at each possible break-date (TB).

The ZA test has been applied widely in macroeconomics data sets to identify the break date. Here we apply the test to rv_t, vol_t and nt_t to investigate the following issues: 1) whether the volatility measure and trade activity measures have the same break date for the same stock; 2) whether there is any particular date emerging across sectors and across volatility and market measures.

2.5.2 OLS Regression

The second method uses Ordinary Least Squares (OLS) regressions to investigate the explanatory powers of market volatility measures on realized measures of volatility. This method can be viewed as a direct test of the MDH as volume (number of trades) is proportional to the information flow, as dealers conduct trading activities in the given trading time and in response to new market information. The model was first introduced by Jones *et al* (1994) in regressing daily number of trades and trading volumes on absolute return residuals, a proxy of volatility first proposed by Schwert (1990). Similar works followed: Ané and German (2000), Chan and Fong (2000, 2006) and most recently, by Giot *et al* (2010). Applying different data sources and different measures of volatility and market activity, the papers all reach a similar conclusion: there is a positive relation between volatility measures and market activity measures. Market activity measures explain proportions of volatility, ranging from 5% to 45%, depending on the measures used. In Chan and Fong (2006), the authors compare absolute return residuals with realized variance and find that the volatility-volume relation is much stronger when realized variance is used. It is generally accepted that the number of trades explains more volatility than trading volume. Here we follow closely the approach by Chan and Fong (2006) by regressing realized variance, realized range, realized power variation and realized bipower variation on trading volume, number of trades and trade decompositions. Across all the literature, the number of trades is shown to reflect more of the daily information arrivals as it explains the greatest amount of the volatility.

The models are specified as

$$rv_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} rv_{it-j} + \begin{Bmatrix} \phi_i vol_{it} \\ \beta_i nt_{it} \\ \gamma_i ut_{it} \\ \theta_i dt_{it} \\ \tau_i st_{it} \end{Bmatrix} + \varepsilon_{it} \qquad [2.8.1]$$

$$rr_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} rr_{it-j} + \begin{Bmatrix} \phi_i vol_{it} \\ \beta_i nt_{it} \\ \gamma_i ut_{it} \\ \theta_i dt_{it} \\ \tau_i st_{it} \end{Bmatrix} + \varepsilon_{it} \qquad [2.8.2]$$

$$pv_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} pv_{it-j} + \begin{Bmatrix} \phi_i vol_{it} \\ \beta_i nt_{it} \\ \gamma_i ut_{it} \\ \theta_i dt_{it} \\ \tau_i st_{it} \end{Bmatrix} + \varepsilon_{it} \qquad [2.8.3]$$

$$bv_{it} = \alpha_i + \alpha_{im} M_t + \sum_{j=1}^{12} \rho_{ij} bv_{it-j} + \begin{Bmatrix} \phi_i vol_{it} \\ \beta_i nt_{it} \\ \gamma_i ut_{it} \\ \theta_i dt_{it} \\ \tau_i st_{it} \end{Bmatrix} + \varepsilon_{it} \qquad [2.8.4]$$

rv_{it} , rr_{it} , pv_{it} and bv_{it} are realized variance, realized range, realized power variation and realized bi-power variation of stock *i* on day *t*. M_t is a Monday dummy, vol_{it}, nt_{it}, ut_{it} , dt_{it} , st_{it} are respectively trading volume, number of trades, number of up trades, number of down trades, and number of same trades of stock *i* on day *t*; and ρ_{ij} is a measure of volatility shock persistence at lag *j*. The Monday dummy is set to account for the well-documented Monday effects in equity markets and the 12 lags is to account for the persistence in the volatility measures.

As some of the market activity measures are trend stationary as indicated by the Augmented Dickey-Fuller test statistics, we form a stationary series of those market activity measures following a nonlinear trend model as used by Chen *et al* (2001) and Chiang (2010):

$$A_t = \alpha + \beta_1 t + \beta_2 t^2 + \varepsilon_t \qquad [2.9]$$

where A_t denotes the raw market activity measures and the residual ε_t is the de-trended market activity measures. A paper by Chan and Fong (2006) considers the situation of both unfiltered and detrended series without reporting the stationarity of the market activity measures. The results from both series reach the same conclusion.

2.5.3 GARCH and Augmented GARCH

Lamoureux and Lastrapes (1990) test the GARCH and augmented GARCH models to investigate the explanatory power of trading volume. They find that when volume, which is used as a measure of the daily information flow to the market, is used at the weakly exogenous mixing variable and augmented in the GARCH model, the ARCH effects tend to disappear and the persistence in the volatility will alleviate. In turn, these findings suggest trading volume is a good proxy for rate of information arrivals in the market. Similar models are adopted by, for instance, Bessembinder and Seguin (1993) which discuss both expected and unexpected trading. Marten and Luu (2003) extend trading volume and realized variance with the GARCH. There are also studies using the number of trades, rather than volume, which is shown to be a better measure to explain MDH under the Augmented GARCH models.

Here we will follow Lamoureux and Lastrapes (1990) but extend the augmented GARCH models, using various market activity measures to see which explains the volatility best.

$$
\begin{aligned}
& r_t = \mu + \varepsilon_t \\
& \varepsilon_t \mid \Phi_{t-1} \sim N(0, h_t) \\
& h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma \begin{Bmatrix} vol_{t-1} \\ nt_{t-1} \\ ut_{t-1} \\ dt_{t-1} \\ st_{t-1} \end{Bmatrix}
\end{aligned}
\qquad [2.10]
$$

where r_t is daily returns calculated from intra-day returns, ε_t is the residual which follows a conditional Gaussian distribution with zero mean and variance h_t . Φ_{t-1} stands for a one day lag information set. All the market activity measures are lagged one term, following the conclusion of Najand *et al* (1991) that contemporaneous market activity measures may cause a simultaneity problem leading to inconsistent estimation of the coefficients. Therefore, we assume one-period lagged volume to be exogenous. The one-period lagged volume is also used in Chan et al (2001) and Luu *et al* (2003).

2.5.4 Bivariate VAR models

It is well documented in the literature that a dynamic structure exists between market activity measures and realized measures of volatility. Lagged volatility measures may affect current market activity measures and lagged market activity measures may affect current volatility. This dynamic could be investigated by implementing a VAR *(p)* for realized measures of volatility V_t and market activity measures, A_t . A commonly used model to investigate the causal relationship between the variables is the Granger Causality test. Previous studies find either or both directions of the causality for different data. For example, Fung and Patterson (1999) and Chiang *et al* (2010) find volatility Granger causes volume; Darrat *et al* (2003) find significant Granger causality from volume to volatility; and bi-directional causality is reported by Brooks (1998) and Marten and Luu (2003).

To date, no conclusive results have been reached. Hence it is interesting to test the dynamic structure between different realized measures of volatility and market activity measures, to see whether bidirectional Granger causality, which supports the MDH, can be found. The bivariate VAR is given as:

$$V_t = \sum_{j=1}^{p} \alpha_{1j} V_{t-j} + \sum_{j=1}^{p} \beta_{1j} A_{t-j} + \varepsilon_{1t} \qquad [2.11.1]$$

$$A_t = \sum_{j=1}^{p} \alpha_{2j} V_{t-j} + \sum_{j=1}^{p} \beta_{2j} A_{t-j} + \varepsilon_{2t} \qquad [2.11.2]$$

We test either $H_0 : \alpha_{11} = \alpha_{12} ... = \alpha_{1p} = 0$ against the alternative hypothesis that market activity measures Granger cause realized measures of volatility; or $H_0 : \beta_{21} = \beta_{22} = ... = \beta_{2p} = 0$ against the alternative that realized measures of volatility Granger cause market activity measures. The optimal lag length p is determined according to the Schwarz Information Criterion (SIC) and varies across the different volatility measures.

2.5.5 Generalized Method of Moments (GMM)

The fifth method estimates the moments of information flow using the Generalized Methods of Moments. The estimated moments are then matched against the moments of trading volume and the number of trades. The GMM produces as a by-product the J-test of over-identifying restrictions which evaluates the goodness of fit of the moment restrictions underlying the MDH. The GMM procedure has been featured in earlier works such as Richardson and Smith (1994), Andersen (1996), Ané and Geman (2000), and Murphy and Izzeldin (2010).

The use of the GMM serves three main purposes in highlighting the information content of volume, number of trades and the constituents of trades. First, to estimate the moments of information conditional on market activity measures. Second, to compare whether the estimated moments are matched with the empirical moments and to see which activity measure best approximates its theoretical counterpart. Third, to observe system convergence and to test the overidentifying restrictions. Although we would not expect all market activity measures to result in system convergence, the degree of convergence using different activity measures has its own interest which we do not pursue here.

We focus on specifically the bivariate model. There are several advantages in using GMM to test the MDH. For instance, it imposes no distributional restrictions on information flow. Allowing for the joint test of restrictions also can also be evaluated using the Hansen J-test.

According to the MDH, when returns r_t and the observed market activity measures a_t are conditional on the information flow i_t, they are independently and normally distributed as:

$$\begin{pmatrix} r_t \\ a_t \end{pmatrix} | i_t \sim N\left(\begin{pmatrix} \mu_{r_t} i_t \\ \mu_{a_t} i_t \end{pmatrix}, \begin{pmatrix} \sigma_{r_t}^2 i_t & 0 \\ 0 & \sigma_{a_t}^2 i_t \end{pmatrix}\right) \qquad [2.13]$$

The model implies a set of moment restrictions which could be applied to the data and tested using the GMM J-test of over-identifying restrictions. Given the first four moments of returns and market activity measures, the general moment restrictions for both skewness and kurtosis for the sample moment vector $s_T(\theta)$ can be written as

$$s_T(\theta) = \begin{pmatrix} eq1r = r_t - \mu'_{r_t} \\ eq2r = (r_t - \mu'_{r_t})^2 - \sigma'^2_{r_t} - \mu'^2_{r_t} m'_2 \\ eq3r = (r_t - \mu'_{r_t})^3 - 3\mu'_{r_t}\sigma'^2_{r_t} m'_2 - \mu'^3_{r_t} m'_3 \\ eq4r = (r_t - \mu'_{r_t})^4 - 3\sigma'^4_{r_t}(1 + m'_2) - 6\mu'^2_{r_t}\sigma'^2_{r_t}(m'_3 + m'_2) - \mu'^4_{r_t} m'_4 \\ eq1a = a_t - \mu'_{a_t} \\ eq2a = (a_t - \mu'_{a_t})^2 - \sigma'^2_{a_t} - \mu'^2_{a_t} m'_2 \\ eq3a = (a_t - \mu'_{a_t})^3 - 3\mu'_{a_t}\sigma'^2_{a_t} m'_2 - \mu'^3_{a_t} m'_3 \\ eq4a = (a_t - \mu'_{a_t})^4 - 3\sigma'^4_{a_t}(1 + m'_2) - 6\mu'^2_{a_t}\sigma'^2_{a_t}(m'_3 + m'_2) - \mu'^4_{a_t} m'_4 \\ eq1r1a = (r_t - \mu'_{r_t})(a_t - \mu'_{a_t}) - \mu'_{r_t}\mu'^2_{a_t} m'_2 \\ eq2r1a = (r_t - \mu'_{r_t})^2(a_t - \mu'_{a_t}) - \sigma'^2_{r_t}\mu'_{a_t} m'_2 - \mu'^2_{r_t}\mu'_{a_t} m'_3 \\ eq1r2a = (r_t - \mu'_{r_t})(a_t - \mu'_{a_t})^2 - \sigma'^2_{a_t}\mu'_{r_t} m'_2 - \mu'^2_{a_t}\mu'_{r_t} m'_3 \\ eq3r1a = (r_t - \mu'_{r_t})^3(a_t - \mu'_{a_t}) - 3\mu'_{r_t}\mu'_{a_t}\sigma'^2_{r_t}(m'_3 + m'_2) - \mu'_{a_t}\mu'^3_{r_t} m'_3 \\ eq2r2a = (r_t - \mu'_{r_t})^2(a_t - \mu'_{a_t})^2 - \sigma'^2_{r_t}\sigma'^2_{a_t}(1 + m'_2) - (\sigma'^2_{a_t}\mu'^2_{r_t} + \sigma'^2_{r_t}\mu'^2_{a_t})(m'_3 + m'_2) - \mu'^2_{r_t}\mu'^2_{a_t} m'_4 \\ eq1r3a = (r_t - \mu'_{r_t})(a_t - \mu'_{a_t})^3 - 3\mu'_{a_t}\mu'_{r_t}\sigma'^2_{a_t}(m'_3 + m'_2) - \mu'_{r_t}\mu'^3_{a_t} m'_3 \end{pmatrix}$$

[2.14]

The first four equations denote the first four moments of return; the next four equations denote the first four moments of market activity measures and the last six equations denote the cross moments of return and market activity measures. The first four moments of considered here is written as

$$m_1^r = \mu_r m_1^i$$

$$m_2^r = \sigma_r^2 m_1^i + \mu_r^2 m_2^i$$

$$m_3^r = 3\mu_r \sigma_r^2 m_1^i + \mu_r^3 m_3^i$$

$$m_4^r = \mu_r^4 m_4^i + 6\sigma^2 m_3^i + 6\sigma^2\mu^2 m_1^i m_2^i + 3\sigma^4(m_2^i + (m_1^i)^2)$$

$$m_1^a = \mu_a m_1^i$$

$$m_2^a = \sigma_a^2 m_1^i + \mu_a^2 m_2^i$$

$$m_3^a = 3\mu_a\sigma_a^2 m_1^i + \mu_a^3 m_3^i$$

$$m_4^a = \mu_a^4 m_4^i + 6\sigma^2 m_3^i + 6\sigma^2\mu^2 m_1^i m_2^i + 3\sigma^4(m_2^i + (m_1^i)^2)$$

We also assume a non-stochastic mean for THE information flow process i_t which $m_1 = 1$ to overcome the identification problem. Therefore, we have

$$r_t' = r_t \cdot m_1, a_t' = a_t \cdot m_1, \sigma_{a_t}'^2 = \sigma_{a_t}^2 \cdot m_1, m_2' = m_2 / m_1^2, m_3' = m_3 / m_1^3,\ m_4' = m_4 / m_1^4$$

Following the normalization, we consider the combined restrictions (non-zero mean case) which have 14 moment conditions in total as follow:

Non-Zero & SKEWNESS+ = {eq1r, eq2r, eq3r, eq4r, eq1a, eq2a, eq3a, eq4a, eq1r1a, eq2r1a, eq1r2a, eq3r1a, eq2r2a, eq1r3a}

With 7 estimated parameters:

$$\theta = (\mu_{t_t}, \sigma_{r_t}'^2, \mu_{a_t}', \sigma_{a_t}', m_2', m_3', m_4')$$

which leaves us with 7 over-identifying restrictions.

The market activity measures whose moment restrictions fit the data best are taken as the best proxy of market activity.

2.5.6 Recovering Returns Normality

Clark (1973) argues that returns viewed in "volume" time are normal. Ané and Geman (2000) argue that returns standardized by number are normal. Izzeldin and Murphy (2010) follow the standardization procedure of Ané and Geman (2000) yet find that the number of trades standardized returns are not normally distributed. Here we shall replicate Ané and German (2000)'s procedure by standardizing returns on different market activity measures. Since the results of Ané *et al* (2000) cannot be replicated in any other paper, the task *per se* would not be expected to recover normality via standardization but rather would be to show which market activity measure(s) recover the return normality most.

Mathematically, this standardization is expressed as

$$r_t / \sqrt{a_t} \sim N(\mu_{r_t}, \sigma_{r_t}) \qquad [2.15]$$

where a_t is the market activity measures.

2.6 Empirical Results

2.6.1 Single Structural Break

Among 100 stocks, only two have the same structural break date for rv_t, vol_t and nt_t. Although in different sectors, both GPS and COST share the same breaking date of 24/07/2007. The commonality of break dates between vol_t and nt_t is greater than that for rv_t and vol_t / nt_t. 18/100 stocks are detected with the same break date for vol_t and

nt_t, whereas 7/100 stocks are detected with the same date for rv_t and vol_t and the number of stocks which have same date between rv_t and nt_t is only 3. In general, therefore, our results suggest that market activity measures do not tend to share break dates with rv_t; and that market activity measures (nt_t and vol_t) do tend to share break dates.

Break dates detected in rv_t, together with nt_t and vol_t, show quite distinct pictures from each other. **Figure 2.3a-e** plot the break dates for rv_t, vol_t and nt_t of every single stock and the detailed dates are provided in **Table 2.4**. The break dates are most discrete in rv_t and most condensed in nt_t. Another interesting finding from the ZA test is that each sector appears to have a common or closely proximate break date for stocks within the sector, at least for market activity measures. This conclusion holds especially in the case of nt_t. For instance, 03/01/2007 is detected as the break date of NT for 4/10 in the ENG sector. Within the same sector, 05/07/2007 and 02/01/2008 are detected for two stocks. Stocks from the same sectors tend to share close, if not the same date. One example could be the break date of rv_t from IT sectors. 7/10 stocks have break dates near 2002, when the internet bubble was prevailing.

To summarize, the single most significant structural break date does not necessarily share among rv_t, vol_t and nt_t. Nevertheless, for the same market activity measure, stocks within one certain sector are detected some common break dates. Even though the 2008 financial crisis has huge effect, when we estimate volatility measures, there are very few stocks that are detected with a break date during the peak period of the crisis. The most volatile days are not always associated with the most significant structural break point.

These results have important implications empirically. Since the break date is different from one measure to another, as well as from one stock to another, the detected date is vital for accurate evaluation of any economic/econometrics model which are intended to account for structural changes.

[Table 2.4 here]

[Figure 2.3a-e here]

2.6.2 OLS Regressions

The results of volume regressions were discussed in Chapter 1 and reported again in **Table 2.5a**. To summarize, pv_t is the most explained realized measures of volatility in the volume regression, with $\bar{R}^2$ at 67.8%. At sector level, ENG and MAR report the highest $\bar{R}^2$ across realized measures of volatility[8].

The results of regressing realized measures of volatility on the number of trades are reported in **Table 2.5b**. Similar to vol_t, nt_t explains a high proportion of realized measures of volatility. The average $\bar{R}^2$, unlike many existing literature, are not significantly higher than the average $\bar{R}^2$ from the trading volume regression. pv_t, again, is the most explained among all realized measures of volatility. The coefficient for nt_t is significant for all the stocks, with the exception of one stock when nt_t is

[8] One of worries regarding the model specification lies in the estimated residuals. In our case, most of the residual series from the regressions are not i.i.d. In the Jones et al (1994) paper which the model was firstly proposed, the authors stated that estimation method (OLS) provides consistent yet not necessarily efficient estimators of the parameters. Meanwhile, the Newey-West standard errors (not reported in the chapter) also confirm that the more than 95% of parameters are significant. Since the main concern in this section is to evaluate the performance of different volatility measures on an established econometrical model, we do not alter the specification of the model. Furthermore, the works of Chan and Fong (2006) and Giot et al (2010) do not report the performance of the residuals and treat them as i.i.d.s. Here we follow the same procedure to make our results more comparable to the literature.

regressed on realized range. In addition, 4 out of 10 sectors have higher $\bar{R}^2$ when regressing realized measures of volatility on vol_t than on nt_t, namely CD, CS, IT and UTL. The rest sectors though have higher $\bar{R}^2$ in nt_t regressions, the gain in $\bar{R}^2$ is marginal.

[Table 2.5a & 2.5b here]

Although nt_t does not necessarily have higher explanatory power than vol_t, we find that decomposing nt_t can enhance the goodness of fit. **Table 2.5c and 2.5d** report the results for up trades and down trades regressions, respectively. Comparing the results between up/down trades and number of trades as well as volume, the $\bar{R}^2$ of ut_t / dt_t regressions is approximately 2% higher over average, suggesting additional explanatory power on realized measures of volatility than vol_t and nt_t. The percentage of significant coefficients is between 98% and 100%. At sector level, all sectors show a higher $\bar{R}^2$ from ut_t / dt_t regressions than that from vol_t regressions.

[Table 2.5c & 2.5d here]

Table 2.5e shows the results for the same trades as the explanatory variable. In general, st_t has a lower explanatory power than the other 4 market activity measures. Even at sector level, the $\bar{R}^2$ from the st_t regression are consistently lower than that from the rest three trade measures and only 3 out of 10 sectors have higher $\bar{R}^2$ from st_t regressions than from vol_t regressions, namely, FIN, MAR and TEL.

[Table 2.5e here]

Combining the results from **Table 2.5a-e,** we find that the well-documented positive volatility-volume relation holds for all the realized measures of volatility and market activity measures. Market activity measures explain over 50% of variation in various realized measures of volatility on average. Nevertheless, in contrast to existing literature, our analysis shows the explanatory powers of these two measures are very similar, with nt_t slightly higher in terms of the number of significant coefficients. When we decompose nt_t into up, down and same trades, the regression results imply that it is the ut_t and dt_t that provide additional information, thereby giving nt_t the appearance of providing more information than vol_t. This result is intuitive, because, when prices change, the corresponding market activity measures shall carry more market information than the market measure when the prices remain the same. On the other hand, we find that, among different realized measures of volatility, the $\bar{R}^2$ of rv_t regressions are generally the lowest and the pv_t regressions generally provide the highest $\bar{R}^2$ across sectors and across market actively measures. The lower $\bar{R}^2$ of rv_t regressions might be explained by the existence of jumps components in the rv_t. Giot *et al* (2010) show that the positive volatility-volume relation does not hold for jump part. Thus, jump-contained rv_t may prove to be more difficult to be explained by the market activity measures than other realized measures of volatility. Izzeldin and Shi (2012) find the same conclusion as Giot *et al* (2010) using a different jump construction method[9].

2.6.3 GARCH and Augmented GARCH

[9] We also run the regression using the pre-filtering market activity measures (no-detrending), the conclusions remain the same. However, the regressions of detrended market activity measures produce higher adj. R^2.

The results of GARCH and Augmented GARCH models for the overall sample are provided in **Table 2.6a**. First of all, the ARCH coefficient α and GARCH coefficient β are significant for all 100 stocks in the sample for the GARCH (1, 1). The sum of $(\alpha + \beta)$ across all the sample stocks is positive and close to 1. This ensures the stationarity of the model and indicates a high level of persistence which is known as the "volatility clustering" in GARCH models. However, the GARCH effect remains significant when lagged market activity measures are augmented with GARCH (1, 1). The persistence of volatility remains strong and only decreases marginally when comparing $(\alpha + \beta)$ from augmented GARCH models with that from GARCH (1, 1). Augmenting market activity measures does not reduce the persistence of GARCH volatility. In our case, the significance of the augmented variables is rejected at the 5% level for many stocks. The number of market activity measures that significance is not rejected at the 5% level is 36/100 for vol_{t-1}, 60/100 for nt_{t-1}, 68/100 for ut_{t-1}, 67/100 for dt_{t-1} and 57/100 for st_{t-1}. The maximized log-likelihood (*LOGL*) value and Jarque-Bera statistics of standardized errors of trades-augmented GARCH equations are also smaller than those of vol_{t-1} augmented GARCH, indicating a better goodness of fit of trade measures over vol_{t-1}. However, *LOGL* and JB statistics of GARCH (1, 1) suggest when GARCH augmented with market activity measures, the market activity measures do not necessarily improve the overall performance of the augmented models.

At the sector level, the conclusion that market activity measures are not necessarily significant variables is further strengthened. The reports are presented in **Table 2.6b**. For instance, in the sector of Consumer Staples (CS), the augmented vol_{t-1} are not significantly greater than zero for all 10 stocks within the sector. In sector ENG and sector MAR, the number of augmented vol_{t-1} that is significantly greater than zero is one and two stocks respectively (both out of 10). FIN sector has the lowest rejection, 40%, when vol_{t-1} is augmented in the GARCH.

The results of trade measures augmented GARCH equations suggest a degree of improvement. When nt_{t-1} is used as the augmented variable, the number of rejections decreases for 8 out of 10 sectors. Taking CS sector as an example, the probability of significant augmented nt_{t-1} increases to 80% while that of augmented vol_{t-1} is 0%. The results of augmented ut_{t-1}, dt_{t-1} and st_{t-1} provide the same conclusion as nt_{t-1}.

Our finding contradicts that of L-L (1990) but is consistent with that of Chan *et al* (2001) and Marten *et al* (2003): the former paper uses contemporaneous volume while the latter two papers use lagged volume. L-L find a significant reduction of persistence (according to the sum of $(\alpha+\beta)$ and all the augmented vol_{t-1} in the model appear to be significant). On the other hand, Chan *et al* (2001) suggest that, although augmented volume contributes some explanation to the GARCH model, it does not reduce the persistence of the original model. Marten *et al* (2003) also finds that augmenting volume does not reduce the value of $(\alpha+\beta)$ as the augmented volume is not a significant variable in a paper where high frequency data is used to construct the return and GARCH series.

Further, our result is mostly in line with that of Martens and Luu (2003). The relatively low significance level for volume than for trade measures, especially up and down trades might be explained by the findings of Andersen (1996), that informational asymmetries and liquidity need motivated trading volume where there is a greater incidence of new information.

To summarize briefly, augmenting lagged market activity measures does not necessarily enhance the goodness of fit of the GARCH model. The persistence of GARCH (1, 1) is not reduced by augmenting extra market information proxies. Augmented trading volume is insignificant for over 60% of the sample stocks while, for the most parts, augmented traded measures are significant. Once again, ut_{t-1} and dt_{t-1} are the best performing market activity measures at overall and sector-average levels.

[Table 2.6a & 2.6b here]

2.6.4 Bivariate VAR Model

The linear Granger causality is first tested, using the most and least actively traded stocks respectively in our sample, namely C and OKE. In order to investigate causality at the overall sample average level, we use the average value of our 100-stock sample to conduct the same test.

The results are ambiguous. For OKE, the bi-directional Granger causality between trading volume and realized measures of volatility are all rejected at the 5% level. With only one exception, we find only unidirectional Granger causality from trading volume to realized measures of volatility, not reversely. This finding is not in line with many studies, which detect either a unidirectional Granger causality from volatility to volume, or bidirectional causality (see Garcia *et al* (1986), Brooks (1998), Fung and Patterson (1999), Martens *et al* (2003), Chiang *et al* (2010)). However, this finding is in line with a paper by Darrat *et al* (2003) which also finds significant causality from volume to volatility.

On the other hand, the test statistics of the Granger causality test, between the number of (up/down/same trades) and volatility measures, generally confirm the existence of bi-directional Granger causality[10].

In the case of OKE, we find only unidirectional Granger causality from volume to volatility measures but bi-directional Granger causality between the number of trades measures and volatility measures.

[Table 2.7a here]

The bi-directional Granger causality between volume and volatility measures is conclusively rejected when we test C. Again, the test statistics point to a unidirectional Granger causality from volume to realized measures of volatility. The results of causality between number of (up/down/same) trades and volatility measures are more ambiguous than the results from OKE. The null hypothesis, that the trade measures do not Granger cause volatility measures are all not rejected at the 5% level. However, for instance, realized range, realized power variation and realized bi-power variation do not Granger cause the number of trades at the 5% level. The similar conclusions could be found in $rv_t \rightarrow ut_t$, $rr_t \rightarrow ut_t$, $pv_t \rightarrow ut_t$, $bv_t \rightarrow ut_t$, $rv_t \rightarrow dt_t$, $pv_t \rightarrow dt_t$ and $pv_t \rightarrow st_t$.

[Table 2.7b here]

Due to ambiguity existing in the individual stocks, we test Granger causality using the average value of aggregated daily data of 100 stocks. For all the test results, the null hypothesis, that marker measures (volatility measures) do not Granger cause volatility

[10] The Granger causality from realized power variation to number of trades, from realized power variation to number of same trades, and from realized bi-power variation to number of trades, are all rejected at the 5% level.

measures (market measures), are rejected. The results point to a clear bi-directional Granger causality between market activity measures and volatility measures, and so, support MDH.

[Table 2.7c here]

2.6.5 GMM

Table 2.8 reports the GMM results of the number of estimated moments of the information flow i that the significance cannot be rejected and the number of over-identifying test that the significance cannot be rejected, both at the 5% level.

We first use J-test to evaluate the validity of MDH under GMM. Given the number of restrictions in Section 4, for $\chi_7^2 > 14.067$, we reject the null that the MDH restrictions imposed on information moments is valid. For some stocks, the estimated moments show significant *negative* signs, where positive signs are expected as volatility and market activity measures are positively correlated. Similar negative results for second and third moments are also reported by Richardson and Smith (1994). Most negative estimates occur in the fourth moment. The higher are the moments, the more difficult it is to obtain accurate estimates for GMM. Negative signs do not necessarily imply the rejection of the model. Therefore, we still use J-test statistics as one of the benchmarks for model validity.

The bivariate moments with ut_t produce the highest number of stocks for which the J-test is not rejected at the 5% level (35/100), followed by dt_t (32/100), st_t (32/100),

nt_t (28/100) and vol_t (19/100). These results give more support for ut_t and dt_t as dominant factors in MDH. An interesting finding from the J-test result at sector level, is that IT and TEL, the sectors with the largest and smallest number of jumps, are the sectors having the highest number of J-test rejections. This result implies that the presence of jumps has little impact in recovering the estimated moments of information flows.

We further consider the estimated moments to assess the validity of the MDH. In the 1994 paper by Richardson and Smith, several stylized facts of the bivariate model are documented. Our results from different market volatility measures support all the stylized facts: the unobservable information flow shows: 1) small variations relative to the mean ($m2$); 2) positive skewness ($m3$); and 3) large kurtosis ($m4$). Furthermore, in line with the literature, the second and third moments are better recovered using bivariate model (over 95% for the second moment and over 85% for the third). The fourth moment is not recovered for more than one third of the sample stocks, many of which are estimated with extremely large and implausible values. For the second and third moments, bivariate results using ut_t and dt_t appear the best. st_t performs best in recovering the fourth moments. vol_t and nt_t produce very similar results with nt_t slightly better in the second and third moments.

Finally, the empirical moments of the re-centred market activity measures are computed and compared with the estimated moments. For the second and third moments, the difference between empirical and estimated moments is smaller than the difference for the fourth moments. Moreover, the second moments of the empirical and of the bivariate methods are closer for those stocks that the J test is not rejected at the 5% level in many cases. Among all the market activity measures, the second moment, both empirical and estimated, of vol_t, is the closet.

[Table 2.8 here]

2.6.6 Recovering Returns Normality

Similar to papers that followed Ané and Geman (2000), our results contrast with the 2000 paper. We standardize daily return series by re-centered market activity measures. Also different from the results reported in a previous chapter, where realized measures of volatility are used as the standardization factor, the normality of market activity measures standardized returns is rejected at the 5% level for all the stocks according to the Jarque-Bera statistics for normality. The normality of return series standardized by volume has been improved greatly across the sample. On the other hand, when return series are standardized by trade measures, the normality is further distorted in most cases. **Table 2.9** shows results for the sector average returns and standardized returns. The JB statistics further confirm the results from the individual stocks that all five market activity measures are not able to recover the normality of daily returns under the direct standardization procedure. The normality of volume standardized returns is a further improvement in comparison with that of raw returns and of returns standardized by trade measures.

The results suggest volume is a better market information proxy than trade measures according to the standardization procedure. The conclusion of Ané and Geman (2000) that returns standardized by re-centered number of trades could be normal is not found in our data.

[Table 2.9 here]

2.7 Conclusion

We apply a series of tests and models to test the validity of MDH using high frequency data. The volatility-volume relation is re-examined, using various realized measures of volatility and different market activity measures. Four realized measures of volatility, namely realized variance, realized range, and realized power and bipower variations, and five market activity measure, namely trading volume, number of trades and its constituents (up/down/same trades), are tested for 100 stocks over an 11-year sample period . We address following issues: 1) correlations between realized measures of volatility and market activity measures; 2) common break-dates; 3) OLS regressions proposed by Jones *et al* (1994) and Chan and Fong (2006); 4) GARCH augmented with realized volatility measures; 5) Granger Causality; 6) bivariate estimation of MDH model using GMM; 7) recovering Return Normality.

The conclusions reached are as follow:

In most sectors, volume correlates more with volatility than the number of trades. Trading volume is highly correlated with trade measures. On the other hand, the correlation between the number of up/down trades and various realized measures of volatility dominates that of volume.

Realized variance, trading volume and number of trades usually do not have common break-dates. Break-dates are more common between realized variance and trading volume relative to the number of trades. Within the same sector, stocks tend to have close or same break- dates, at lease for market activity measures.

The numbers of up trades and down trades explain more volatility than volume or the number of trades. The number of same trades explains volatility most poorly. This conclusion holds for every stock in the whole sample.

Augmenting lagged market activity measures does not necessarily improve the model fitness of GARCH (1, 1). The significant reduction in model persistence is no longer found. Lagged volume is not a significant variable in more than 60% of stocks. The number of up trades and the number of down trades also outperform the rest market activity measures in terms of the percentage of significant variables and Log Likelihood statistics.

Bi-directional Granger causality is found when the overall average realized measures of volatility and market activity measures are tested. At individual stock level, the test results point to unidirectional Granger causality from market activity measures to realized measures of volatility, yet not in reverse for any of the realized measures of volatility to market activity measures.

The bivariate MDH model is generally supported. In the case of a non-zero mean, the second and third moments of the information are much easier to recover than the fourth moments. Again, the best results are given, when the numbers of up/down trades are used in the bivariate model. When comparing the empirical moments of the re-centred market activity measures with the estimated ones.

The normality of returns standardized by market activity measures is rejected in every case. However, returns that are standardized by trading volume are closer to the normal distribution.

The test results do not always lead to the same conclusion yet patterns emerge. Overall, the MDH is favourably supported by realized volatility measures across sectors and remains strong in the presence of crises. The volatility-volume relation is held no matter the level of the market activity measures yet the relation tends to appear stronger in those highly actively traded sectors. Realized power variation provides the strongest support of the MDH among the realized volatility measures while the number of up trades and the number of down trades are the main driving factors that make the number of trades a more informative market activity measure than trading volume. The number of trades is only more informative than trading volume when number of trades happens as a result of a price moving.

Tables

Table 2.1a: Summary statistics of market activity measures

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					Trading Volume (vol_t)					
Mean	6.674	4.325	6.249	24.828	9.462	7.387	30.564	5.347	6.585	2.062
S.D.	2.678	1.806	3.296	38.161	3.984	5.375	8.471	4.005	4.711	1.018
Skewness	1.180	1.877	1.617	2.641	1.519	3.663	1.108	1.615	1.684	1.636
Kurtosis	4.893	8.953	8.348	11.024	6.788	28.916	6.087	5.965	7.358	7.672
Jarque-Bera	1056	5710	4503	10640	2719	83623	1665	2216	3498	3751
ADF	-3.082	-3.902	-3.056	-2.493	-4.213	-3.105	-7.209	-2.210	-1.929	-5.387
p value	0.028	0.002	0.030	0.117	0.001	0.026	0.000	0.203	0.319	0.000
ACF(36)	0.553	0.475	0.638	0.72	0.463	0.561	0.058	0.787	0.698	0.494
Q(36)	36939	32675	50899	60010	29896	43285	5822	68584	54502	34364
					Number of Trades (nt_t)					
Mean	1.873	1.319	2.021	4.527	2.195	1.766	5.702	1.750	1.283	0.670
S.D.	1.943	1.441	2.458	7.244	2.168	2.206	3.805	2.178	1.509	0.733
Skewness	1.417	1.635	1.664	2.438	1.358	2.040	1.703	1.378	1.279	1.489
Kurtosis	4.704	5.941	6.585	10.284	4.627	8.375	7.390	4.091	3.935	5.395
Jarque-Bera	1261	2230	2758	8858	1156	5250	3560	1013	856	1683
ADF	-1.673	-2.029	-1.851	-1.957	-1.841	-1.988	-2.672	-1.807	-1.404	-1.862
p value	0.445	0.274	0.356	0.306	0.361	0.292	0.079	0.378	0.582	0.351
ACF(36)	0.861	0.84	0.853	0.8	0.839	0.81	0.725	0.87	0.865	0.826
Q(36)	79775	77974	80187	69604	77269	73930	63733	81559	80452	75796
					Number of Up Trades (ut_t)					
Mean	0.243	0.201	0.323	0.532	0.297	0.266	0.666	0.284	0.146	0.114
S.D.	0.203	0.205	0.391	0.773	0.224	0.278	0.364	0.322	0.145	0.118
Skewness	1.965	3.027	2.718	3.196	2.126	2.317	2.216	1.661	1.631	2.315
Kurtosis	9.680	19.115	16.449	20.768	12.098	12.055	13.025	6.861	6.525	12.616
Jarque-Bera	6925	34165	24261	41108	11627	11930	13850	2991	2659	13133
ADF	-2.105	-2.516	-2.369	-2.638	-2.515	-2.354	-4.294	-2.409	-1.690	-2.414
p value	0.243	0.112	0.151	0.085	0.112	0.155	0.001	0.139	0.436	0.138
ACF(36)	0.768	0.717	0.762	0.689	0.722	0.755	0.547	0.795	0.808	0.735
Q(36)	66052	63437	69263	55953	62136	66327	43993	71944	73460	64367
					Number of Down Trades (dt_t)					
Mean	0.243	0.201	0.322	0.531	0.296	0.264	0.666	0.282	0.146	0.114
S.D.	0.203	0.206	0.392	0.774	0.225	0.279	0.365	0.323	0.145	0.118
Skewness	1.955	3.007	2.702	3.186	2.118	2.303	2.212	1.655	1.628	2.313
Kurtosis	9.552	18.772	16.101	20.629	11.911	11.844	12.957	6.751	6.468	12.485
Jarque-Bera	6713	32849	23154	40511	11225	11464	13686	2884	2609	12840
ADF	-2.108	-2.517	-2.370	-2.636	-2.508	-2.344	-4.288	-2.408	-1.696	-2.410
p value	0.242	0.112	0.151	0.086	0.114	0.158	0.001	0.140	0.433	0.139
ACF(36)	0.768	0.717	0.763	0.69	0.723	0.756	0.548	0.796	0.809	0.735
Q(36)	66280	63630	69391	56033	62318	66569	44041	72085	73589	64413
					Number of Same Trades (st_t)					
Mean	1.387	0.917	1.376	3.465	1.602	1.235	4.370	1.183	0.991	0.442
S.D.	1.548	1.050	1.708	5.755	1.738	1.666	3.226	1.564	1.225	0.507
Skewness	1.341	1.361	1.384	2.388	1.260	2.065	1.387	1.424	1.243	1.301
Kurtosis	4.203	4.230	4.534	9.604	3.949	8.442	5.767	4.104	3.708	3.998
Jarque-Bera	996	1029	1155	7659	836	5380	1770	1076	770	895
ADF	-1.626	-1.800	-1.704	-1.848	-1.731	-1.984	-2.431	-1.650	-1.333	-1.750
p value	0.469	0.381	0.429	0.358	0.415	0.294	0.133	0.457	0.616	0.406
ACF(36)	0.876	0.870	0.881	0.819	0.858	0.818	0.769	0.889	0.871	0.857
Q(36)	82205	81725	83757	72305	79956	75243	69011	84214	81234	79732

Note: Summary statistics of daily market activity measures for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation. ADF is the augmented Dickey-Fuller statistics for the null of a unit root with 5% and 1% critical values of 2.862 and -3.433 respectively. ACF (36) is the autocorrelation at 36th lags and Q (36) is the Q statistics at 36th lag. Volume is rescaled by dividing by 1,000,000 and trade measures are rescaled by dividing by 10,000.

Table 2.1b: Summary statistics of logarithmic market activity measures

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					Trading Volume (vol_t)					
Mean	1.824	1.392	1.709	2.455	2.170	1.836	3.383	1.443	-0.370	0.617
Maximum	3.166	2.892	3.610	5.729	3.636	4.437	4.416	3.605	2.484	2.229
S.D.	0.383	0.369	0.494	1.105	0.388	0.532	0.275	0.666	1.448	0.458
Skewness	0.087	0.412	0.131	1.050	0.231	0.823	-0.363	0.389	0.199	0.124
Kurtosis	2.916	3.560	2.530	2.763	3.115	3.757	5.463	2.319	1.653	2.915
Jarque-Bera	4.316	114.311	33.369	515.115	26.097	378.249	760.169	123.024	227.609	7.956
ADF	-3.353	-3.807	-2.637	-1.167	-4.307	-2.815	-6.898	-1.838	-1.142	-3.492
p value	0.013	0.003	0.086	0.691	0.000	0.056	0.000	0.362	0.701	0.008
ACF(36)	0.562	0.484	0.693	0.886	0.485	0.664	0.059	0.825	0.957	0.552
Q(36)	36469	31630	56151	83436	31607	51415	4966	73249	94560	38792
					Number of Trades (nt_t)					
Mean	0.105	-0.291	-0.193	0.289	0.288	-0.143	1.550	-0.550	-2.362	-1.043
Maximum	2.631	2.377	3.052	4.057	2.792	2.903	3.546	2.225	0.300	1.646
S.D.	1.023	1.068	1.449	1.575	1.024	1.194	0.609	1.305	0.924	1.186
Skewness	0.366	0.351	0.097	0.561	0.152	0.397	0.229	0.355	0.419	0.085
Kurtosis	1.728	1.780	1.594	1.842	1.815	1.845	2.426	1.514	1.870	1.885
Jarque-Bera	248.336	228.382	232.343	299.655	172.571	226.588	62.168	312.776	228.266	146.783
ADF	-1.046	-1.237	-1.377	-0.962	-1.328	-1.196	-2.371	-0.602	-0.923	-1.556
p value	0.739	0.660	0.595	0.769	0.619	0.678	0.150	0.868	0.781	0.505
ACF(36)	0.938	0.937	0.955	0.954	0.926	0.942	0.802	0.957	0.924	0.934
Q(36)	90415	91197	94523	93994	89613	91790	71414	94296	89273	91378
					Number of Up Trades (ut_t)					
Mean	-1.690	-1.967	-1.817	-1.421	-1.459	-1.770	-0.530	-1.965	-2.362	-2.653
Maximum	0.829	0.967	1.572	2.249	1.005	1.150	1.463	1.010	0.300	0.193
S.D.	0.718	0.825	1.242	1.219	0.705	0.936	0.492	1.258	0.924	1.028
Skewness	0.506	0.416	-0.002	0.553	0.001	0.281	0.137	0.062	0.419	-0.103
Kurtosis	2.207	2.480	1.930	2.118	2.613	2.192	2.911	1.800	1.870	2.383
Jarque-Bera	190.651	111.000	132.008	230.728	17.252	111.847	9.610	167.779	228.266	48.877
ADF	-1.691	-1.780	-1.702	-1.500	-1.933	-1.569	-3.145	-1.422	-0.923	-1.977
p value	0.436	0.3907	0.4302	0.5336	0.3171	0.4984	0.0235	0.5728	0.7813	0.2972
ACF(36)	0.874	0.892	0.936	0.923	0.854	0.909	0.68	0.943	0.924	0.907
Q(36)	79797	84461	91720	89453	78809	86871	55584	92545	89273	87780
					Number of Down Trades (dt_t)					
Mean	-1.697	-1.978	-1.834	-1.434	-1.465	-1.785	-0.531	-1.983	-2.372	-2.671
Maximum	0.824	0.957	1.560	2.248	0.996	1.141	1.460	0.998	0.290	0.181
S.D.	0.722	0.832	1.250	1.229	0.709	0.944	0.493	1.269	0.930	1.038
Skewness	0.509	0.423	0.015	0.553	0.010	0.302	0.137	0.075	0.425	-0.079
Kurtosis	2.196	2.453	1.911	2.102	2.592	2.172	2.909	1.782	1.862	2.334
Jarque-Bera	194.141	117.101	136.711	233.912	19.253	121.189	9.621	173.528	232.399	54.038
ADF	-1.688	-1.764	-1.683	-1.484	-2.106	-1.544	-3.145	-1.393	-0.918	-1.951
p value	0.437	0.3986	0.44	0.5421	0.2423	0.511	0.0235	0.5875	0.7831	0.3088
ACF(36)	0.875	0.893	0.936	0.924	0.855	0.91	0.681	0.944	0.924	0.908
Q(36)	79954	84654	91772	89591	78944	87029	55639	92646	89374	87891
					Number of Same Trades (st_t)					
Mean	-0.347	-0.798	-0.727	-0.223	-0.201	-0.685	1.202	-0.914	-1.009	-1.009
Maximum	2.232	1.736	2.449	3.660	2.387	2.639	3.260	2.224	1.882	1.882
S.D.	1.193	1.225	1.586	1.771	1.216	1.364	0.763	1.572	1.502	1.502
Skewness	0.282	0.308	0.140	0.531	0.144	0.394	-0.103	0.250	0.308	0.308
Kurtosis	1.630	1.626	1.491	1.749	1.646	1.729	2.077	1.620	1.444	1.444
Jarque-Bera	253.307	261.273	271.514	310.248	220.959	257.720	103.094	248.259	322.959	322.959
ADF	-0.874	-1.030	-1.180	-0.799	-1.111	-1.008	-1.996	-0.979	-0.502	-0.502
p value	0.797	0.7444	0.6851	0.8189	0.7139	0.7525	0.2888	0.7627	0.8884	0.8884
ACF(36)	0.951	0.949	0.962	0.961	0.944	0.952	0.866	0.961	0.964	0.964
Q(36)	92706	93050	95493	95166	92296	93337	80808	95249	95362	95362

Note: Summary statistics of daily logarithmic market activity measures for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation. ADF is the augmented Dickey-Fuller statistics for the null of a unit root with 5% and 1% critical values of 2.862 and -3.433 respectively. ACF (36) is the autocorrelation at 36th lags and Q (36) is the Q statistics at 36th lag. Volume is rescaled by dividing by 1,000,000 and trade measures are rescaled by dividing by 10,000.

Table 2.2: Long memory test

		CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	Overall
Market Measures	vol_t	0.342	0.347	0.419	0.443	0.358	0.379	0.334	0.445	0.419	0.389	0.388
	nt_t	0.469	0.471	0.480	0.465	0.461	0.479	0.461	0.480	0.474	0.468	0.471
	ut_t	0.462	0.464	0.475	0.454	0.447	0.474	0.452	0.473	0.474	0.460	0.463
	dt_t	0.462	0.464	0.475	0.454	0.447	0.475	0.452	0.473	0.474	0.460	0.464
	st_t	0.468	0.472	0.481	0.469	0.466	0.480	0.468	0.481	0.472	0.472	0.473

Note: Table reports the sector average and overall average Robinson's *d* statistics of long memory test. VOL= volume, NT= number of trades, UT = number of up trades, DT = number of down trades, ST = number of same trades. All *d* statistics are significant at 5%.

Table 2.3: Correlation matrix

	vol_t	nt_t	ut_t	dt_t	st_t	rv_t	rr_t	pv_t	bv_t
vol_t	**1.000**								
p value	-								
nt_t	**0.739**	**1.000**							
p value	(0.000)	-							
nt_t	**0.738**	**0.924**	**1.000**						
p value	(0.000)	(0.000)	-						
dt_t	**0.737**	**0.924**	**0.999**	**1.000**					
p value	(0.000)	(0.000)	(0.000)	-					
st_t	**0.719**	**0.991**	**0.874**	**0.874**	**1.000**				
p value	0.004	(0.000)	(0.000)	(0.000)	-				
rv_t	**0.394**	**0.312**	**0.414**	**0.413**	**0.263**	**1.000**			
p value	(0.000)	(0.029)	(0.011)	(0.011)	0.036	-			
rr_t	**0.416**	**0.362**	**0.470**	**0.470**	**0.310**	**0.924**	**1.000**		
p value	(0.000)	(0.013)	(0.019)	(0.019)	(0.041)	(0.000)	-		
pv_t	**0.426**	**0.345**	**0.446**	**0.445**	**0.295**	**0.955**	**0.899**	**1.000**	
p value	(0.000)	(0.025)	(0.009)	(0.010)	(0.027)	(0.000)	(0.000)	-	
bv_t	**0.404**	**0.325**	**0.429**	**0.429**	**0.274**	**0.963**	**0.920**	**0.958**	**1.000**
p value	(0.000)	(0.029)	(0.010)	(0.010)	(0.031)	(0.000)	(0.000)	(0.000)	-

Note: This table reports the overall average correlation between realized volatility measures and market activity measures.

Table 2.4: Common structural break dates for volatility and market activity measures

		Zivot & Andrews Single Structure Break Test							
		rv_t	vol_t	nt_t			rv_t	vol_t	nt_t
Consumer Discretionary	AMZN	20011130	20070423	20070423	Consumer Staples	AVP	20080728	20080102	20071019
	BBY	20021227	20070618	20070618		BFB	20070723	20090421	20070227
	DIS	20080902	20080102	20080102		COST	20070724	20070724	20070724
	GPS	20070724	20070724	20070724		EL	20071018	20080103	20080102
	HD	20071017	20060628	20070620		KMB	20071019	20070709	20070709
	IPG	20080409	20071031	20071031		KO	20080101	20070212	20070709
	MAR	20080104	20080528	20080523		PEP	20070711	20070710	20070717
	MCD	20030507	20060908	20070720		PG	20080102	20070419	20070709
	NWSA	20080627	20081002	20080902		UL	20080104	20080903	20080613
Energy	BHI	20080619	20060421	20070103	Financials	ALL	20080905	20070717	20070717
	CHK	20080619	20080416	20080208		AXP	20071031	20071031	20071009
	CVX	20090903	20090511	20070705		BAC	20080906	20090112	20080602
	DVN	20080102	20090511	20080102		BK	20080107	20080102	20080102
	HAL	20080702	20050722	20070103		C	20080627	20090219	20071009
	OXY	20071011	20090427	20070103		GS	20070724	20070718	20070710
	SUN	20090319	20090511	20070103		JPM	20071031	20080102	20080102
	TEA	20080902	20040211	20070718		MS	20080707	20071031	20071031
	WMB	20030327	20030813	20080102		TRV	20070720	20080619	20080102
	XOM	20070720	20090511	20070705		WFC	20071231	20080603	20080603
Health Care	ABT	20070710	20080102	20080102	Industrials	BA	20080612	20070720	20070720
	AMGN	20021008	20011120	20090427		CAT	20080902	20080902	20080102
	BSX	20080707	20090512	20080619		CMI	20070725	20090511	20070420
	GILD	20010924	20040202	20070709		GD	20070724	20030717	20070521
	HUM	20080115	20080102	20080102		GE	20080902	20080904	20080102
	JNJ	20080904	20070216	20070216		HON	20080625	20010924	20070417
	MDT	20080829	20060323	20070717		LUV	20080102	20061019	20080102
	MRK	20080117	20080104	20070718		MMM	20071009	20030930	20070710
	PFE	20080625	20080417	20080102		UPS	20080618	20090424	20070709
	UNH	20080619	20080311	20080102		UTX	20080102	20080102	20070709
Information Technology	AAPL	20070720	20090123	20090123	Materials	AA	20080902	20080925	20070705
	CSCO	20021031	20030324	20071031		AKS	20080701	20080902	20080624
	DELL	20020726	20011119	20071016		DD	20080102	20080102	20070706
	EMC	20020512	20061109	20070524		DOW	20080915	20080902	20080618
	HPQ	20030318	20080104	20080102		FCX	20070720	20070226	20070221
	IBM	20071011	20070103	20070222		IP	20080902	20080304	20080102
	INTC	20021107	20020607	20080102		NEM	20071031	20090323	20070706
	MSFT	20071011	20071015	20071015		NUE	20090320	20090424	20070719
	ORCL	20021209	20040305	20080102		WY	20070720	20061214	20070119
	XRX	20021024	20090506	20080506		X	20071231	20080902	20080701
Telecommunications	AMT	20030128	20090508	20070718	Utilities	AEP	20030127	20030528	20080102
	BT	20071224	20090428	20061127		CEG	20080812	20080424	20070523
	CTL	20070725	20081002	20080929		DUK	20030214	20070103	20070522
	FTR	20021216	20090512	20071231		ETR	20070719	20070518	20070129
	Q	20030325	20021121	20070607		EXC	20070521	20030529	20070524
	S	20080108	20080116	20070103		OKE	20080908	20030109	20070203
	T	20030403	20060921	20070601		PCG	20010924	20070606	20070605
	TEF	20080104	20010924	20071011		PEG	20021018	20020708	20071231
	VOD	20021114	20080915	20080903		PGN	20070604	20070522	20070523
	VZ	20071211	20080104	20080102		SO	20021114	20070523	20070522

Table 2.5a: OLS regressions of volume on realized volatility measures (sector and overall average)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	OVERALL
						Realized Variance (rv_t)					
$\overline{\phi}_{it}$	0.652	0.884	0.862	0.754	0.366	0.546	0.148	0.502	0.987	6.524	1.222
% $\overline{R}^2$	45.209	47.722	63.934	58.489	48.804	56.552	58.540	61.338	41.284	45.708	52.758
% Significant	90	100	100	100	100	100	100	100	90	80	96
						Realized Range (rr_t)					
$\overline{\phi}_{it}$	0.492	0.742	0.754	0.684	0.336	0.500	0.147	0.440	0.784	5.305	1.018
% $\overline{R}^2$	53.103	48.450	64.393	62.929	51.572	53.684	60.287	66.735	47.463	37.977	54.659
% Significant	90	100	100	100	100	100	100	100	90	80	96
						Realized Power Variation (pv_t)					
$\overline{\phi}_{it}$	0.488	0.845	0.737	0.540	0.368	0.585	0.161	0.485	0.953	2.691	0.785
% $\overline{R}^2$	65.675	64.767	73.242	71.911	64.822	70.210	70.751	72.310	59.947	64.193	67.783
% Significant	100	100	100	100	100	100	100	100	90	90	98
						Realized Bipower Variation (bv_t)					
$\overline{\phi}_{it}$	0.416	0.699	0.792	0.679	0.344	0.529	0.140	0.473	0.865	6.831	1.177
% $\overline{R}^2$	53.658	52.003	64.847	59.713	51.452	57.629	58.568	62.895	47.391	51.795	55.995
% Significant	100	100	100	100	100	100	100	100	90	90	98

Note: Volume is rescaled by dividing by 1,000,000. ρ_{ij} measures the persistence of volatility shock at lag *j*. M_t is the Monday dummy. $\overline{\phi}_{it}$ is equally-weighted cross sectional mean coefficients for volume. The $\overline{R}^2$ is the mean value of 100 stocks. The last row reports the percentage of $\hat{\phi}_{it}$ coefficients which are statistically significantly from zero at 5% level.

Table 2.5b: OLS regressions of number of trades on realized volatility measures (sector and overall average)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	OVERALL
						Realized Variance (rv_t)					
$\bar{\beta}_{it}$	1.915	3.336	1.888	2.144	1.069	1.916	0.359	2.267	3.348	11.081	2.932
% $\bar{R}^2$	44.814	46.410	63.657	60.495	49.642	57.389	57.328	63.129	42.025	44.460	52.935
% Significant	100	100	100	100	100	100	100	100	100	100	100
						Realized Range (rr_t)					
$\bar{\beta}_{it}$	1.665	2.753	1.784	1.910	1.043	1.866	0.368	2.024	3.056	10.400	2.687
% $\bar{R}^2$	52.930	48.066	64.681	64.550	52.762	55.233	59.109	68.639	48.610	36.871	55.145
% Significant	100	100	100	100	100	100	100	100	100	99	99
						Realized Power Variation (pv_t)					
$\bar{\beta}_{it}$	1.693	3.113	1.869	1.536	1.074	1.952	0.406	2.124	3.601	5.161	2.253
% $\bar{R}^2$	65.357	64.306	73.162	73.389	65.238	70.642	69.603	73.754	60.649	62.963	67.906
% Significant	100	100	100	100	100	100	100	100	100	100	100
						Realized Power Variation (bv_t)					
$\bar{\beta}_{it}$	1.493	2.575	1.745	1.948	0.983	1.767	0.343	2.117	2.980	11.572	2.752
% $\bar{R}^2$	53.680	51.040	64.668	61.694	52.238	58.160	57.418	64.617	48.108	49.922	56.154
% Significant	100	100	100	100	100	100	100	100	100	100	100

Note: Number of trades is rescaled by dividing by 10,000. ρ_{ij} measures the persistence of volatility shock at lag *j*. M_t is the Monday dummy. $\bar{\beta}_{it}$ is equally-weighted cross sectional mean coefficients for number of trades. The $\bar{R}^2$ is the mean value of 100 stocks. The last row reports the percentage of $\hat{\beta}_{it}$ coefficients which are statistically significantly from zero at 5% level.

Table 2.5c: OLS regressions of number of up trades on realized volatility measures (sector and overall average)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	OVERALL
	Realized Variance (rv_t)										
$\bar{\gamma}_{it}$	23.635	16.598	9.979	15.703	10.304	12.914	5.554	15.725	30.323	57.447	19.818
% $\bar{R}^2$	46.283	48.116	64.286	62.626	52.889	59.579	59.637	64.283	43.613	46.890	54.820
% Significant	100	100	100	100	100	100	100	100	100	90	99
	Realized Range (rr_t)										
$\bar{\gamma}_{it}$	21.463	15.035	9.767	14.146	10.331	12.863	6.051	14.547	27.129	52.268	18.360
% $\bar{R}^2$	54.986	50.445	65.510	66.486	57.213	58.251	61.646	70.052	50.656	39.331	57.458
% Significant	100	100	100	100	100	100	100	100	100	80	98
	Realized Power Variation (pv_t)										
$\bar{\gamma}_{it}$	20.409	16.601	9.676	11.478	10.253	13.000	5.880	14.428	28.353	26.955	15.703
% $\bar{R}^2$	66.500	65.507	73.615	75.084	67.340	72.240	71.436	74.827	61.746	64.926	69.322
% Significant	100	100	100	100	100	100	100	100	100	100	100
	Realized Power Variation (bv_t)										
$\bar{\gamma}_{it}$	19.369	13.423	9.189	14.502	9.380	12.268	5.432	14.737	26.486	60.692	18.548
% $\bar{R}^2$	55.381	52.950	65.225	63.928	55.453	60.441	59.752	65.794	49.633	52.351	58.091
% Significant	100	100	100	100	100	100	100	100	100	90	99

Note: Number of up trades is rescaled by dividing by 10,000. ρ_{ij} measures the persistence of volatility shock at lag *j*. M_t is the Monday dummy. $\bar{\gamma}_{it}$ is equally-weighted cross sectional mean coefficients for number of up trades. The $\bar{R}^2$ is the mean value of 100 stocks. The last row reports the percentage of $\hat{\gamma}_{it}$ coefficients which are statistically significantly from zero at 5% level.

Table 2.5d: OLS regressions of number of down trades on realized volatility measures (sector and overall average)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	OVERAL
						Realized Variance (rv_t)					
$\bar{\theta}_{it}$	24.060	16.213	9.976	15.632	10.298	12.877	5.548	15.664	30.227	56.856	19.735
% $\bar{R}^2$	46.900	47.984	64.258	62.552	52.874	59.482	59.643	64.253	43.620	46.811	54.838
% Significant	100	100	100	100	100	100	100	100	100	80	98
						Realized Range (rr_t)					
$\bar{\theta}_{it}$	22.006	14.754	9.738	14.107	10.317	12.837	6.050	14.478	27.071	52.062	18.342
% $\bar{R}^2$	55.985	50.275	65.455	66.441	57.165	58.185	61.660	70.014	50.665	39.307	57.515
% Significant	100	100	100	100	100	100	100	100	100	90	99
						Realized Power Variation (pv_t)					
$\bar{\theta}_{it}$	20.409	16.601	9.672	11.424	10.248	12.960	5.874	14.376	28.279	26.732	15.674
% $\bar{R}^2$	66.500	65.507	73.604	75.031	67.331	72.176	71.443	74.805	61.758	64.877	69.344
% Significant	100	100	100	100	100	100	100	100	100	100	100
						Realized Power Variation (bv_t)					
$\bar{\theta}_{it}$	19.754	13.156	9.188	14.440	9.376	12.214	5.426	14.679	26.499	60.054	18.479
$\overline{se}_\theta$	4.931	3.046	2.524	4.870	2.212	2.873	1.155	3.510	8.459	29.037	6.262
% $\bar{R}^2$	56.193	52.828	65.199	63.863	55.440	60.335	59.760	65.766	49.649	52.271	58.130
% Significant	100	100	100	100	100	100	100	100	100	100	100

Note: Number of down trades is rescaled by dividing by 10,000. ρ_{ij} measures the persistence of volatility shock at lag *j*. M_t is the Monday dummy. $\bar{\theta}_{it}$ is equally-weighted cross sectional mean coefficients for number of down trades. The $\bar{R}^2$ is the mean value of 100 stocks. The last row reports the percentage of $\hat{\theta}_{it}$ coefficients which are statistically significantly from zero at 5% level.

Table 2.5e: OLS regressions of no. of same trades on realized volatility measures (sector and overall average)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	OVERALL
						Realized Variance (rv_t)					
$\overline{\tau}_{it}$	2.120	5.005	2.077	2.513	1.232	2.292	0.367	2.462	4.072	26.186	4.833
% $\overline{R}^2$	43.931	44.673	62.177	58.778	48.130	55.377	56.665	61.711	41.378	43.331	51.615
% Significant	100	100	100	100	100	100	100	100	100	80	98
						Realized Range (rr_t)					
$\overline{\tau}_{it}$	1.814	3.822	1.935	2.245	1.178	2.156	0.365	2.138	3.736	22.511	4.190
% $\overline{R}^2$	51.804	45.734	63.000	63.044	50.952	52.894	58.433	67.170	47.730	35.583	53.634
% Significant	100	100	100	100	100	100	100	100	100	80	98
						Realized Power Variation (pv_t)					
$\overline{\tau}_{it}$	1.876	4.260	2.079	1.848	1.234	2.331	0.424	2.330	4.431	10.091	3.090
% $\overline{R}^2$	64.604	62.821	71.881	72.167	64.080	69.070	69.041	72.498	60.139	61.977	66.828
% Significant	100	100	100	100	100	100	100	100	100	80	98
						Realized Power Variation (bv_t)					
$\overline{\tau}_{it}$	1.631	3.644	1.924	2.259	1.140	2.074	0.345	2.288	3.618	28.640	4.757
% $\overline{R}^2$	52.725	49.178	63.220	60.077	50.726	56.170	56.762	63.248	47.504	48.861	54.847
% Significant	100	100	100	100	100	100	100	100	100	80	98

Note: Number of same trades is rescaled by dividing by 10,000. ρ_{ij} measures the persistence of volatility shock at lag j. M_t is the Monday dummy. $\overline{\tau}_{it}$ is equally-weighted cross sectional mean coefficients for number of same trades. The $\overline{R}^2$ is the mean value of 100 stocks. The last row reports the percentage of $\hat{\tau}_{it}$ coefficients which are statistically significantly from zero at 5% level.

Table 2.6a: GARCH and augmented GARCH results (overall average)

	GARCH(1,1)	Augmented with vol_t	Augmented with nt_t	Augmented with ut_t	Augmented with dt_t	Augmented with st_t
			Overall			
$\bar{\omega}$	0.0299	0.0321	0.0287	0.0276	0.0283	0.0271
$\bar{\alpha}$	0.0719	0.0789	0.0836	0.0833	0.0822	0.0825
$\bar{\beta}$	0.9230	0.9025	0.8849	0.8710	0.8824	0.8752
$\bar{\gamma}$	-	-0.0059	0.0577	0.6025	0.7297	0.1325
$\overline{LOGL}$	-5334.8974	-5360.0646	-5348.0031	-5346.4766	-5342.9512	-5347.5545
% significance	100	36	60	68	67	57

Note: The Augmented GARCH model is given as

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t | \Phi_{t-1} \sim N(0, h_t)$$
$$h_t = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma MT_{t-1}$$

Where MT_{t-1} denotes the market activity measures, namely, volume, no. of trades, no. of up trades, no. of down trades, and no. of same trades.
All the coefficients and stats are reported as the overall average. *LOGL* denotes the log likelihood. Finally, % significance is the percentage of significance when market activity measures are augmented in to GARCH (1, 1) model.

Table 2.6b: GARCH and Augmented GARCH results (by Sector)

	GARCH(1,1)	Augmented with vol_t	Augmented with nt_t	Augmented with ut_t	Augmented with dt_t	Augmented with st_t
			Consumer Discretionary			
$\overline{\omega}$	0.0517	0.0546	0.0522	0.0541	0.0425	0.0509
$\overline{\alpha}$	0.0613	0.0682	0.0748	0.0847	0.0932	0.0602
$\overline{\beta}$	0.9346	0.8891	0.8958	0.9005	0.8454	0.9335
$\overline{\gamma}$	-	-0.0226	0.0259	-0.1671	-0.0732	0.0055
$\overline{LOGL}$	-5659.2815	-5682.5424	-5661.4010	-5700.8682	-5697.7368	-5654.2061
% significance		50	70	60	60	80
			Consumer Staples			
$\overline{\omega}$	0.0594	0.0592	0.0566	0.0589	0.0588	0.0592
$\overline{\alpha}$	0.0653	0.0652	0.0773	0.0651	0.0649	0.0648
$\overline{\beta}$	0.9231	0.9225	0.8691	0.9176	0.9178	0.9193
$\overline{\gamma}$	-	0.0010	-0.0155	0.1092	0.1094	0.0237
$\overline{LOGL}$	-4476.1458	-4473.4490	-4508.0289	-4470.4460	-4470.3890	-4470.7534
% significance		0	80	70	70	70
			Energy			
$\overline{\omega}$	0.0209	0.0214	0.0213	0.0158	0.0226	0.0130
$\overline{\alpha}$	0.0688	0.0696	0.0688	0.0763	0.0704	0.1026
$\overline{\beta}$	0.9228	0.9216	0.9173	0.8601	0.8999	0.7593
$\overline{\gamma}$	-	-0.0014	0.0254	2.0677	0.3812	0.3931
$\overline{LOGL}$	-5635.8358	-5632.3467	-5630.5658	-5635.6407	-5626.4629	-5676.9409
% significance		10	60	80	80	60
			Financials			
$\overline{\omega}$	0.0199	0.0162	0.0195	0.0133	0.0199	0.0203
$\overline{\alpha}$	0.0883	0.0973	0.0900	0.0982	0.0894	0.0889
$\overline{\beta}$	0.9093	0.8613	0.9042	0.8509	0.9038	0.9062
$\overline{\gamma}$	-	0.0206	0.0132	0.4989	0.0807	0.0151
$\overline{LOGL}$	-5416.4291	-5445.8499	-5410.7547	-5432.8209	-5409.6799	-5410.9254
% significance		60	40	60	60	40

Table 2.6b: GARCH and Augmented GARCH results (by Sector) (continued)

	GARCH	Augmented with vol_t	Augmented with nt_t	Augmented with ut_t	Augmented with dt_t	Augmented with st_t
				Health Care		
$\bar{\omega}$	0.0140	0.0140	0.0134	0.0130	0.0130	0.0134
$\bar{\alpha}$	0.0587	0.0671	0.0578	0.0583	0.0583	0.0577
$\bar{\beta}$	0.9306	0.8861	0.9284	0.9267	0.9267	0.9290
$\bar{\gamma}$	-	-0.0129	0.0013	0.0248	0.0255	0.0002
$\overline{LOGL}$	-5254.6396	-5321.1110	-5248.8244	-5248.5231	-5248.4882	-5249.0200
% significance		50	30	40	30	50
				Industrials		
$\bar{\omega}$	0.0399	0.0393	0.0393	0.0355	0.0342	0.0315
$\bar{\alpha}$	0.0837	0.0686	0.0673	0.0939	0.0822	0.0964
$\bar{\beta}$	0.9244	0.9212	0.9197	0.8354	0.8599	0.8016
$\bar{\gamma}$	-	0.0022	0.0227	0.5811	0.4017	0.2696
$\overline{LOGL}$	-5172.1900	-5167.6292	-5166.5138	-5187.5230	-5180.2644	-5193.7272
% significance		50	60	60	60	70
				Information Technology		
$\bar{\omega}$	0.0683	0.0673	0.0711	0.0628	0.0693	0.0658
$\bar{\alpha}$	0.0643	0.0631	0.1324	0.0976	0.0970	0.0942
$\bar{\beta}$	0.9326	0.9329	0.8232	0.8407	0.8692	0.8728
$\bar{\gamma}$	-	0.0002	0.0136	0.4943	0.6765	0.0469
$\overline{LOGL}$	-5671.6072	-5667.3012	-5714.6977	-5706.9399	-5690.8365	-5689.7323
% significance		40	60	60	60	30
				Materials		
$\bar{\omega}$	-0.0371	-0.0373	-0.0370	-0.0366	-0.0367	-0.0372
$\bar{\alpha}$	0.0542	0.0537	0.0538	0.0510	0.0510	0.0538
$\bar{\beta}$	0.9380	0.9374	0.9339	0.9368	0.9367	0.9350
$\bar{\gamma}$	-	0.0018	0.0241	0.1590	0.1606	0.0227
$\overline{LOGL}$	-6025.1570	-6022.2661	-6019.8252	-6018.9239	-6018.8325	-6020.3489
% significance		20	80	80	80	70

Table 2.6b: GARCH and Augmented GARCH results (by Sector) (continued)

	GARCH	Augmented with vol_t	Augmented with nt_t	Augmented with ut_t	Augmented with dt_t	Augmented with st_t
				Telecommunication Services		
$\overline{\omega}$	0.0308	0.0187	0.0237	0.0306	0.0294	0.0238
$\overline{\alpha}$	0.0649	0.0735	0.0933	0.0875	0.1051	0.0786
$\overline{\beta}$	0.9333	0.8957	0.8387	0.8096	0.8022	0.8874
$\overline{\gamma}$	-	-0.0049	0.3600	1.5901	5.1552	0.3915
$\overline{LOGL}$	-5255.9304	-5349.3049	-5333.9312	-5284.4665	-5316.8194	-5318.0448
% significance		40	50	80	80	50
				Utilities		
$\overline{\omega}$	0.0309	0.0675	0.0264	0.0283	0.0302	0.0302
$\overline{\alpha}$	0.1096	0.1628	0.1209	0.1205	0.1107	0.1275
$\overline{\beta}$	0.8814	0.8574	0.8190	0.8319	0.8627	0.8080
$\overline{\gamma}$	-	-0.0425	0.1058	0.6667	0.3796	0.1570
$\overline{LOGL}$	-4781.7575	-4838.8453	-4785.4882	-4778.6137	-4770.0019	-4791.8464
% significance	100	40	70	90	90	50

Note: The Augmented GARCH model is given as

$$r_t = \mu + \varepsilon_t$$
$$\varepsilon_t | \Phi_{t-1} \sim N(0, h_t)$$
$$h_t = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma MT_{t-1}$$

Where MT_{t-1} denotes the market activity measures, namely, volume, no. of trades, no. of up trades, no. of down trades, and no. of same trades.

All the coefficients and stats are reported as the sector average (10 at each sector). *LOGL* denotes the log likelihood.Finally, % significance is the percentage of significance when market activity measures are augmented in to GARCH (1, 1) model.

Table 2.7a: Granger Causality test of OKE (least active)

Causality	Lags	F-Stats	Prob.	Causality	Lags	F-Stats	Prob.	Causality	Lags	F-Stats	Prob.
Volume				*No. of Trades*				*No. of Up Trades*			
$rv_t \rightarrow vol_t$	5	0.381	**(0.862)**	$rv_t \rightarrow nt_t$	15	3.930	(0.000)	$rv_t \rightarrow ut_t$	14	7.556	(0.000)
$vol_t \rightarrow rv_t$		4.085	(0.001)	$nt_t \rightarrow rv_t$		10.440	(0.000)	$ut_t \rightarrow rv_t$		12.786	(0.000)
$rr_t \rightarrow vol_t$	3	1.444	**(0.228)**	$rr_t \rightarrow nt_t$	5	6.359	(0.000)	$rr_t \rightarrow ut_t$	5	13.557	(0.000)
$vol_t \rightarrow rr_t$		8.376	(0.000)	$nt_t \rightarrow rr_t$		18.419	(0.000)	$ut_t \rightarrow rr_t$		21.436	(0.000)
$pv_t \rightarrow vol_t$	5	1.395	**(0.223)**	$pv_t \rightarrow nt_t$	5	2.068	**(0.067)**	$pv_t \rightarrow ut_t$	5	4.228	(0.001)
$vol_t \rightarrow pv_t$		1.049	**(0.387)**	$nt_t \rightarrow pv_t$		16.636	(0.000)	$ut_t \rightarrow pv_t$		19.653	(0.000)
$bv_t \rightarrow vol_t$	5	1.943	**(0.084)**	$bv_t \rightarrow nt_t$	5	3.904	(0.002)	$bv_t \rightarrow ut_t$	5	9.000	(0.000)
$vol_t \rightarrow bv_t$		0.786	**(0.559)**	$nt_t \rightarrow bv_t$		11.237	(0.000)	$ut_t \rightarrow bv_t$		12.273	(0.000)
No. of Down Trades				*No. of Same Trades*							
$rv_t \rightarrow dt_t$	15	8.432	(0.000)	$rv_t \rightarrow st_t$	15	2.028	(0.011)				
$dt_t \rightarrow rv_t$		12.252	(0.000)	$st_t \rightarrow rv_t$		8.170	(0.000)				
$rr_t \rightarrow dt_t$	15	10.344	(0.000)	$rr_t \rightarrow st_t$	5	2.390	(0.036)				
$dt_t \rightarrow rr_t$		7.262	(0.000)	$st_t \rightarrow rr_t$		15.110	(0.000)				
$pv_t \rightarrow dt_t$	15	3.509	(0.000)	$pv_t \rightarrow st_t$	5	0.685	**(0.635)**				
$dt_t \rightarrow pv_t$		8.916	(0.000)	$st_t \rightarrow pv_t$		12.792	(0.000)				
$bv_t \rightarrow dt_t$	15	5.862	(0.000)	$bv_t \rightarrow st_t$	5	1.494	**(0.188)**				
$dt_t \rightarrow bv_t$		5.881	(0.000)	$st_t \rightarrow bv_t$		9.830	(0.000)				

Note: The null hypothesis is A does not Granger cause B. No. of lags is decided according to the Schwarz information Criterion. The numbers in bold are the ones of which p-values are not significant at 5%.

Table 2.7b: Granger Causality test of C (most active)

Causality	Lags	F-Stats	Prob.	Causality	Lags	F-Stats	Prob.	Causality	Lags	F-Stats	Prob.
Volume				*No. of Trades*				*No. of Up Trades*			
$rv_t \rightarrow vol_t$	6	1.554	**(0.157)**	$rv_t \rightarrow nt_t$	5	2.591	(0.024)	$rv_t \rightarrow ut_t$	3	1.749	**(0.155)**
$vol_t \rightarrow rv_t$		5.299	(0.000)	$nt_t \rightarrow rv_t$		32.434	(0.000)	$ut_t \rightarrow rv_t$		18.057	(0.000)
$rr_t \rightarrow vol_t$	6	1.063	**(0.383)**	$rr_t \rightarrow nt_t$	5	1.965	**(0.081)**	$rr_t \rightarrow ut_t$	5	0.876	**(0.496)**
$vol_t \rightarrow rr_t$		5.542	(0.000)	$nt_t \rightarrow rr_t$		33.596	(0.000)	$ut_t \rightarrow rr_t$		11.185	(0.000)
$pv_t \rightarrow vol_t$	6	1.163	**(0.323)**	$pv_t \rightarrow nt_t$	6	1.167	**(0.321)**	$pv_t \rightarrow ut_t$	3	0.900	**(0.440)**
$vol_t \rightarrow pv_t$		6.325	(0.000)	$nt_t \rightarrow pv_t$		22.646	(0.000)	$ut_t \rightarrow pv_t$		15.100	(0.000)
$bv_t \rightarrow vol_t$	6	0.793	**(0.576)**	$bv_t \rightarrow nt_t$	6	1.781	**(0.099)**	$bv_t \rightarrow ut_t$	6	0.475	**(0.827)**
$vol_t \rightarrow bv_t$		6.291	(0.000)	$nt_t \rightarrow bv_t$		33.317	(0.000)	$ut_t \rightarrow bv_t$		11.371	(0.000)
No. of Down Trades				*No. of Same Trades*							
$rv_t \rightarrow dt_t$	3	1.761	**(0.152)**	$rv_t \rightarrow st_t$	5	3.950	(0.001)				
$dt_t \rightarrow rv_t$		18.136	(0.000)	$st_t \rightarrow rv_t$		38.755	(0.000)				
$rr_t \rightarrow dt_t$	5	2.856	(0.014)	$rr_t \rightarrow st_t$	5	2.856	(0.014)				
$dt_t \rightarrow rr_t$		40.089	(0.000)	$st_t \rightarrow rr_t$		40.089	(0.000)				
$pv_t \rightarrow dt_t$	3	0.875	**(0.453)**	$pv_t \rightarrow st_t$	13	1.519	**(0.103)**				
$dt_t \rightarrow pv_t$		15.127	(0.000)	$st_t \rightarrow pv_t$		16.424	(0.000)				
$bv_t \rightarrow dt_t$	6	2.950	(0.007)	$bv_t \rightarrow st_t$	6	2.950	(0.007)				
$dt_t \rightarrow bv_t$		39.835	(0.000)	$st_t \rightarrow bv_t$		39.835	(0.000)				

Note: The null hypothesis is A does not Granger cause B. No. of lags is decided according to the Schwarz information Criterion. The numbers in bold are the ones of which p-values are not significant at 5%.

Table 2.7c: Granger Causality test of overall average market activity measures and realized volatility measures

Causality	Lags	F-Statistic	Prob.	Causality	Lags	F-Statistic	Prob.	Causality	Lags	F-Statistic	Prob.
Volume				*No. of Trades*				*No. of Up Trades*			
$rv_t \rightarrow vol_t$	12	3.927	(0.000)	$rv_t \rightarrow nt_t$	10	13.553	(0.000)	$rv_t \rightarrow ut_t$	12	18.084	(0.000)
$vol_t \rightarrow rv_t$		7.057	(0.000)	$nt_t \rightarrow rv_t$		13.164	(0.000)	$ut_t \rightarrow rv_t$		11.319	(0.000)
$rr_t \rightarrow vol_t$	10	3.467	(0.000)	$rr_t \rightarrow nt_t$	10	11.262	(0.000)	$rr_t \rightarrow ut_t$	16	12.133	(0.000)
$vol_t \rightarrow rr_t$		8.174	(0.000)	$nt_t \rightarrow rr_t$		21.018	(0.000)	$ut_t \rightarrow rr_t$		17.213	(0.000)
$pv_t \rightarrow vol_t$	10	3.892	(0.000)	$pv_t \rightarrow nt_t$	10	7.864	(0.000)	$pv_t \rightarrow ut_t$	10	8.999	(0.000)
$vol_t \rightarrow pv_t$		6.168	(0.000)	$nt_t \rightarrow pv_t$		10.429	(0.000)	$ut_t \rightarrow pv_t$		7.917	(0.000)
$bv_t \rightarrow vol_t$	10	4.351	(0.000)	$bv_t \rightarrow nt_t$	10	12.662	(0.000)	$bv_t \rightarrow ut_t$	10	17.487	(0.000)
$vol_t \rightarrow bv_t$		6.532	(0.000)	$nt_t \rightarrow bv_t$		15.661	(0.000)	$ut_t \rightarrow bv_t$		13.939	(0.000)
No. of Down Trades				*No. of Same Trades*							
$rv_t \rightarrow dt_t$	12	18.411	(0.000)	$rv_t \rightarrow st_t$	12	10.358	(0.000)				
$dt_t \rightarrow rv_t$		12.021	(0.000)	$st_t \rightarrow rv_t$		12.875	(0.000)				
$rr_t \rightarrow dt_t$	16	12.223	(0.000)	$rr_t \rightarrow st_t$	15	6.745	(0.000)				
$dt_t \rightarrow rr_t$		17.623	(0.000)	$st_t \rightarrow rr_t$		13.831	(0.000)				
$pv_t \rightarrow dt_t$	10	9.538	(0.000)	$pv_t \rightarrow st_t$	10	5.871	(0.000)				
$dt_t \rightarrow pv_t$		8.778	(0.000)	$st_t \rightarrow pv_t$		10.331	(0.000)				
$bv_t \rightarrow dt_t$	10	17.972	(0.000)	$bv_t \rightarrow st_t$	10	9.439	(0.000)				
$dt_t \rightarrow bv_t$		14.819	(0.000)	$st_t \rightarrow bv_t$		14.908	(0.000)				

Note: The null hypothesis is A does not Granger cause B. No. of lags is decided according to the Schwarz information Criterion. The numbers in bold are the ones of which p-values are not significant at 5%.

Table 2.8: GMM Results

	m2	*m3*	*m4*	*J Test*	*m2*	*m3*	*m4*	*J Test*	*m2*	*m3*	*m4*	*J Test*	*m2*	*m3*	*m4*	*J Test*
	CD				**CS**				**ENG**				**FIN**			
Bivariate Moments with vol_t	9	9	6	3	9	9	8	1	10	9	7	2	9	8	8	5
Bivariate Moments with nt_t	8	8	4	3	10	10	7	3	10	9	5	5	10	9	8	4
Bivariate Moments with ut_t	10	10	3	5	10	10	6	4	10	10	7	4	10	8	8	6
Bivariate Moments with dt_t	10	10	4	3	10	10	6	5	10	9	7	4	8	8	7	4
Bivariate Moments with st_t	9	9	5	3	10	10	8	3	10	10	7	6	10	9	8	6
	HC				**IND**				**IT**				**MAR**			
Bivariate Moments with vol_t	8	6	5	3	10	9	7	0	10	9	6	0	10	10	6	1
Bivariate Moments with nt_t	10	10	7	1	10	8	6	3	9	9	9	0	8	7	4	3
Bivariate Moments with ut_t	10	10	6	2	10	10	5	6	10	9	7	2	10	10	7	3
Bivariate Moments with dt_t	9	10	7	2	10	10	5	5	10	10	7	3	10	10	9	4
Bivariate Moments with st_t	10	9	6	2	9	9	6	7	10	10	7	0	10	9	5	2
	TEL				**UTL**				**OVERALL**							
Bivariate Moments with vol_t	10	8	4	2	9	7	4	2	94	84	61	19				
Bivariate Moments with nt_t	10	8	6	3	10	9	5	3	95	87	61	28				
Bivariate Moments with ut_t	9	8	6	0	10	10	3	3	99	95	58	35				
Bivariate Moments with dt_t	10	9	9	0	10	10	6	2	97	96	67	32				
Bivariate Moments with st_t	9	9	8	2	10	9	9	1	97	93	69	32				

Note: This tables reports the number of stocks that the recovery of the estimated information flow moments and the over-identifying J test are not rejected at the 5% level.

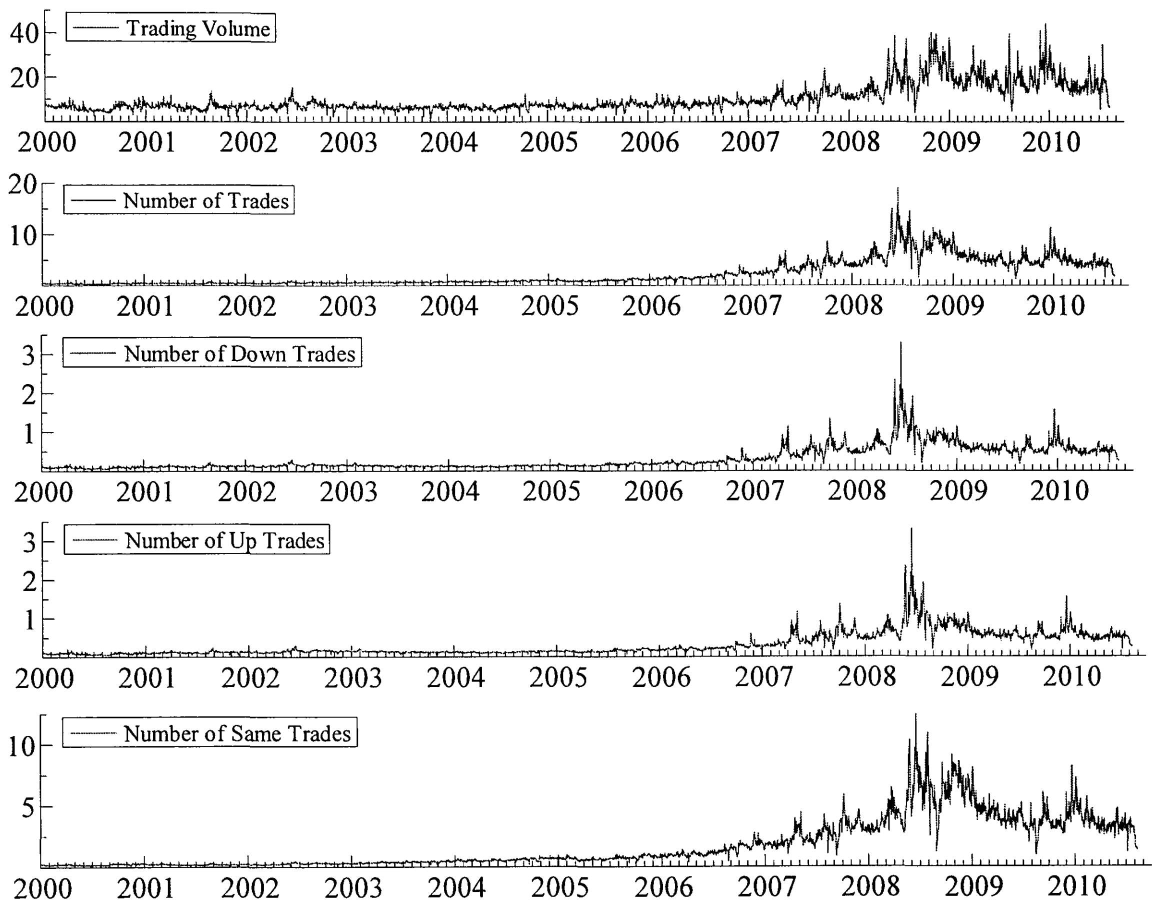
Table 2.9: Standardization results of sector average returns on market activity measures

	r_t	$r_t/\sqrt{vol_t}$	$r_t/\sqrt{nt_t}$	$r_t/\sqrt{ut_t}$	$r_t/\sqrt{dt_t}$	$r_t/\sqrt{st_t}$	r_t	$r_t/\sqrt{vol_t}$	$r_t/\sqrt{nt_t}$	$r_t/\sqrt{ut_t}$	$r_t/\sqrt{dt_t}$	$r_t/\sqrt{st_t}$
	CD						**CS**					
Skewness	0.445	0.219	0.554	0.418	0.483	0.622	0.317	0.234	0.670	0.627	0.706	0.655
Kurtosis	9.455	6.455	12.592	10.359	10.579	13.503	10.551	5.895	11.642	10.931	11.133	12.107
Jarque-Bera	4895.004	1398.580	10749.640	6323.878	6729.497	12897.440	6620.154	991.804	8816.615	7433.480	7856.215	9759.462
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	ENG						**FIN**					
Skewness	-0.404	-0.167	-0.098	-0.117	0.018	-0.135	-0.240	-0.021	0.105	0.070	0.157	0.091
Kurtosis	11.575	4.536	9.311	9.235	9.211	10.003	14.159	5.055	9.311	8.799	8.779	9.480
Jarque-Bera	8553.035	284.912	4597.029	4488.294	4447.414	5662.960	14382.610	487.158	4597.525	3879.747	3861.926	4844.706
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	HC						**IND**					
Skewness	0.047	0.053	0.120	0.053	0.113	0.168	-0.198	-0.016	0.170	0.133	0.206	0.166
Kurtosis	9.164	5.437	8.987	8.284	8.294	9.248	7.090	4.234	7.742	7.238	7.269	8.108
Jarque-Bera	4381.467	686.122	4139.674	3220.375	3237.143	4513.316	1946.997	175.692	2605.981	2078.401	2121.178	3020.663
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	IT						**MAR**					
Skewness	0.277	0.061	0.105	0.010	0.023	0.182	-0.534	0.047	0.100	-0.173	0.308	0.110
Kurtosis	7.656	5.187	7.522	5.308	5.311	8.496	8.621	6.205	11.905	13.905	14.277	11.134
Jarque-Bera	2535.008	552.995	2362.796	614.312	615.902	3498.387	3774.455	1185.332	9146.702	13723.540	14706.270	7632.859
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	TEL						**UTL**					
Skewness	-0.096	0.135	0.322	0.272	0.375	0.303	0.033	0.029	-0.062	-0.249	0.077	-0.019
Kurtosis	8.674	6.254	9.361	8.217	8.333	10.022	10.989	6.226	11.584	11.806	11.456	12.143
Jarque-Bera	3716.473	1229.379	4713.158	3171.796	3344.000	5727.778	7359.564	1200.139	8497.934	8967.929	8246.561	9637.859
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis as well as the Jarque-Bera statistics for normality of sector average returns standardized by different market activity measures.

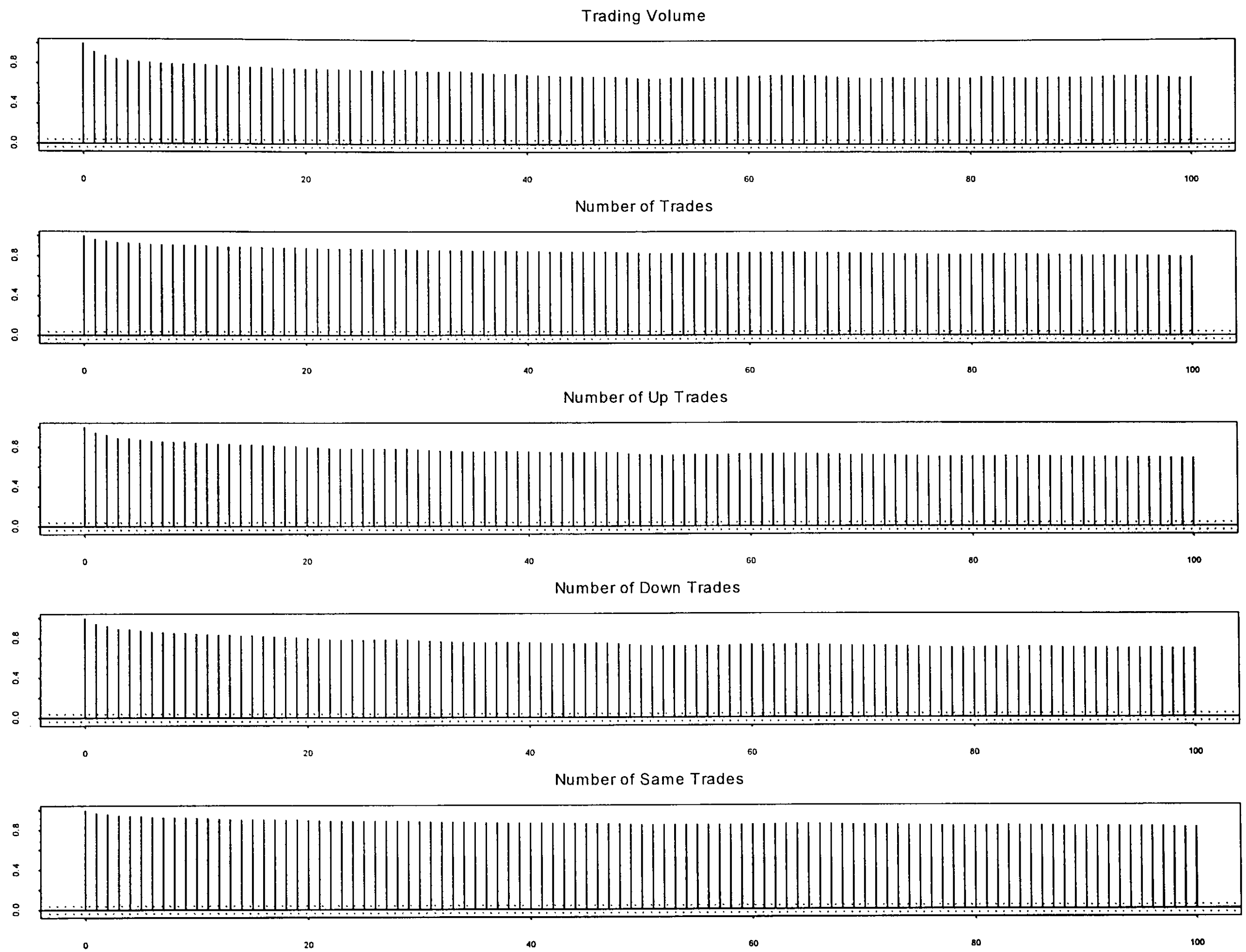
Figures

Figure 2.1: Market activity measures plots



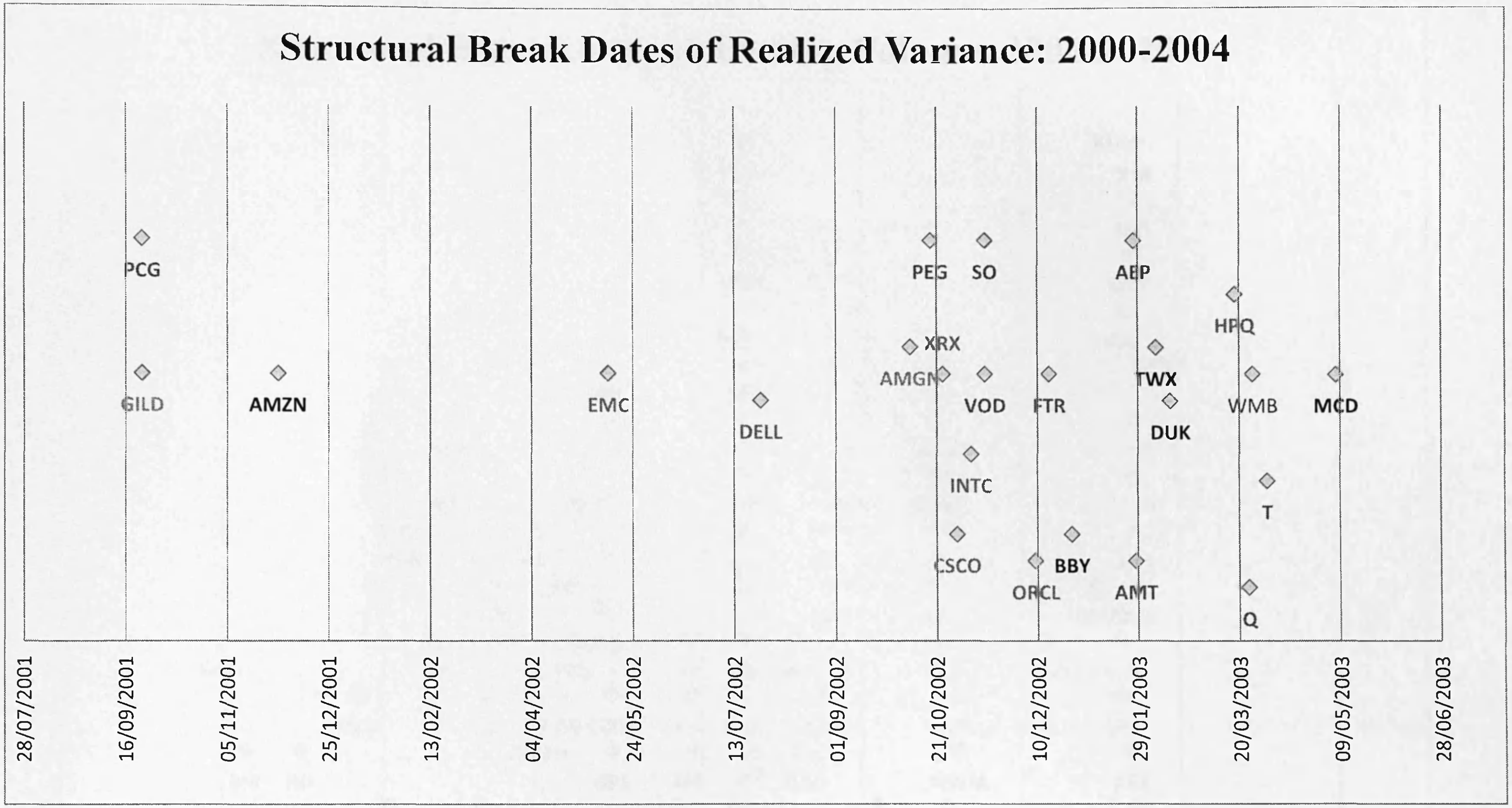
Note: The figures present the overall sample average market activity measures. Volume is rescaled by dividing by 1,000,000 and trade measures are rescaled by dividing by 10,000.

Figure 2.2: ACFs of Market Activity Measures



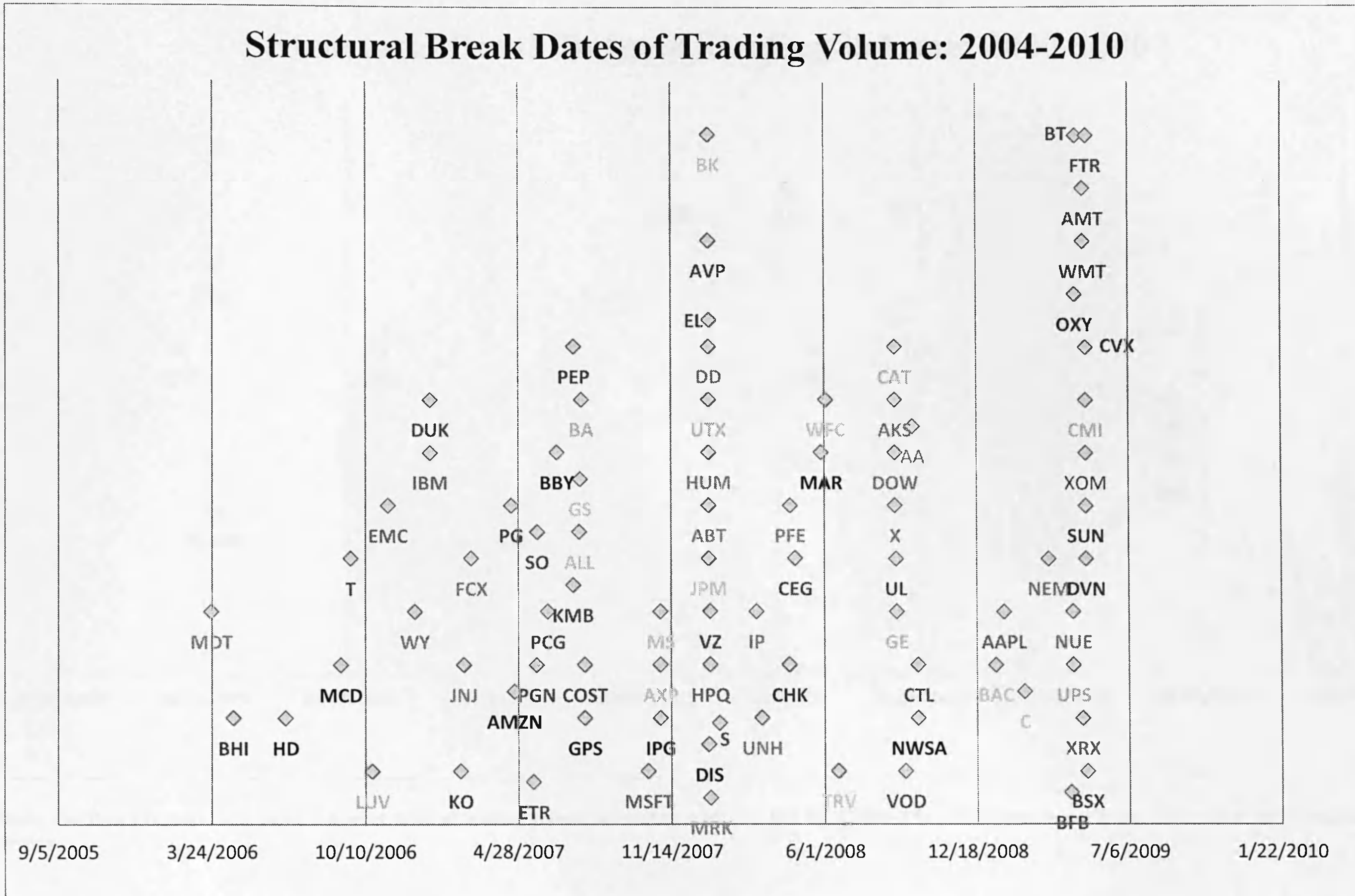
Note: ACFs up to 100 lags are plotted for overall average market activity measures.

Figure 2.3a: Structural break dates of realized variance: 2000-2004



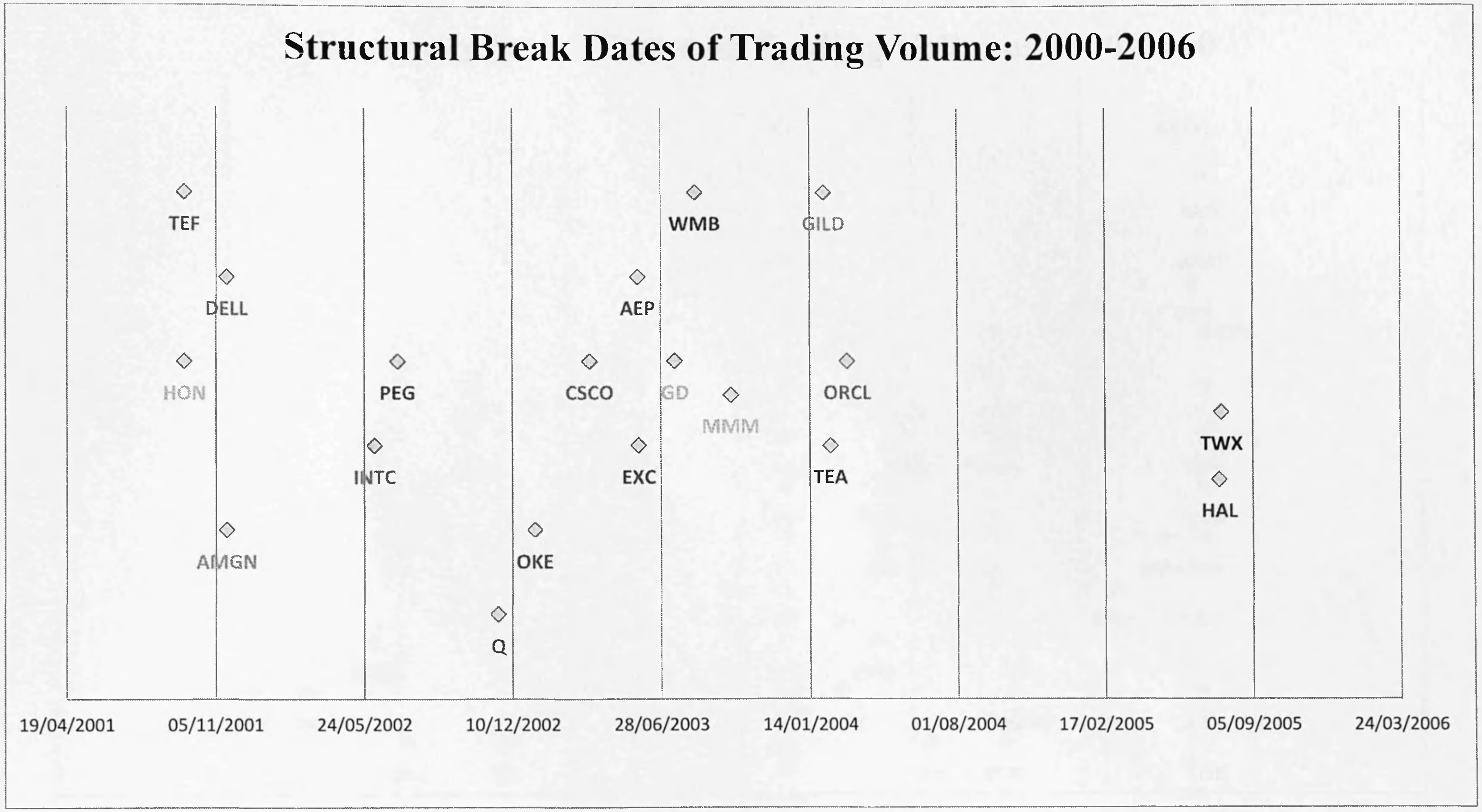
Note: The figure reports the structural breaks dates of realized variance according to the ZA test from 2000-2004. The stocks shown in the same colour are categorized in the same sectors.

Figure 2.3b: Structural break dates of realized variance: 2004-2010



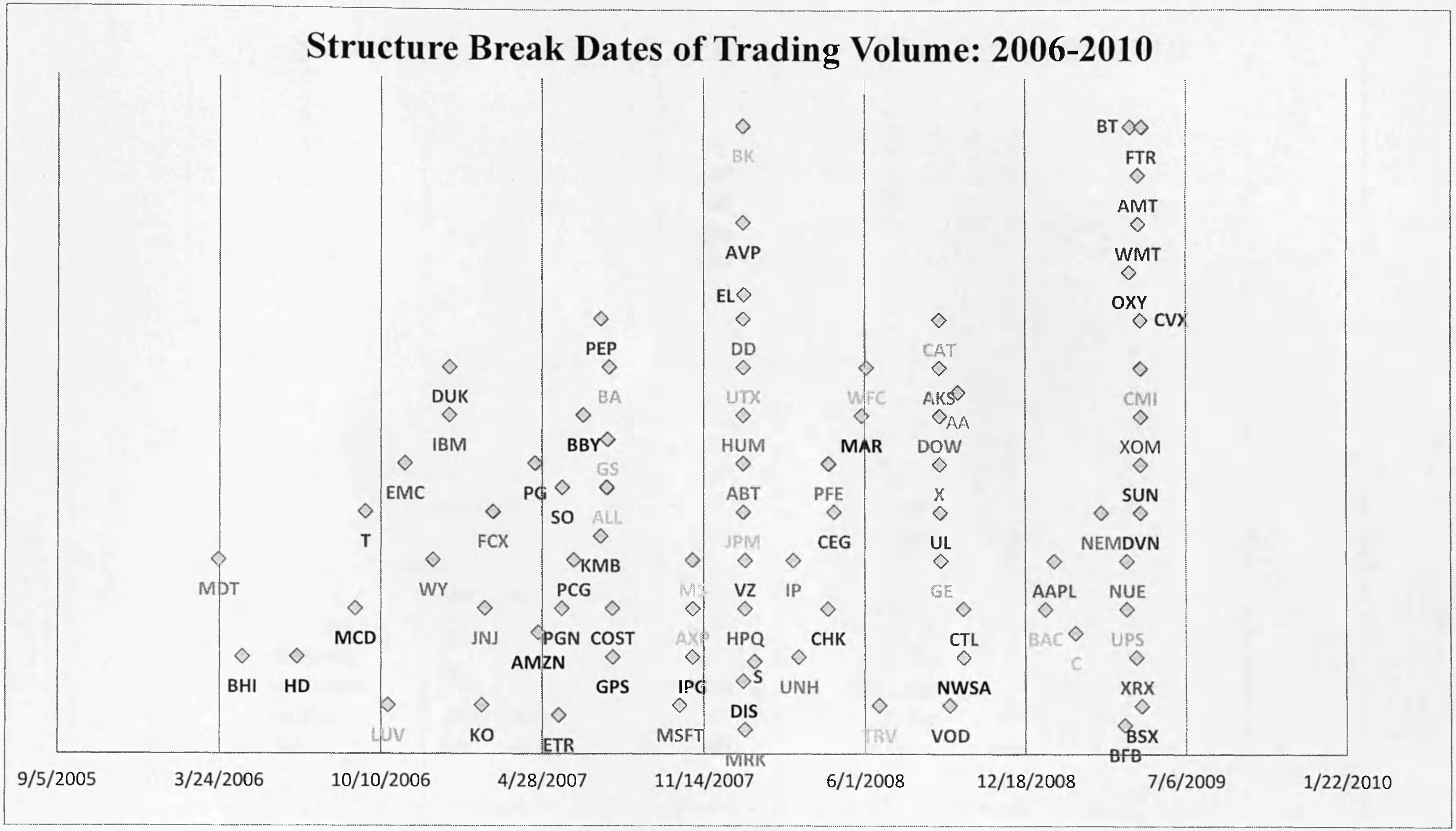
Note: The figure reports the structural breaks dates of realized variance according to the ZA test from 2004-2010. The stocks shown in the same colour are categorized in the same sectors.

Figure 2.3C. Structural break dates of trading volume: 2000-2006



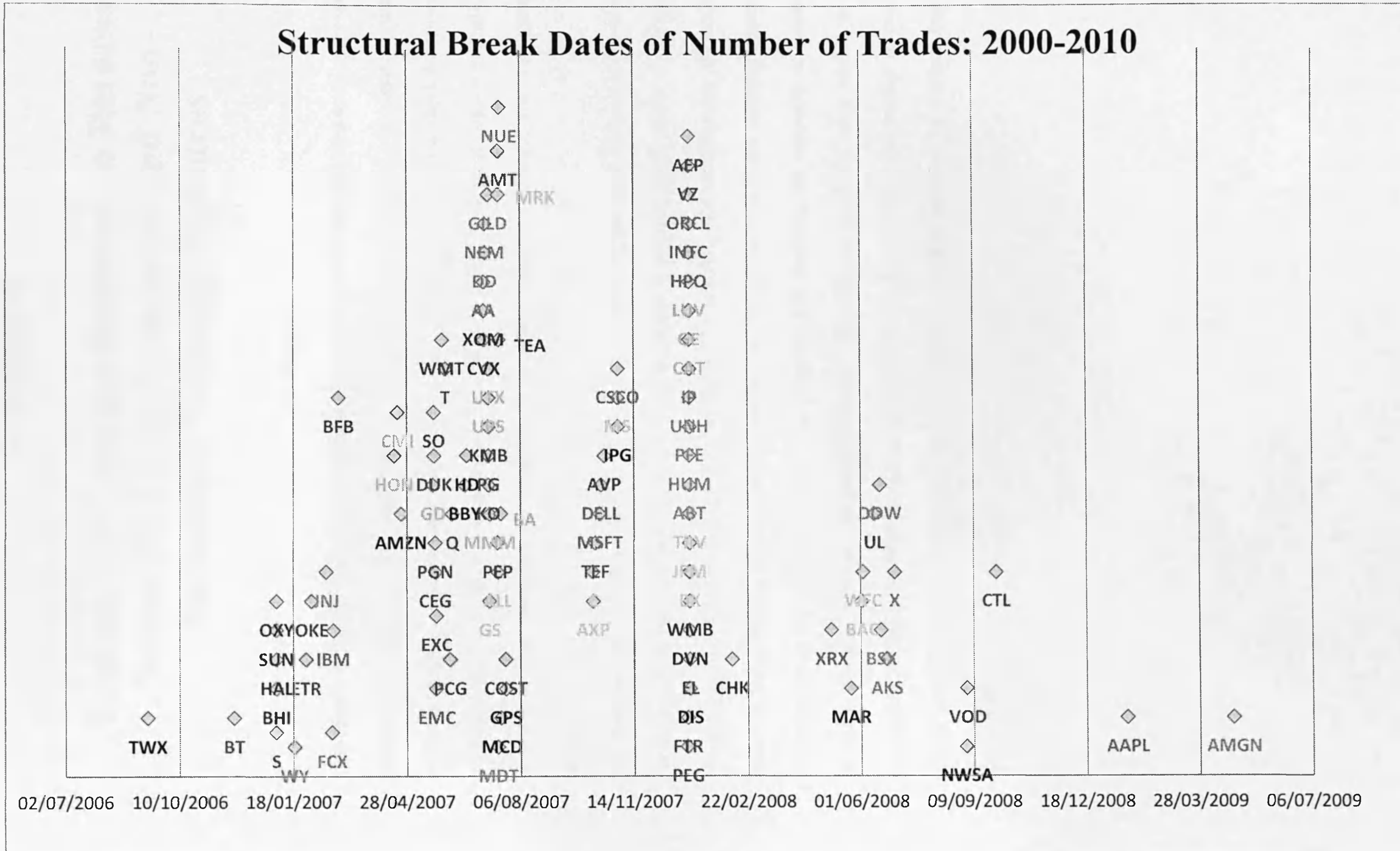
Note: The figure reports the structural breaks dates of trading volume according to the ZA test from 2000-2006. The stocks shown in the same colour are categorized in the same sectors.

Figure 2.3d: Structure break dates of trading volume: 2006-2010



Note: The figure reports the structural breaks dates of trading volume according to the ZA test from 2006-2010. The stocks shown in the same colour are categorized in the same sectors.

Figure 2.3e: Structural break dates of number of trades: 2000-2010.



Note: The figure reports the structural breaks dates of number of trades according to the ZA test from 2000-2010. The stocks shown in the same colour are categorized in the same sectors.

Chapter 3

Factors Affecting the Recovery to Returns Normality Using Parametric and Non-Parametric Volatility Measures

Abstract

This chapter investigates the issue of achieving returns normality using high frequency non-parametric measures (realized variance, range, and power and bipower variations) and classical parametric measures (stochastic volatility and GARCH). Our task is twofold. The first is to check the relevance of stock type and the level of activity on the power of the volatility measure to achieve normality. The second task is to investigate the impact of: a) sampling frequency; b) jumps; and c) microstructure noise.

Our findings can be summarized as follows: 1) the nature of the stock is relevant to recovering normality – It is more difficult to achieve return normality in actively traded stocks; 2) the sampling frequency affects the recovery of returns normality. For example, the rejections rate of standardized returns is higher at the 1-second sampling frequency as opposed to the 5-minute sampling frequency; 3) excluding jumps from realized variance has little effect in distorting the normality of standardized returns at least at 5-minute sampling frequency. The effect of jumps is more visible at higher frequencies; 4) applying filter to counter microstructure noise enhances the process of recovering returns normality; 5) the performance of non-parametric volatility measures far exceeds that of the classical measures.

3.1 Introduction

The assumption that asset returns are normally distributed is fundamental in many asset pricing models, such as the Black and Scholes (1973) financial derivatives pricing model, the VaR evaluation in the portfolio management, and so on. It also has important implications for financial hedging and risk management. However, asset returns are rarely observed to be normally distributed. This well-established fact is extensively reported in a large number of empirical studies. In the distant past we have the papers of Mandelbrot (1963, 67), Fama (1965), and Pratez (1972). Mandelbrot (1963) attributed the non-normality of stock returns to the fact that the variance of returns, which is the sum of elementary logarithmic prices changes, is infinite, which in turn implies that the Central Limit Theorem is violated. Another explanation of the non-normality of returns is provided by the seminal paper of Clark (1973) which attempts to explain the non-normality without sacrificing the assumption of finite variance. He argues that the transactions are not evenly spreader across the trading hours and therefore the distribution of returns can be viewed as a mixture distribution through the subordination process. This mixture distributions hypothesis has attracted the most attention and paved the way for a number of studies that follow. The univariate mixture distribution model is first proposed by Clark (1973), and then is modified as a bivariate model by Tauchen and Pitts (1983), Andersen (1996) and Liesenfeld (2001). According to Mixture of Distributions Hypothesis, although the returns are not Gaussian, returns subordinated with trading volume could be Gaussian as both financial prices and trading volume are driven by the same latent information flow arrivals. Furthermore, Monroe (1978) asserts that any semi-martingale can be written as a time changed Brownian motion. The Monroe result in essence indicates that, as long as we are willing to change the time of the process, there will exist a filtration that to which the return process can be adapted and be brought back to normal distribution.

Based on MDH, Ané and Geman (2000) revisit the Clark (1973) paper by replacing the trading volume by number of the trades as the subordinator. Under the non-arbitrage assumptions and the conclusion reached by Monroe, Ané and Geman (2000) find that returns adapted to the time of number of trades are normally distributed.

Nevertheless, recent empirical studies question the results of Clark (1973) or Ané and Geman (2000) with respect to recovering return normality. The failure to recovering return normality by volume/number of trades could be attributed to the irregular arrival of information flows. Prices evolve at different rates during a given time period (day, hour, 5 minutes, and so on) because the arrival of information is assumed to be random.

With the availability of high frequency data and the development of continuous-time models, the non-normality issue has been addressed to better effect by studies which primarily use high frequency data. Many find that unconditional distributions of raw daily returns have fat tails yet the distributions appear close to Gaussian when the returns are standardized by the corresponding realized volatility measures (See for instance ABDL (2001) ABDE(2001) Areal (2001), BN-S(2004), ABFN(2010) Fleming and Paye(2007, 2011). ABFN (2010) claims that "*the (true) realized volatility standardized returns should be indistinguishable from a Gaussian if the true price process belongs to a certain class of pure diffusive processes and market microstructure frictions are negligible*". Nonetheless, there are situations that easily invalidate the above statement and hence make the realized volatility measures unable to restore returns normality. The failure to recover normality by realized volatility measures may be attributed to market microstructure noise, the presence of jumps or also known as the discontinuity in the price path, the leverage effect, the correlation between price and volatility innovations which might induce the asymmetry in the standardized return distribution, and the sampling frequency at which the realized volatility is calculated.

Given the many reasons different that might explain the non-normality of standardized returns, empirical studies reach a variety of conclusions. For instance, Fleming and Paye (2011) argue that the presence of jumps affect the normality of standardized returns, whereas ABFN (2010) suggest that jumps plays little part in the standardization process.

In this chapter, we discuss the distributional properties of daily returns under the framework of continuous-time price models. We specifically follow ABFN (2010) and Fleming and Paye (2011) and extend the number of realized volatility measures used. In order to compare the power of different realized volatility measures to recover normality, we apply a series of tests and transform the daily return series to account for jumps and the effects of microstructure noise: We construct the volatility series at the optimal sampling frequency according to the volatility signature plots which help show the microstructure noise and the price jumps. We separate the continuous variance path and jumps path in the realized variance, relying on the realized bipower variation proposed by BN-S (2004) to detect the significant daily jumps based on BN-S (2006) and Huang and Tauchen (2005). We use the exponentially weighted moving average as a filtering process for daily returns and realized volatility measures to smooth the market microstructure noise.

The main findings of this chapter are summarized as follows:

1. Volatility measures estimated using non-parametric methods are superior to those using parametric methods in capturing the dynamics of the return process and hence produce more normal returns. Realized power variation performs the best in the realized volatility measures.
2. Jumps should be included in the price process yet it has little effect on the normality of returns. Microstructure noise is by far the dominant factor.
3. The level of trading actively has an impact upon the distributional properties of daily returns. We find that the sectors that are most difficult to standardize are those sectors which contain high trading volume stocks.
4. Appling an exponentially weighted moving average filter can, in some instances, enhance the power of (certain) realized volatility measures to restore normality. This moving average filter can be applied to all the realized volatility measures discussed in this chapter. It is also easy to model. The exponentially weighted moving average filter is more successful with the highly actively traded stocks/sectors.

In section 2, we review the literature. In section 3, we outline the theoretical framework for Gaussianity of the standardized returns distributions. Section 4 provides a brief discussion of the data and some preliminary descriptive statistics. The outcomes of distributional tests are summarized in Section 5. Section 6, finally, presents our conclusions.

3.2 Literature Review

Some early attempts in using realized volatility measures to standardize returns can be traced to ABDL (2000) and ABDE (2001). Using foreign exchange data and equity stock data respectively, both papers find that returns standardized by realized volatility (calculated as the squared root of realized variance) are (near) Gaussian. In contrast, return standardized by GARCH, ARCH and SV are far from normal distribution although Gaussianity of the parametric volatility standardized returns are better than the raw returns.

Areal and Taylor (2001) reach similar conclusion in using 8-year FTSE-100 futures prices data. The paper assigns both equal weights and optimal weights to realized volatility and finds that returns standardized by optimal weighted realized volatility are closer to the normal distribution. However, the normality of standardized returns remains rejected at the 5% level.

ABD (2007b) test the returns distribution using both simulated and 17-year 2-minute S&P 500 futures data. The simulation procedure considers following assumptions of high frequency stochastic volatility: no-leverage pure diffusion, no-leverage jump diffusion, leverage pure diffusion and leverage jump diffusion. Moreover, the effect of microstructure noise is also considered in the simulation. Then real empirical data is tested. Contrasting with conclusions of ABDL (2000), that prices follow a pure diffusion process, the results from both simulated and real data show that the price follows the jump-diffusive representation. It is also revealed that microstructure noise may play a critical part in determining the distributional properties of intraday aggregated returns. Accounting for both the leverage effect and daily and intraday jumps in the volatility is crucial in determining the return distributions.

Fleming and Paye (2007) investigate the impact of microstructure noise on the distributions of returns standardized by realized volatility. Using 10-year data of 20 stocks traded in the Major Market Index, the authors conclude that microstructure noise leads to an upward bias of realized volatility, hence making rv_t standardized returns artificially normal distributed. However, as long as the bias is corrected, rv_t standardized returns are substantially departure for normal distribution and appear to be platykurtotic. Another important finding is that the choice of sampling frequency has an important effect upon the standardization result. The volatility signature plots show that the standard deviation of rv_t decreases with sampling frequency and the kurtosis exhibits an increasing trend when the sampling frequency is higher.

After the ABD (2007b) paper, which discusses the impact of jumps in the rv_t on the distributional properties of standardized returns, there were similar publications. ABFN (2010) use 30 DJIA stocks to revisit the impact of jumps, the leverage effect and market microstructure noise. Individual stocks have higher volatility, contain more jumps and are more greatly influenced by the microstructure noise than index futures data. This has enabled researchers to obtain new evidence on the validity of previous studies. To examine the role of jumps, ABFN (2010) use two different daily jump detection methods: The first is the widely used single daily jumps detection method first proposed by Huang and Tauchen (2005). This heavily relies on another realized volatility measure: realized bipower variation (BN-S, 2004, 2006). The second jump detection method allows for the presence of multiple jumps within one trading day. ABFN (2000) also suggest 5-minute as an optimal sampling frequency. The findings support the ABD (2007) paper and point to the conclusion that prices may be described by a jump-diffusion process, but after allowing for leverage and/or feedback effects. The presence of jumps in the prices has little impact on the distributional properties of standardized returns.

Khalifa *et al* (2011) follow the methodology proposed by ABFN (2010), but further consider absolute returns, realized bipower variation and integrated volatility via Fourier transformation (IVFT) for high frequency US gold, silver and copper futures data between 1999 and 2008. Their findings include: the normality of metal futures returns is more difficult to achieve. None of the three series recover the normality of returns at the 1% level, no matter which volatility measures are used. Realized bipower variation performs the best among all the measures. The presence of microstructure noise might be the explanation. However, when jumps and the leverage effect are accounted for in the manner suggested by AFBN (2010), the normality of demeaned financial-time returns of three metal markets cannot be rejected at the 1% level.

Another work which also looks at the impact of jumps on distributional properties of standardized returns is that of Fleming and Paye (2011). Using 20 stocks in MMI, the authors construct realized variance and realized bipower variation and compare the properties of returns standardized by these two measures. Their findings favour bv_t standardized returns, which appear to be closer to normal distribution. According to BN-S (2004, 2006), realized bipower variation is jump robust. When jumps are excluded, the Gaussianity of standardized returns is greatly improved. This paper uses 3-minute aggregated bipower variation as their volatility signature plots suggest. On the other hand, the better performance of realized bipower variation in the standardization procedure, shows realized volatility as a noisy estimator of the quadratic variation.

Chevallier and Sevi (2011) use EXC CO2 emission 2008 futures data to discuss the distributional properties of returns and standardized returns. As distinct from the extensive studied financial data, the environmental economic data show an optimal sampling frequency of 15-minute. Realized volatility is calculated and then used to standardize daily returns. The paper compares returns standardized by rv_t and by GARCH (1, 1). The latter provides a closer fit to the normal distribution.

Apart from realized variance and realized bipower variation, the distributional properties of returns standardized by other realized volatility measures have also been studied. Martens and van Dijk (2007) look at the unconditional distributions of daily returns standardized by realized range using S&P 500 index futures data from 1999-2004. The results show that the normality of returns standardized by the (squared root of the rescaled) realized range cannot be rejected at the 5% level. However, returns standardized by realized volatility, which are constructed from the same data set, lead to a rejection of the null that rv_t standardized returns are normally distributed.

Fuertes *et al* (2009) investigate standardized returns distributions by considering four realized volatility measures: realized variance, realized range and realized power and bipower variation and one parametric volatility, GARCH (1, 1). They use 14 actively traded US equity data from 1997 to 2003. Realized range is the most successful volatility measure and brings 13 out of 14 returns back to normal at the 5% level. Realized power variation is the second best, followed by realized power variation and realized variance.

From the above literature, it is evident that the power of realized volatility measures to achieve the Gaussianity of daily returns is generally stronger than the traditional GARCH/SV measures. However, conclusions differ as different datasets are considered. Findings have not converged to any conclusion as to which realized volatility measure(s) under what condition(s) are most successful in recovering returns normality under the standardization procedure. In the following sections, we shall discuss this issue by considering different volatility measures (parametric and nonparametric) and different market conditions.

3.3 Theoretical Framework

In line with the literature, we consider a jump-diffusion process which views asset prices as containing a continuous sample path and a non-continuous part (jump part). Recall [1.1], assume price process P_t, which is semi-martingale, follows a geometric Brownian Motion,

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + k(t)dq(t) \qquad 0 \le t \le T$$

where $\mu(t)$ denotes a continuous and locally bounded process, $\sigma(t)$ is the constant volatility parameters, $W(t)$ denotes a standard Brownian Motion. Both $\mu(t)$ and $\sigma(t)$ are jointly independent of the Brownian motion. $q(t)$ is the counting process and $k(t)$ the size of the corresponding jumps. When $dq(t) = 1$ a jump is present at time t and $dq(t) = 0$ otherwise.

In reality, the empirical tests are conducted in discrete time. Implied by the jump diffusion process from [1.1], the one-period continuously compounded discrete time asset returns is calculated as

$$r_t = p_t - p_{t-1}, \quad t= 1, 2\ldots \qquad [3.1]$$

The 'one-period' equals one day. The distribution of daily return r_t depends on the continuous-time model. Here we consider three general model conditions which show how standardized returns should be standard normally distributed under each.

3.3.1 Pure Diffusion Case

This is the simplest continuous-time model assume there is no jumps, no leverage and feedback effects in the prices. $q(t) \equiv 0$, $\sigma(t)$ and $W(u)$ re independent of each other for all $t \geq 0$ and $u \geq 0$ .Therefore,

$$r_t / (\int_{t-1}^{t} \sigma^2(u) du)^{-1/2} \sim N(0,1), \quad t=1, 2, 3... \qquad [3.2]$$

where $\int_{t-1}^{t} \sigma^2(u) du$ is the integrated variance (IV).

The distributional properties under [3.1] are conditioned on *ex post* sample path realization of $\sigma(t)$ over the corresponding discrete time return interval, $(t-1, t]$. The integrated variance is latent, yet various estimators of IV has been discussed and studied extensively, starting with the realized variance by ABDL (2001).

3.3.2 Jumps

The presence of jumps in the asset data has been extensively reported. Allowing for jumps when modelling asset prices returns has been explicitly argued. See for example, Eraker *et al* (2003) and ABD (2003, 2007a, b). [3.2] is no longer valid when an additional component is added to the *ex post* price. If the size of jumps is assumed known, and then the corresponding jump-adjusted returns can be written as

$$\tilde{r}_t = p_t - p_{t-1} - \sum_{s=q_{t-1}}^{q_t} k(s), \qquad t=1, 2, 3.... \qquad [3.3]$$

As jumps are assumed to be independent of the Brownian Motion $W(t)$, when they are excluded from the return series, the adjusted returns link only with the diffusion component and hence should again be normally distributed after standardization:

$$\tilde{r}_t / (\int_{t-1}^{t} \sigma^2(u) du)^{-1/2} \sim N(0,1), \quad t=1, 2, 3... \qquad [3.4]$$

It is important to separate the continuous sample path and jumps by adopting a jump detection method. Here we use the non-parametric jump detection method by Bollerslev *et al* (2009) and allow for the presence of jumps of different magnitudes. The detailed discussion of the jump-detection method is provided in Chapter 1, Section 1.5.2.

3.3.3 Market Microstructure Noise

Another force that drives the returns from normal distribution is the presence of market microstructure noise. With high frequency data analysis, noise is more prevalent in the tick-by-tick data. Microstructure noise has been attributed to such as the bid-ask bounce, latency, and information asymmetry. Eliminating microstructure noise is essential to remove bias in the estimation. A large number of studies have addressed this problem, some of which focus on removing or filtering microstructure noise. See for example, Hansen and Lunde (2005), Bandi and Russell (2006, 2008). Specifically, Hansen and Lunde (2005) assume that the observed price (p_t) is the sum of true price (p_t^*) and noise (ω_t),

$$p_t^* = p_t - \omega_t \qquad [3.5]$$

From which noise-adjusted returns is denoted as

$$r_t^* = p_t^* - p_{t-1}^*, \qquad t = 1, 2, 3 \ldots \qquad [3.6]$$

Assume that market microstructure noise is independent. As for the case of jumps, noise-adjusted returns standardized by the IV should also result in Gaussianity.

$$r_t^* / (\int_{t-1}^{t} \sigma^2(u)du)^{-1/2} \sim N(0,1), \quad t=1, 2, 3\ldots \qquad [3.7]$$

However, the independent noise assumption may not always hold in reality especially when sampling data at the highest frequency. If we assume the condition of Section 3.1, such that prices follow the Brownian semi-martingale pure diffusion process without leverage effects, the sampling frequency m converges to infinity, Hansen and Lunde (2006) claim that

$$RV_t - IV_t \xrightarrow{p} \omega_t \qquad [3.8]$$

and that the standardized returns follow a distribution of

$$r_t / (RV_t)^{-1/2} \sim N(0, \frac{IV_t}{IV_t + \omega_t}) \qquad [3.9]$$

This distribution has mixing weights jointly determined by the integrated variance and the noise term and so should be leptokurtotic.

3.4 Data and Descriptive Statistics

The dataset used here is the same as the previous two chapters. We consider four realized volatility measures: realized variance, realized range and realized power and bipower variations. Our main concentration, in line with the previous two chapters, is the 5-minute sampling frequency. It has been argued that sampling frequency influences the outcome of standardization (Fleming and Paye, 2007). However, the extent to which the sampling frequency may distort the distribution of standardized returns has not yet been examined using a large data set. To fill this gap, we further construct the realized volatility measures and returns using a sampling frequency of 1 second, which is the highest possible frequency for this dataset. The summary statistics of 1 second aggregated returns and realized volatility measures are given in **Table 3.1**.

[Table 3.1 here]

We first look at the properties of 1-second returns. The difference, not surprisingly, between 1-second and 5-minute aggregated returns is not that deviated from each other. The mean values of returns are similar between the two sampling frequencies. The standard deviation of 1-second returns is lower, as is the kurtosis. 5-minute returns have lower skewness in general. The Jarque-Bera statistics of 1-second returns is also lower than that of 5-minute returns.

In comparing **Table 3.2** with **Table 1.4**, three of four realized volatility measures constructed at 1-second sampling frequency show much higher values for the mean and standard deviation. This is because market microstructure noise is overwhelmed at such a high sampling frequency. The exception of rr_t arises from the fact that rr_t is constructed with a bias-correction factor in the equation. The price change between the highest and lowest prices is at a minimum for the 1-second frequency. rr_t is also downward biased when the trading is infrequent. Our results are consistent with those reported by Fleming and Paya (2011) who also find an upward bias for realized volatility measures (rv_t in their case) when sampling at 1 second using trade data. On average, the means of realized volatility measures are at least twice as large at 1-second sampling frequency as at 5-minute sampling frequency. This difference is even larger in actively traded sectors. The sector average rv_t in IT sector is 5.722, it increases to 37.810 when the sampling frequency increases to 1 second. Similar results are reported for FIN and HC sectors. The dramatic difference between realized volatility measures sampled at different frequencies shows the impact of market microstructure noise to be most severe at the highest sampling frequency.

[Table 3.2 here]

We construct the continuous variance and jumps series from rv_t and bv_t, according to [1.14] and [1.15]. **Table 3.3** reports the summary statistics of cv_t, jumps and rv_t. When sampling at 1-second frequency, the size of jumps increases accordingly. Jumps also become strictly positive. However, large jumps are still not necessarily associated with high volatility days – a finding we have reported in Chapter 1 under 5-minute sampling frequency.

In summary, at 1-second sampling frequency, returns, realized volatility measures, as well as jumps, are different from the 5-minute sampled counterparts. 5-minute is the most commonly used as various studies have shown that 5-minute sampling frequency most efficiently balances the intraday information content and the microstructure noise. In Chapter 1, we drew the volatility signature plots of two stocks and showed that 5-minute is an optimal sampling frequency for rv_t, bv_t and rr_t. As stated earlier, it is still interesting to compare the standardization results of 1-second data with those of 5-minute data. This comparison allows us to investigate the extent of the influence of jumps and microstructure noise arising from sampling frequencies. In the following empirical results section, we start reporting the standardization results from returns standardized by realized volatility measures from various sampling frequencies.

[Table 3.3 here]

[Figure 3.1 and 3.2 here]

3.5 Empirical Results

In this section, we report the results of the distributional properties of standardized returns. We first investigate the aggregational Gaussianity of intraday returns using 10 stocks. Then we compare the returns standardized by realized volatility measures at 1-second and 5-minute sampling frequencies, respectively. 5-minute sampled realized volatility measures work more successfully than 1-second sampled ones for all the sectors. Next, we compare the standardization power between realized volatility measures and two parametric volatility measures, namely, GARCH and stochastic volatility. This comparison is based on 5-minute sampled data only. We find that the non-parametrically estimated realized volatility measures have stronger power to achieve the returns normality. Fourth, we investigate the impact of jumps. The presence of jumps in the realized volatility has little impact of recovery the return normality, at least at 5-minute sampling frequency. At 1-second sampling frequency, excluding jumps from realized volatility shows an improvement of recovering normality. Fifth, we use a moving average filter to microstructure noise in the realized volatility measures. The MA filtration removes the microstructure noise contained in the realized volatility measures and enhances the performance of realized volatility measures when used to standardize the returns. In the last part of this section, we select the 30 most actively traded stocks (T30 stocks) in the whole sample and summarize the standardization results of these stocks. The results from the T30 stocks suggest that the level of stock activity matters in recovering the normality.

3.5.1 Aggregational Gaussianity of Returns

Aggregational Gaussianity is a well-established stylized fact of asset returns. It implies that the distribution of returns converges more closely to a standard normal as the time scale that used to calculate the returns increases. This statistical property describes returns distribution behaviour and has important applications in modelling tail risk behaviour, see for example Eberlein and Keller (1995), Barndorff-Nielsen (2001), Conti (2001), and Roger *et al* (2011).

In this section, we aim to establish whether stocks from different market sectors converge to Gaussianity at similar rates. We are also interested to see whether the capability of the realized measures to restore normality is associated with the degree returns conforms to aggregational Gaussianity.

We select 10 representative stocks (one from each sector), where normality could be restored for 5/10 at both the 1-second and 5-minute frequency and not for the remaining. For each of the stocks we observe returns at 5, 30, 60 and 300 seconds. For the same stocks we calculate realized volatility at the 1-second and 5-minute frequencies to standardize the daily returns.

Table 3.4 shows the results of aggregational Gaussianity (left panel) and the standardized returns (right panel). The left panel displays the percentage change in the JB test statistic relative to the previous sampling frequency whereas the right panel reports the standardized test statistics.

The results from the left panel shows return distributions gradually converge to normality as the sampling frequency decreases. We observe a significant decrease of the JB statistics as we move from 1-second to 5-minute, with two exceptions of PG (CS sector) and AA (MAR sector). When the time scale increases, the assumptions underlying the CLM starts to hold as the time scale is more uniformly.

The standardized returns (right panel) all show lower JB test statistics at 5-minute frequency than their counterparts at 1-second frequency. This conclusion holds for both stock groups irrespective of whether normality is achieved. The largest difference in the JB test statistics between the 1-second and 5-minute frequencies statistics are shown in "Sprint Nextel" happens to be for the most jumps-contained sector.

Despite normality converging rate is relatively constant, the average converging rate for the stocks where returns normality could be recovered higher. This finding holds for all the intraday sampling frequencies that have been considered. This result is interesting as it may identify a general criterion as whether returns normality could be restored. It also shows that assumptions of the central limit theorem tend to be affected to a certain extent by the nature of the stock under consideration.

[Table 3.4 here]

3.5.2 Standardized Returns: Different Sampling Frequencies

Table 3.5 shows the results of returns standardized by realized volatility measures which are aggregated both at the 5-minute and 1-second sampling frequencies, respectively. In the table, we report the percentage of the stocks that return normality is rejected at the 5% level according to Jarque-Bera normality test. The results are presented both at sector level and overall level. [11]

[Table 3.5 here]

[11] For brevity, we only report the percentage of rejections based on the JB test statistics at 5% level. The detailed statistics of JB test together with skewness and kurtosis of each single stock is provided in the appendix. We also test the normality using the Kolmogrov-Smirnov (KS) density test. The KS test provides the identical conclusion as JB test.

The result of standardized returns aggregated at 5min has been reported and discussed in Chapter 1. To summarize briefly: pv_t outperforms the three other realized volatility measures in most of the sectors and overall. bv_t works second best. rv_t is the worst performed realized volatility measure. The two most successfully recovered sectors are ENG and UTL whereas the two least successful sectors are MAR and IT. Among these four sectors, UTL is the sector by which a large number of daily and intraday jumps are detected and IT is detected with least daily and intraday jumps.

The result of returns standardized by 1-second aggregated realized volatility measures is in the lower panel. Comparing the 1-second result with the 5-minute result, the percentage of rejections has increased at overall level for all the realized volatility measures. The increase is more dramatic for pv_t, bv_t and rr_t, of which the rejections increase from 34%, 42% and 48% to 77%, 86% and 94%, respectively. Compared with other realized volatility measures, the percentage of rejections based on rv_t standardized returns increases to a lesser extent, from 60% to 66%.

One explanation for the smaller increase of normality rejections on rv_t standardized returns is provided by Fleming and Paye (2011). They suggest that upward bias in realized variance due to the microstructure noise, which is more prominent in 1-second aggregated rv_t, can reduce the standard deviation and increase the kurtosis of standardized returns and hence lead to a false appearance of normality. This at least partially explains why the returns normality of some stocks is rejected when sampled at 5-minute, is not rejected at when sampled at 1-second frequency.

Figures 3.3a, b and **3.4a, b** show respectively the density plots and QQ plots of sector-average raw returns sampled at 5-minute and 1-second frequencies. The distributions are not normal and have long tails. The density plots and QQ plots of returns standardized by realized volatility measures sampled at 1-second and 5-minute frequencies are given respectively in **Figure 3.5 to 3.12**. Here we also use sector average returns and realized volatility measures. This is for both simplicity and the representation of the sectors.

Although providing less formal evidence than the statistics, the density and QQ plots are nevertheless informative. It is evident that the realized volatility measures standardized returns are much closer to the standard normal distribution than raw returns both at 1-second and 5-minute levels, especially for the tails of the QQ plots. The tails have been greatly shrunk than raw returns and show only small deviations from the 45-degree lines. In comparing the density and QQ plots of 1-second and 5-minute standardized returns, the plots show that returns standardized by 5-minute sampled returns are closer to the Gaussian distribution than their 1-second counterparts. This finding is most evident for rr_t standardized returns. At the 5-minute sampling frequency, the density and QQ plots of rr_t standardized returns are better fitted than the rv_t standardized. For the 1-second sampling frequency, both plot types clearly deviate from the standard normal.

[Figure 3.3a, b to 3.12a, b here]

Here we use the sector-average realized volatility measures to standardize sector-average daily returns. The figures are expected to be different from the results of individual stocks. This is because taking average value might remove the idiosyncratic risks presented in the individual stocks. However, the sector-average standardized returns also help us observe some trends which are less obvious at individual stock level. The plots confirm that some sectors are more difficult to standardize than others.

By comparing the power to recover returns normality of different realized measures at different sampling frequencies, we find that the effect of microstructure noise, which arises from change to the sampling frequency, plays a more important role when the sampling frequency is ultra-high. In the next section, we further address the effect of the presence of jumps, by comparing the recovery of normality for realized volatility measures which do, and do not contain jumps.

3.5.3 Standardized Returns: with and without Jumps

Table 3.6 reports the percentage of normality rejections for returns standardized by realized variance, realized bipower variation, realized power variation and continuous variance. Except for rv_t , the remaining three measures are all jump robust. Continuous variance, whose construction is based on the theoretical framework of Huang and Tauchen (2005), is by definition, the continuous path of the quadratic variation and converge to the integrated variance.

[Table 3.6 here]

In general, jump robust realized volatility measures recover returns normality more successfully than rv_t. Extracting jumps from rv_t does not always aid the recovery of normality. The percentage of normality rejection of returns standardized by cv_t is higher than for returns standardized by pv_t and bv_t, although cv_t is expected to be a more accurate estimator of the integrated variance. The percentage of normality rejection for $r_t / \sqrt{cv_t}$ is 46%, and for $r_t / \sqrt{bv_t}$ and $r_t / \sqrt{pv_t}$ is 42% and 34%, respectively.

ABFN (2010) report similar findings from their investigation of 30 DJIA stocks. They argue that a large jump tends to increase the (absolute) value of returns and realized volatility of standardized returns. Therefore, the impact of jumps is muted. As discussed in the previous section, the sector that contains the least number of significant jumps (IT) is also the sector that records the highest rejection percentage. This failure not only applies to returns standardized by rv_t, but also to returns standardized by pv_t, bv_t and cv_t.

To further investigate the effects of jumps to recovering normality, we also compute the cv_t from 1-second aggregated data and find that the rejection percentage of cv_t standardized returns is even higher (77%) than that for rv_t standardized returns (66%). However, the rejection of $r_t / \sqrt{bv_t}$ is 86%. At the 1-second aggregation level, cv_t is a more efficient realized volatility measures than bv_t.

Here we plot the density plots and QQ plots of sector-average returns standardized by cv_t. First, we find that the plots of cv_t standardized returns mimic bv_t standardized returns. At the 5-minute level, the plots of $r_t / \sqrt{cv_t}$ and $r_t / \sqrt{bv_t}$ are much closer than their 1-second level counterparts. This finding is also confirmed by the skewness and kurtosis as well as the JB statistics both at individual stock level and sector average level. For some stocks, cv_t is the least successful volatility measure in restoring the return normality. The plots from 1-second data tell a different story. Plots of $r_t / \sqrt{cv_t}$ show closer fitness of Gaussianity than those of $r_t / \sqrt{bv_t}$.

[Figure 3.13a, 13b and Figure 3.14a, 14b here]

To summarize, the standardization results for both sampling frequencies suggest that the effect of jumps in recovering normality is more of a problem at the 1 second aggregation level. At the 5 minute level, the presence of jumps is less of a concern. Instead of jumps, microstructure noise within realized volatility measures from the ultra-high sampling frequency is the predominant factor in biasing returns normality.

3.5.4 Standardized Returns: Moving Average Filtration

The results from previous sections suggest that microstructure noise is the important factor in distorting the normality of standardized returns. Extensive attempts have been made to eliminate the microstructure noises in realized volatility measures. Early attempts can be traced to Zhou (1996) who uses a simple moving average filter. Later attempts include those of Maheu and McCuddy (2002), Ait-Sahalia *et al* (2005), Russell and Bandi (2006, 2008), Oomen (2005), Owens *et al* (2006), Zhang *et al* (2005), Zhang (2006), Hansen, Large and Lunde (2006) and Barndorff-Nielsen *et al* (2008). These papers either construct realized volatility measure that is robust to microstructure noise (realized kernel for instance), or separate the noise from rv_t, or smooth the noise contained in the rv_t. The above-mentioned attempts primarily focus on realized variance only and leave the three other realized volatility measures unexamined. To smooth the microstate noise, we apply a moving average filter to daily realized volatility measures. The use of moving average filter for realized variance has been considered in ABDE (2001), Maheu and McCuddy (2002) and Hansen, Large and Lunde (2006), to smooth intraday returns before constructing realized variance. Hansen, Large and Lunde (2006) show that MA-based realized volatility estimator is consistent and asymptotically Gaussian distributed about the integrated variance under restrictive assumptions. Empirically, improved performance of realized variance has been confirmed using both individual stocks and foreign exchange data. Inspiring from yet contrasting with the MA method used in other papers, we propose the Exponentially Weighted Moving Average (EWMA) filter, which can be applied not only to the realized variance, but also to the reaming three realized volatility measures. Market microstructure is more predominant at the intraday level than at the daily level, as long as MA filter is effective at the intraday level, it is expected to be effective at daily level as the microstructure noise tends to mitigate at the daily level. The EWMA filter is shown to be easy to apply and is especially widely used in volatility forecasting. It generally yields the lowest MSE among other forecasting models. (For a detailed survey of EWMA, see Poon and Granger, 2003).

An *n*-period EWMA of a time series y_t is defined as

$$\mu_t(n) = \sum_{i=0}^{n-1} \omega_i \cdot y_{t-1}, \quad \omega_i = \frac{\lambda^{i-1}}{\sum_{i=0}^{n-1} \lambda^{i-1}} \qquad [3.10]$$

As *n* converges to infinity, $\lambda^n \to 0$, $\omega_n \to 0$ and the EWMA converges to

$$\mu_t(\lambda) = (1-\lambda)\sum_{i=0}^{\infty} \lambda^i y_{t-i} \qquad [3.11]$$

Therefore, the EWMA may be defined independently of the window length *n*. The EWMA in [3.11] may be computed using the recursion

$$\mu_t(\lambda) = (1-\lambda)y_t + \lambda\mu_{t-1}(\lambda) \qquad [3.12]$$

where y_t is the realized volatility measure and λ is known as the decay parameter or smoothing parameter and takes the value between 0 and 1.

It is clear that the closer the decay parameter to 1, the more weight is put on the previous period's estimate relative to the current period's value (day in this case). We consider three values: 0.25, 0.5 and 0.75. The result for $\lambda = 0.5$ is reported as this weighting provides the best outcome in the filtration[12].

We plot the EWMA filtered realized volatility measure in **Figure 3.15**. The figure shows that the filtered volatility measures are less noisy than the original series.

Table 3.7 reports the percentage of normality rejections of returns standardized by MA filtered realized volatility measures. In some cases, using filtered realized volatility series brings the standardized returns closer to normal. In general, the percentage of rejection decreases to 38% for $r_{MA} / \sqrt{rv_{MA}}$ and to 40% for $r_{MA} / \sqrt{bv_{MA}}$ (the percentage of rejection for $r_t / \sqrt{rv_t}$ is 60% and for $r_t / \sqrt{bv_t}$ is 42% respectively). EWMA works exceptionally well on rv_t. Before applying MA filtration on rv_t, the highest rejection of $r_t / \sqrt{rv_t}$ is 80% from FIN, IT and MAR sectors. When standardized by MA filtered rv_t, the percentage of rejection decreases to 20%, 60% and 30% respectively. The rejection of CD and TEL is 70% under raw series and decreases to 40% and 20% respectively after filtration. These results indicate a considerable improvement. However, the percentage of rejection for returns standardized by MA filtered pv_t and rr_t further increase. The MA filter is less able to enhance the returns normality of the stocks that normality is recovered by raw realized volatility measures.

[Table 3.7 here]

[Figure 3.15 here]

[12] Another noise filtration method is proposed by Owens *et al* (2006) who filter the noise within RV using the Kalman filter and Kalman smoother. We investigate these filters as well and find that they are less effective than EWMA in terms of recovery returns normality.

3.5.5 Standardized Returns: Parametric and Nonparametric Volatility Measures

In this section, we standardize returns by the volatility of parametric and non-parametric measures. The extent of normality that can be achieved in the standardized returns acts as an indicator of the performance of volatility measures (parametric and non-parametric) in accounting for the factors causing returns non-normality.

The normality of returns standardized by two parametric volatility measures, GARCH (1, 1) ($garch_t$) and stochastic volatility (sv_t), is rejected at the 5% level for all stocks. Normality is greatly improved as compared with raw returns. There is a contraction in the tails of $garch_t / sv_t$ standardized returns relative to the raw returns, yet they remain significantly leptokurtic. The distributions of $sv_t / garch_t$ standardized returns all show excess kurtosis. When used as the standardized factor, $garch_t$ and sv_t display a similar performance and tend to depart from normality by similar magnitudes. The density and QQ plots of $garch_t / sv_t$ standardized returns from **Figure 3.16** and **3.17** are less satisfactorily fitted when compared with non-parametric volatility standardized returns.

Same conclusion is reached in ABD (2000b) and ABFN (2010). Both papers consider the case of $garch_t$ only. The difference between parametric and non-parametric volatility measures lies mainly in the estimation approach, as ABD (2000) state. The parametric measures are estimated conditional on the discrete path of returns up to day *t*, whereas non-parametric volatility measures are estimated conditional on the continuous path of stochastic volatility up to and including day *t*. The degree of information contained in the two volatility measures is shown to be quite distinct.

These results indicate that the non-parametric volatility measures are more capable in capturing the dynamics of the market and hence in recovering returns normality. If we were to measure volatility accuracy based on recovering normality, we might conclude that $garch_t$ and sv_t were not that accurate in comparison with the non-parametric measures.

[Figure 3.16a, b and 3.17a, b here]

3.5.6 Standardized Returns: 30 Most Actively Traded Stocks

Another focus of this chapter is whether the recovering power of realized volatility measures is associated with the level of stock activity. We investigate the 30 most actively traded stocks within the sample. We are particularly interested in this subsample because most of these stocks are the most extensively studied in the literature. They may also draw more implication for the practitioners.

The standardization results according to the percentage of rejection are reported in **Table 3.8**. To compare the different performance of the top 30 stocks and the overall sample, we also report the results of 100 stocks in the lower panel.

The 30 most actively traded stocks contain fewer significant jumps on average than the overall sample. However, the percentage of normality rejections is obviously higher. This result further suggests that the number of jumps contained in the realized volatility measures and returns has little impact in recovering returns normality through standardization processes. On the other hand, liquid stocks have less microstructure noise than stocks with larger trading volumes. From their examination of the relation between microstructure noise and various liquidity measures using all NYSE traded common stocks over 10-years, Ait-Sahalia and Yu (2009) conclude that

"*Trading volume, which aggregates the information in trade size and number of trades, is positively correlated with noise*". Hence, T30 stocks are expected to have more noise on average.

Here we also consider the effect of MA filtration on the T30 stocks. The percentage of rejections is significantly changed: 80% to 13% from $r_t / \sqrt{rv_t}$ to $r_{MA} / \sqrt{rv_{MA}}$, 67% to 17% from $r_t / \sqrt{bv_t}$ to $r_{MA} / \sqrt{bv_{MA}}$, and 47% to 40% from $r_t / \sqrt{rr_t}$ to $r_{MA} / \sqrt{rr_{MA}}$. The only exception is for pv_t, where rejections increase from 43% to 57%.

In our sample, more than half of the T30 stocks come from FIN, HC and IT sectors (18/30). We reported in the previous section the considerable reduction of normality rejections of $ma\text{-}rv_t$ standardized returns in these sectors. The overall results from T30 stocks further strengthen the augment: except for pv_t, applying MA on the other three realized volatility measures successfully filters the microstructure noise contained in the realized volatility measures and hence enhance the power of the realized volatility measures of standardizing returns. The higher the level of noise that is contained in the stock prices, the more powerful is the MA filter.

Moreover, one explanation for the failure of $ma\text{-}pv_t$ is that pv_t itself is already a consistent and efficient estimator of the integrated variance. Smoothing realized power variation erases valuable dynamics/information that weakens its power to capture the dynamics of the returns process and hence renders it incapable of capturing the factors causing non-normality.

[Table 3.8 here]

3.6 Conclusion

Different realized measures of volatility, namely realized variance, realized range, realized power and bipower variations, sampled at the two frequencies (1-second and 5-minute, respectively) have been compared based on their power to recover the normality of 100 stocks from 10 sectors. The most noteworthy findings are summarized as follows:

Realized power variation is the most efficient unfiltered[13] realized volatility measure in terms of recovering returns normality.

Adjusting for jumps has little (if any) impact upon the restoration of normality to standardized returns at the 5-minute sampling frequency. However, removing jumps from realized variance enhances the power of restoring normality to standardized returns at the 1-second sampling frequency. Jumps are a more likely reason for distortion to the distributions of standardized returns at the higher sampling frequency.

The non-parametric measures are superior in recovering returns normality when compared to the parametric GARCH and stochastic volatility. Both GARCH and stochastic volatility fail to recover normality in all of the cases considered.

The presence of market microstructure noise in realized volatility measures is the main factor distorting the Gaussianity of standardized returns.

The stocks that are highly traded are found to contain fewer jumps yet are more difficult to achieve returns normality under the raw realized volatility measures. Exponentially Weighted Moving Average filtered realized volatility measures bring normality in cases where the unfiltered series fail. This moving average filter works better on more actively traded stocks which are constructed at the 5-minute sampling frequency. This does not apply to the 1-second sampled realized volatility measures.

[13] "Unfiltered" refers to the volatility series that are not filtered by the EWMA.

Tables

Table 3.1: Summary statistics of sector average daily returns sampled at 1-second frequency (in %)

	Daily Return (R_t)									
	CD	**CS**	**ENG**	**FIN**	**HC**	**IND**	**IT**	**MAR**	**TEL**	**UTL**
Mean	0.034	0.050	-0.038	-0.022	-0.002	-0.002	0.019	-0.072	-0.002	0.011
Maximum	12.969	7.935	13.000	11.710	9.480	6.507	14.860	9.972	6.507	11.060
Minimum	-8.744	-6.762	-13.495	-16.987	-8.679	-7.461	-7.506	-13.917	-7.461	-8.170
S.D.	1.512	0.939	1.697	2.000	1.187	1.293	1.685	1.786	1.293	1.255
Skewness	0.437	0.272	-0.398	-0.283	0.056	-0.180	0.283	-0.596	-0.180	0.044
Kurtosis	9.708	10.621	11.325	14.103	9.143	6.776	7.653	8.958	6.776	10.959
JB	5276.202	6730.441	8063.353	14249.520	4352.462	1658.602	2533.426	4256.286	1658.602	7303.437
ADF	-30.796	-41.830	-12.933	-26.495	-14.927	-12.324	-20.800	-10.824	-12.554	-39.580
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: Summary statistics of daily returns for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation. ADF denotes the augmented Dickey-Fuller statistics for the null of a unit root with 5% and 1% critical values of 2.862 and -3.433 respectively.

Table 3.2: Summary statistics of realized volatility measures (1-second frequency)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					Realized Variance (rv_t)					
Mean	19.132	6.325	15.838	23.250	13.097	9.411	37.810	13.764	9.411	8.849
Maximum	1735.920	183.691	926.965	834.948	248.949	318.203	921.619	469.200	318.203	3056.518
Minimum	1.963	0.624	1.258	0.716	1.582	0.977	3.087	1.981	0.977	0.528
S.D.	40.932	9.953	38.510	53.897	16.344	12.864	77.337	20.954	12.864	70.348
Skewness	27.568	6.881	13.693	8.264	3.913	8.915	5.397	10.420	8.915	36.346
Kurtosis	1123.083	85.521	247.210	95.714	30.373	155.993	41.199	171.735	155.993	1452.115
					Realized Range (rr_t)					
Mean	3.121	0.874	2.126	9.626	1.952	1.595	10.349	2.359	1.595	1.340
Maximum	732.703	252.317	357.327	885.489	171.930	192.823	631.591	414.894	192.823	1300.007
Minimum	0.120	0.056	0.024	0.022	0.089	0.026	0.455	0.024	0.026	0.008
S.D.	15.118	5.847	10.148	36.445	4.605	5.475	25.135	11.108	5.475	26.051
Skewness	41.279	33.711	20.196	11.869	20.962	19.989	9.137	24.423	19.989	45.864
Kurtosis	1966.744	1328.622	594.086	211.605	694.267	605.469	160.524	800.140	605.469	2247.235
					Realized Power Variation (pv_t)					
Mean	56.955	23.224	43.250	69.248	44.221	35.201	111.501	43.806	35.201	22.582
Maximum	612.664	330.327	1002.797	1335.332	530.175	646.562	1319.961	920.776	646.562	576.207
Minimum	9.645	4.227	8.068	4.148	7.776	5.219	18.348	8.184	5.219	3.238
S.D.	49.717	22.281	51.527	107.554	34.317	32.319	143.307	48.501	32.319	31.164
Skewness	2.833	4.087	7.950	5.101	3.188	5.160	3.793	6.152	5.160	8.976
Kurtosis	16.416	32.499	100.894	42.011	23.409	61.001	20.669	67.271	61.001	121.064
					Realized Bipower Variation (bv_t)					
Mean	15.307	5.103	12.825	19.237	10.721	7.734	32.261	10.689	7.734	7.007
Maximum	521.104	151.837	875.707	697.669	219.567	282.746	832.096	411.028	282.746	2427.890
Minimum	1.460	0.508	1.036	0.571	1.277	0.740	2.436	1.533	0.740	0.402
S.D.	22.725	8.327	34.300	44.767	13.607	11.028	66.106	17.687	11.028	57.129
Skewness	6.973	7.119	14.421	8.369	4.142	9.543	5.366	11.026	9.543	35.463
Kurtosis	107.949	89.329	271.916	99.409	34.720	176.539	41.567	187.378	176.539	1378.429

Note: Summary statistics of daily volatility measures for the 100 stocks over the period 01/01/2000 to 31/12/2010 are reported at sector average level. S.D. denotes standard deviation.

Table 3.3: Summary statistics of realized variance and its components (1-second frequency)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL
					Realized Variance (rv_t)					
Mean	19.132	6.325	15.838	23.250	13.097	9.411	37.810	13.764	9.411	8.849
Maximum	1735.920	183.691	926.965	834.948	248.949	318.203	921.619	469.200	318.203	3056.518
S.D.	40.932	9.953	38.510	53.897	16.344	12.864	77.337	20.954	12.864	70.348
Skewness	27.568	6.881	13.693	8.264	3.913	8.915	5.397	10.420	8.915	36.346
Kurtosis	1123.083	85.521	247.210	95.714	30.373	155.993	41.199	171.735	155.993	1452.115
					Continuous Variance (cv_t)					
Mean	18.902	6.092	15.593	23.023	12.889	9.197	37.612	13.491	24.138	8.581
Maximum	1734.716	183.501	926.908	834.755	248.824	318.085	921.517	469.068	1091.594	3056.288
S.D.	40.927	9.963	38.522	53.913	16.349	12.876	77.355	20.964	38.577	70.352
Skewness	27.534	6.874	13.690	8.259	3.913	8.901	5.395	10.418	10.684	36.343
Kurtosis	1121.161	85.355	247.098	95.631	30.370	155.611	41.170	171.624	230.886	1451.924
					Jumps (J_t)					
Mean	0.230	0.233	0.245	0.227	0.208	0.215	0.198	0.273	0.243	0.268
Maximum	1.860	1.459	1.035	1.049	0.704	1.073	0.890	1.618	1.251	1.556
S.D.	0.091	0.071	0.076	0.079	0.053	0.053	0.066	0.098	0.077	0.081
Skewness	7.309	5.699	2.279	1.851	1.879	2.710	0.543	2.558	2.344	3.223
Kurtosis	107.876	67.344	17.344	15.752	14.293	37.404	8.747	24.581	29.030	35.616

Note: This table reports the mean, standard deviation, skewness, kurtosis and maximum at sector average realized variance and its continuous and non-contiguous components from 03/01/2000 to 31/12/2010.

Table 3.4: Aggregational Gaussianity of returns

		Intraday Returns					$r_t / \sqrt{rv_t}$		
		1sec	5sec	30sec	60sec	300sec		1sec	300sec
% change in JB	Procter & Gamble	0.00%	-70.93%	-99.20%	28.73%	-100.00%	JB	8.477	0.786
	(PG)						p value	(0.014)	(0.675)
% change in JB	Exxon Mobil	0.00%	-98.66%	-99.16%	-62.57%	-93.43%	JB	5.215	4.968
	(XOM)						p value	(0.074)	(0.083)
% change in JB	Bank of NY Mellon	0.00%	-86.17%	-98.91%	-53.09%	-94.47%	JB	6.120	4.691
	(BK)						p value	(0.047)	(0.096)
% change in JB	Dell Inc.	0.00%	-77.98%	-96.36%	-84.09%	-89.29%	JB	6.944	3.409
	(DELL)						p value	(0.031)	(0.182)
% change in JB	Entergy Corp.	0.00%	-99.97%	-99.86%	-76.84%	-100.00%	JB	7.679	0.963
	(ETR)								
average % change in JB			-86.74%	-98.70%	-49.57%	-95.44%	p value	(0.022)	(0.618)
% change in JB	The Home Depot	0.00%	-81.14%	-97.53%	-75.45%	-98.03%	JB	27.730	7.850
	(HD)						p value	(0.000)	(0.020)
% change in JB	Amgen Inc.	0.00%	-99.11%	-98.86%	-56.66%	-74.76%	JB	38.027	11.105
	(AMGN)						p value	(0.000)	(0.004)
% change in JB	General Electric	0.00%	-81.50%	-96.99%	-54.18%	-82.83%	JB	91.549	19.832
	(GE)						p value	(0.000)	(0.000)
% change in JB	Alcoa	0.00%	-45.41%	-98.94%	16.78%	-87.67%	JB	3325.977	15.766
	(AA)						p value	(0.000)	(0.000)
% change in JB	Sprint Nextel Corp	0.00%	-74.38%	-88.61%	-61.95%	-97.68%	JB	4820.770	15.357
	(S)								
average % change in JB			-76.31%	-96.19%	-46.29%	-88.19%	p value	(0.000)	(0.000)

Note: This table reports the % of change in the Jarque-Bera statistics of the intraday returns calculated at different sampling frequencies and the Jarque-Bera statistics and p values of rv_t standardized returns sampled at 1-second and 5-minute frequencies. The sampling frequencies for aggregational Gaussianity are 1, 5, 30, 60 and 300 second, respectively. The 10 stocks are selected from each market sector.

Table 3.5: Percentage of rejection of normality on standardized returns sampled at different frequencies

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	Overall
	% of Rejection of Normality (5-minute)										
$r_t/\sqrt{rv_t}$	70%	20%	50%	80%	60%	60%	80%	80%	70%	30%	60%
$r_t/\sqrt{bv_t}$	60%	30%	10%	50%	30%	40%	90%	60%	30%	20%	42%
$r_t/\sqrt{pv_t}$	30%	40%	20%	50%	10%	30%	20%	40%	60%	40%	34%
$r_t/\sqrt{rr_t}$	80%	40%	10%	30%	40%	50%	50%	80%	50%	50%	48%
	% of Rejection of Normality (1-second)										
$r_t/\sqrt{rv_t}$	80%	100%	50%	80%	60%	50%	60%	60%	80%	40%	66%
$r_t/\sqrt{bv_t}$	90%	100%	80%	80%	90%	60%	80%	80%	100%	100%	86%
$r_t/\sqrt{pv_t}$	90%	100%	40%	80%	70%	60%	50%	80%	100%	100%	77%
$r_t/\sqrt{rr_t}$	100%	90%	100%	100%	90%	90%	70%	100%	100%	100%	94%

Note: The table reports the percentage of stocks for which the null hypothesis of normality of return is rejected based on the Jarque – Bera test when daily returns are standardized by volatility measures. r_t refers the daily returns, while rv_t, rr_t, pv_t, bv_t denote the realized variance, realized range, realized power variation and realized bipower variation respectively.The higher panel reports the rejection based on 5 minute sampling frequency while the lower panel reports the rejection based on 1 second sampling frequency. The percentage is based on 5% significant level.

Table 3.6: Percentage of rejection of normality on standardized returns sampled at different frequencies (including and excluding jumps)

	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	Overall
					% of Rejection of Normality (5-minute)						
$r_t / \sqrt{rv_t}$	70%	20%	50%	80%	60%	60%	80%	80%	70%	30%	60%
$r_t / \sqrt{bv_t}$	60%	30%	10%	50%	30%	40%	90%	60%	30%	20%	42%
$r_t / \sqrt{cv_t}$	60%	30%	20%	50%	60%	40%	90%	50%	50%	10%	46%
					% of Rejection of Normality (1-second)						
$r_t / \sqrt{rv_t}$	80%	100%	50%	80%	60%	50%	60%	60%	80%	40%	66%
$r_t / \sqrt{bv_t}$	90%	100%	80%	80%	90%	60%	80%	80%	100%	100%	86%
$r_t / \sqrt{cv_t}$	80%	100%	60%	80%	50%	70%	70%	60%	100%	100%	77%

Note: The table reports the percentage of stocks for which the null hypothesis of normality of return is rejected based on the Jarque – Bera test when daily returns are standardized by volatility measures. r_t refers the daily returns, while rv_t, bv_t, cv_t denote the realized variance, realized bipower variation and realized continuous variance, respectively. The higher panel reports the rejection based on 5 minute sampling frequency while the lower panel reports the rejection based on 1 second sampling frequency. The percentage is based on 5% significant level.

Table 3.7: Percentage of rejection of normality on standardized returns (MA filtered realized volatility measures)

	% of Rejection of Normality (5-minute)										
	CD	CS	ENG	FIN	HC	IND	IT	MAR	TEL	UTL	Overall
$r_{t(MA)} / \sqrt{rv_{t(MA)}}$	40%	40%	40%	20%	10%	30%	60%	30%	20%	90%	38%
$r_{t(MA)} / \sqrt{bv_{t(MA)}}$	60%	50%	40%	20%	20%	30%	20%	40%	30%	90%	40%
$r_{t(MA)} / \sqrt{pv_{t(MA)}}$	80%	90%	70%	70%	70%	70%	50%	60%	90%	100%	75%
$r_{t(MA)} / \sqrt{rr_{t(MA)}}$	70%	50%	80%	5%	40%	80%	50%	70%	80%	100%	67%

Note: The table reports the percentage of stocks for which the null hypothesis of normality of return is rejected based on the Jarque – Bera test when daily returns are standardized by EWMA volatility measures. r_t refers the daily returns, while rv_t, rr_t, pv_t, bv_t denote the realized variance, realized range, realized power variation and realized bipower variation respectively. The percentage is based on 5% significant level.

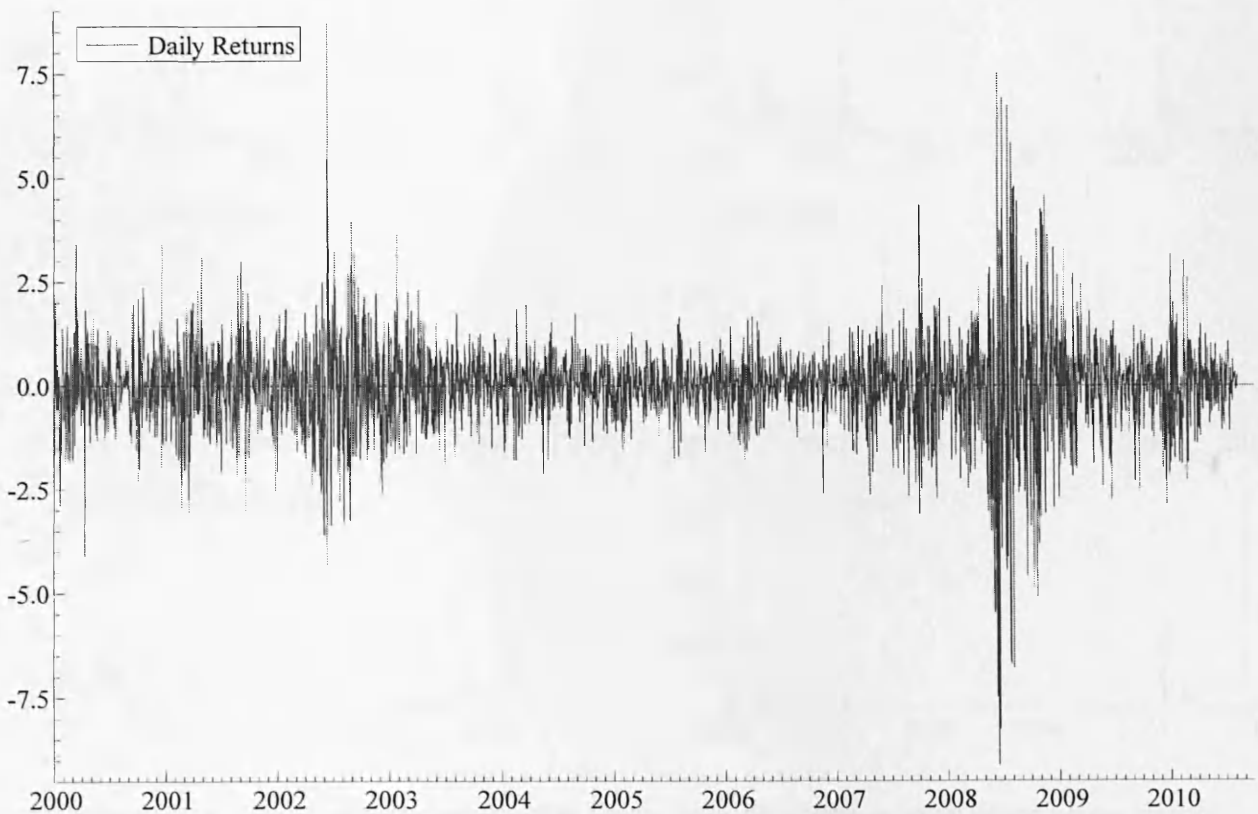
Table 3.8: Percentage of rejection of normality on standardized returns (T30 stocks)

% of Rejection of Normality (T30, 5-minute)			
$r_t / \sqrt{rv_t}$	$r_t / \sqrt{bv_t}$	$r_t / \sqrt{pv_t}$	$r_t / \sqrt{rr_t}$
80%	**67%**	**43%**	**53%**
$r_{t(MA)} / \sqrt{rv_{t(MA)}}$	$r_{t(MA)} / \sqrt{bv_{t(MA)}}$	$r_{t(MA)} / \sqrt{pv_{t(MA)}}$	$r_{t(MA)} / \sqrt{rr_{t(MA)}}$
13%	**17%**	**57%**	**40%**

Note: The table reports the percentage of stocks for which the null hypothesis of normality of return is rejected based on the Jarque – Bera test when daily returns are standardized by EWMA filtered and raw volatility measures for the 30 most actively traded stocks. The percentage is based on 5% significant level.

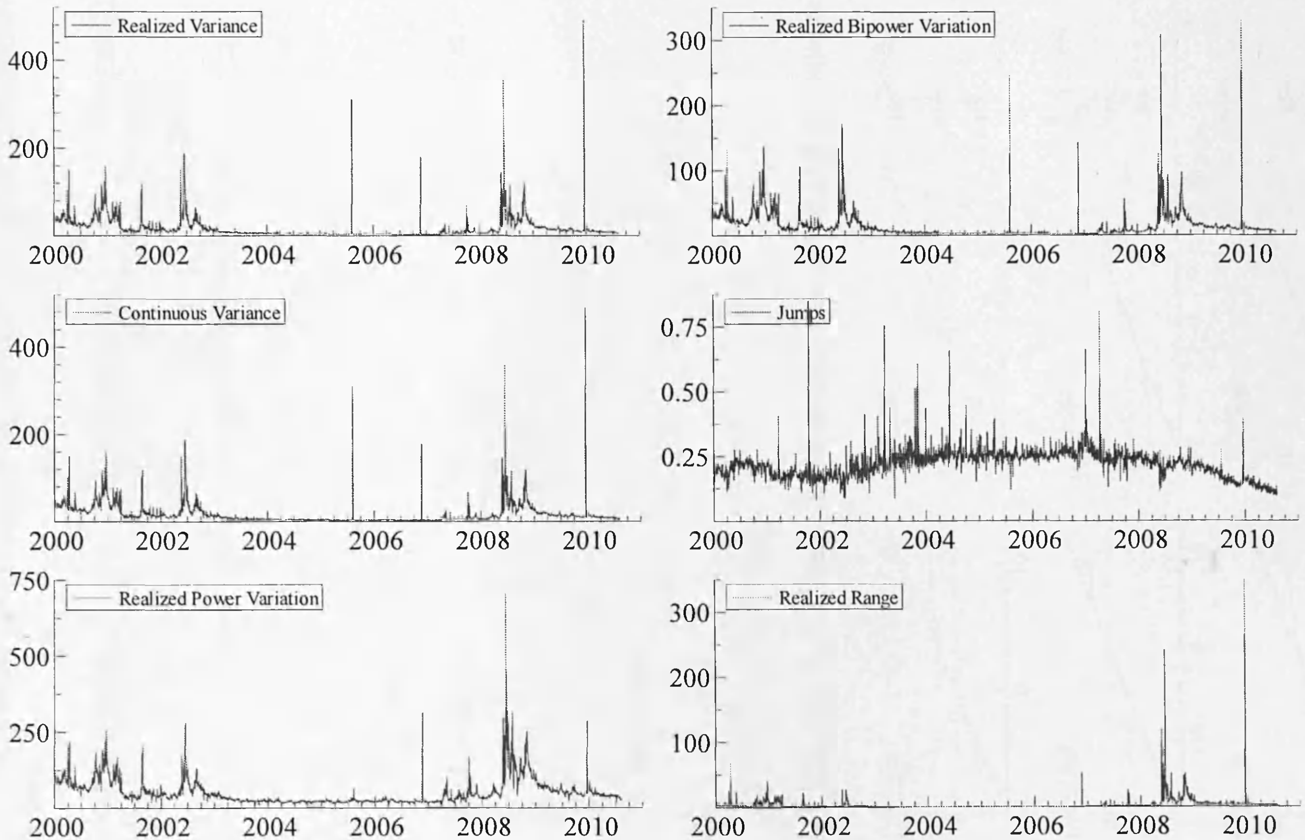
Figures

Figure 3.1: Plots of overall daily returns sampled at 1-second frequency (%)



Note: This figure presents the overall sample average daily returns which are sampled at 1 second frequency. The returns is reported by percentage.

Figure 3.2: Plots of overall average realized volatility measures, continuous variance and jumps (1-second sampling frequency)



Note: The figures reports the overall sample average realized volatility measures, jumps and continuous variance at 1 second sampling frequency.

Figure 3.3a: Density plots of sector-averages daily returns sampled at 5-minute frequency

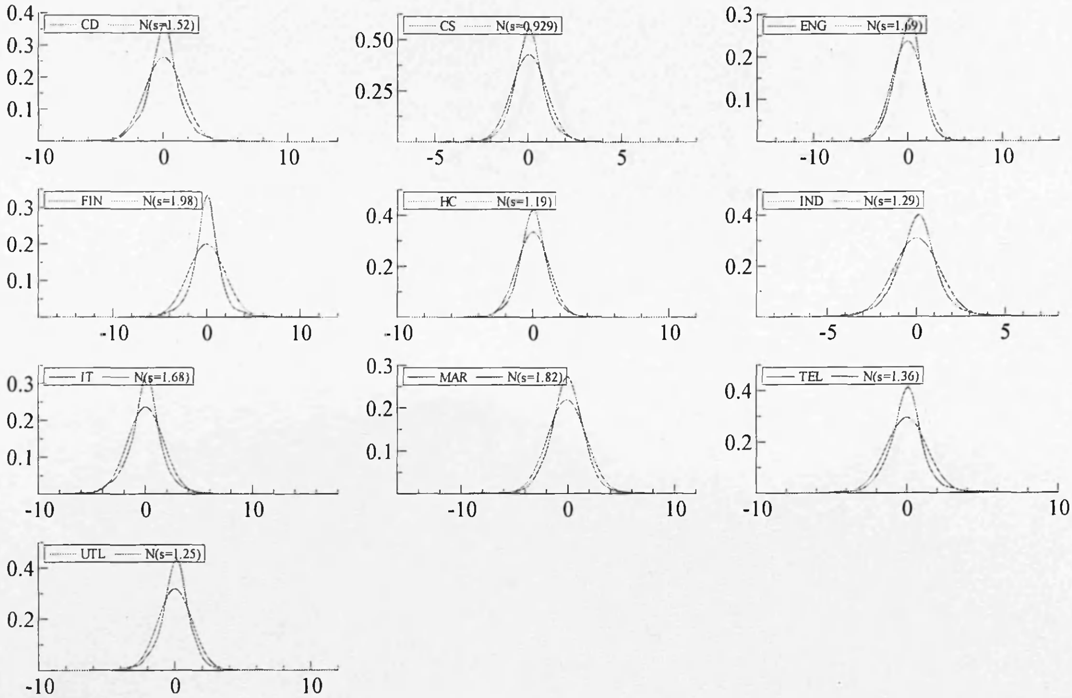


Figure 3.3b: QQ plots of sector-averages daily returns sampled at 5-minute frequency

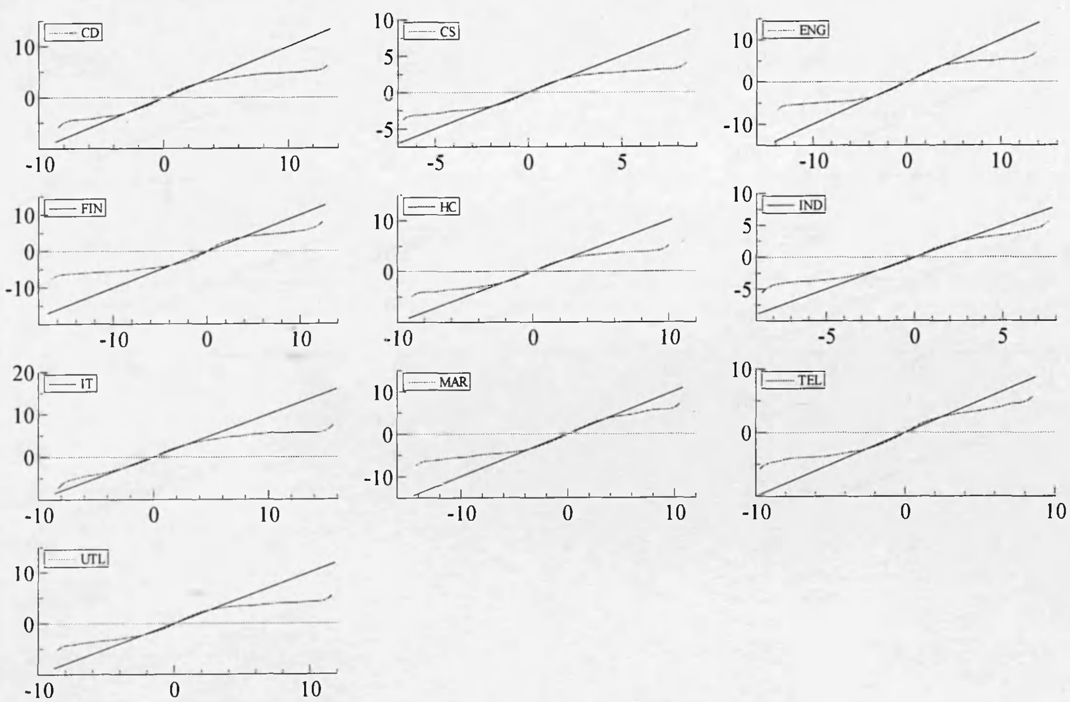


Figure 3.4a: Density plots of sector-averages daily returns sampled at 1-second frequency

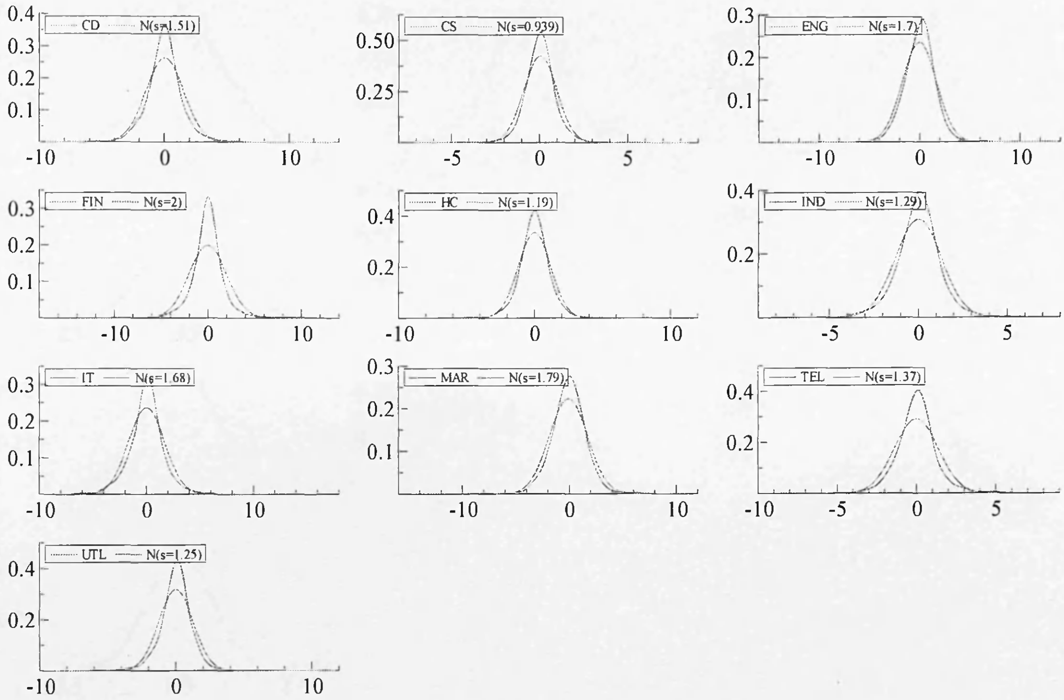


Figure 3.4b: QQ plots of sector-averages daily returns sampled at 1-second frequency

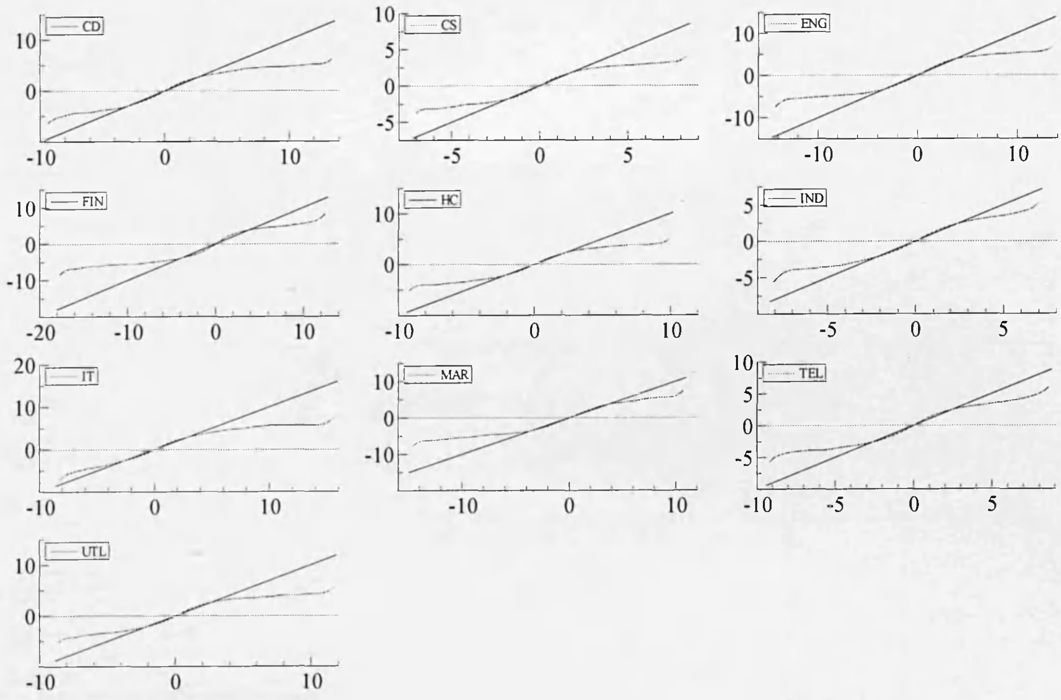


Figure 3.5a: Density plots of sector-averages daily returns standardized by realized variance sampled at 5-minute frequency

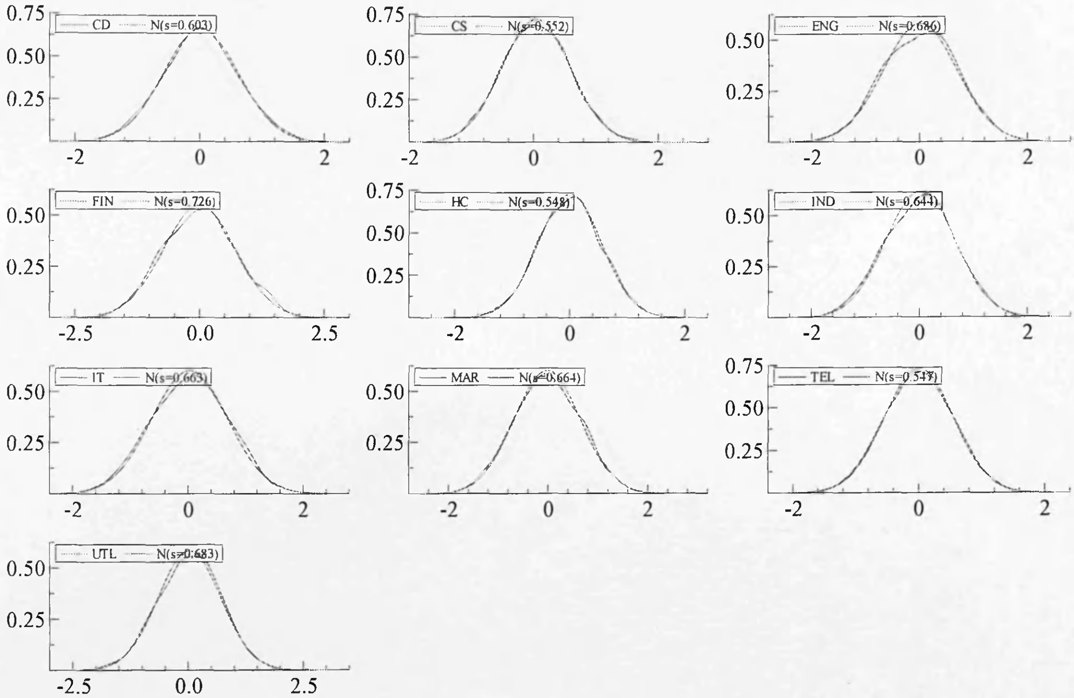


Figure 3.5b: QQ plots of sector-averages daily returns standardized by realized variance sampled at 5-minute frequency

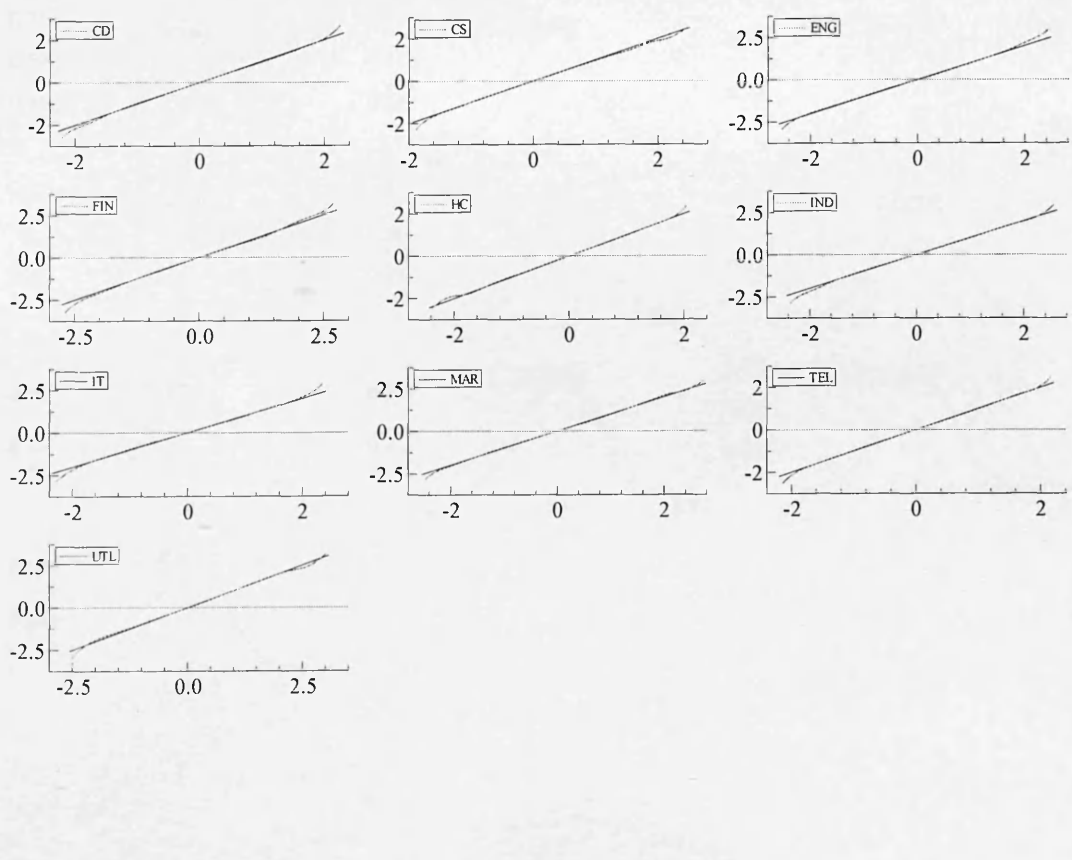


Figure 3.6a: Density plots of sector-averages daily returns standardized by realized bipower variation sampled at 5-minute frequency

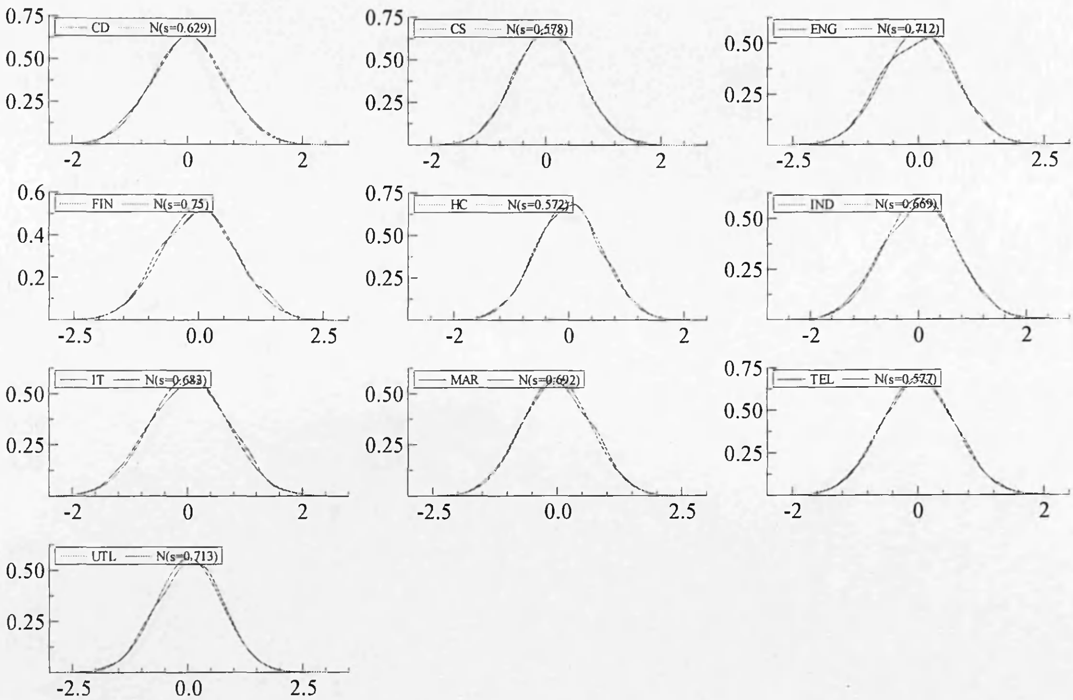


Figure 3.6b: QQ plots of sector-averages daily returns standardized by realized bipower variation sampled at 5-minute frequency

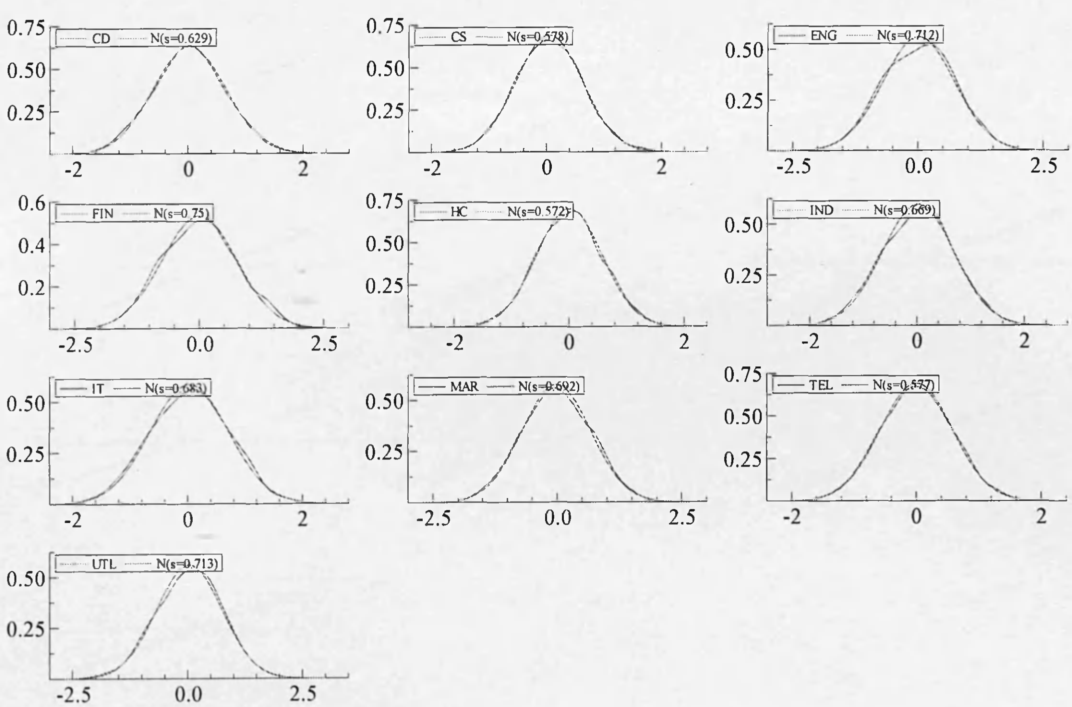


Figure 3.7a: Density plots of sector-averages daily returns standardized by realized power variation sampled at 5-minute frequency

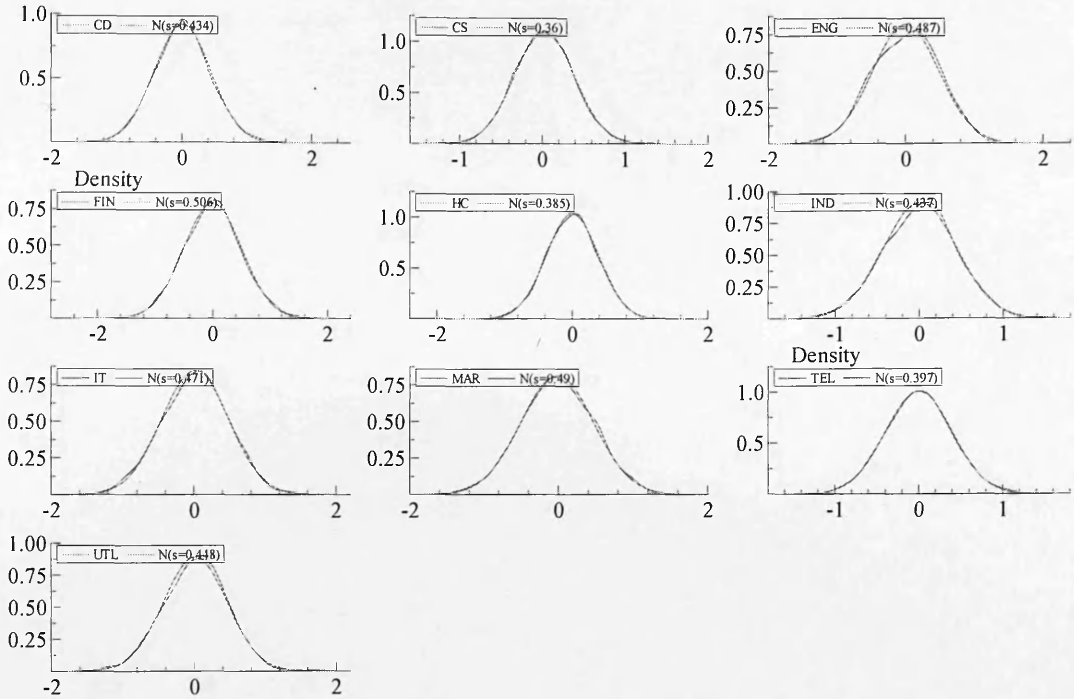


Figure 3.7b: QQ plots of sector-averages daily returns standardized by realized power variation sampled at 5-minute frequency

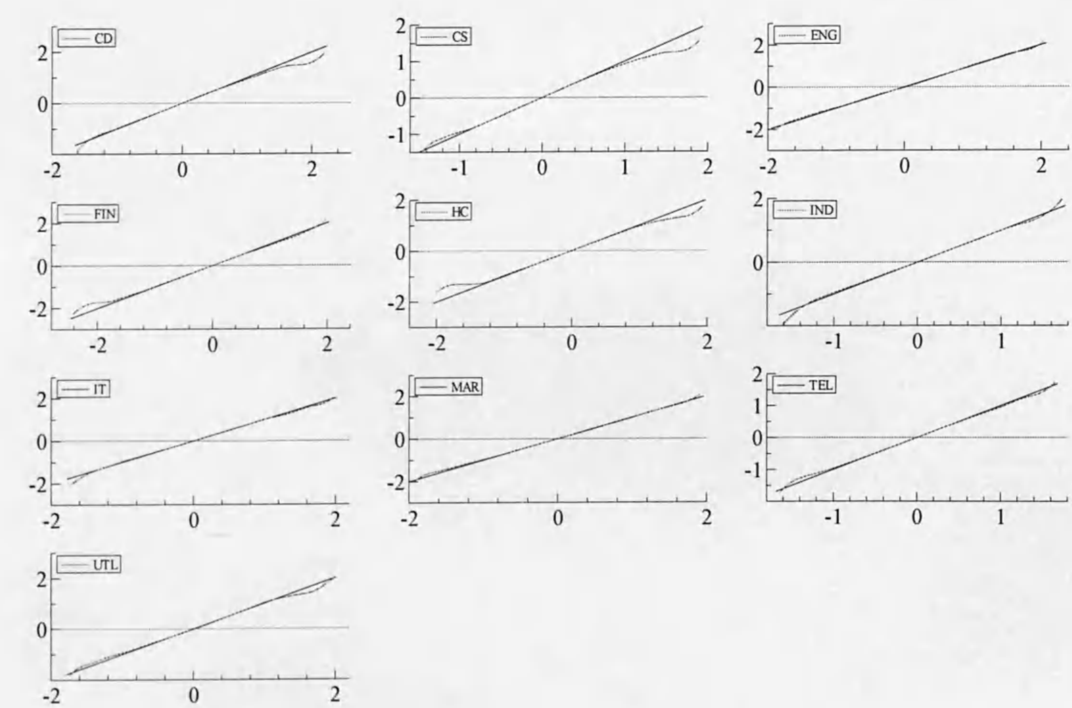


Figure 3.8a: Density plots of sector-averages daily returns standardized by realized range sampled at 5-minute frequency

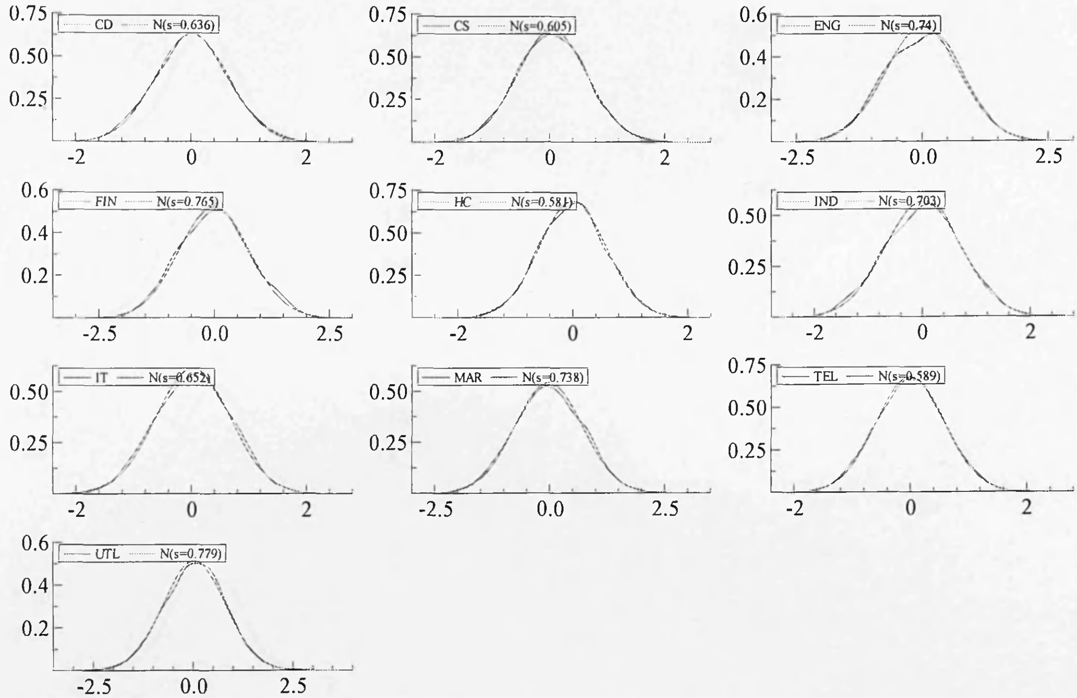


Figure 3.8b: QQ plots of sector-averages daily returns standardized by realized range sampled at 5-minute frequency

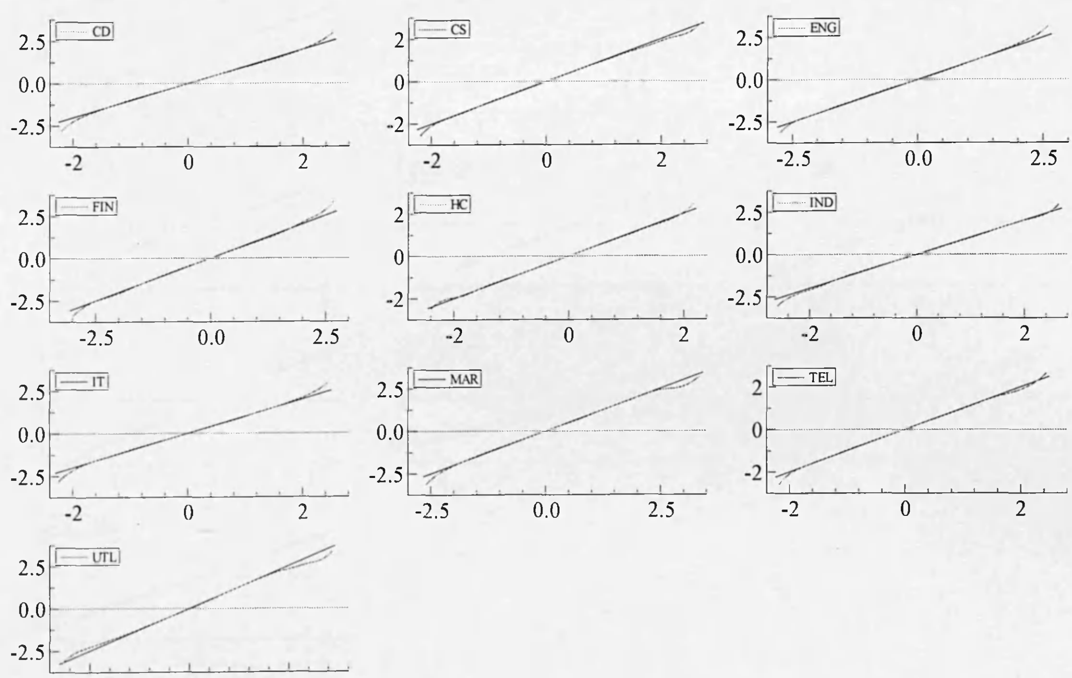


Figure 3.9a: Density plots of sector-averages daily returns standardized by realized variance sampled at 1-second frequency

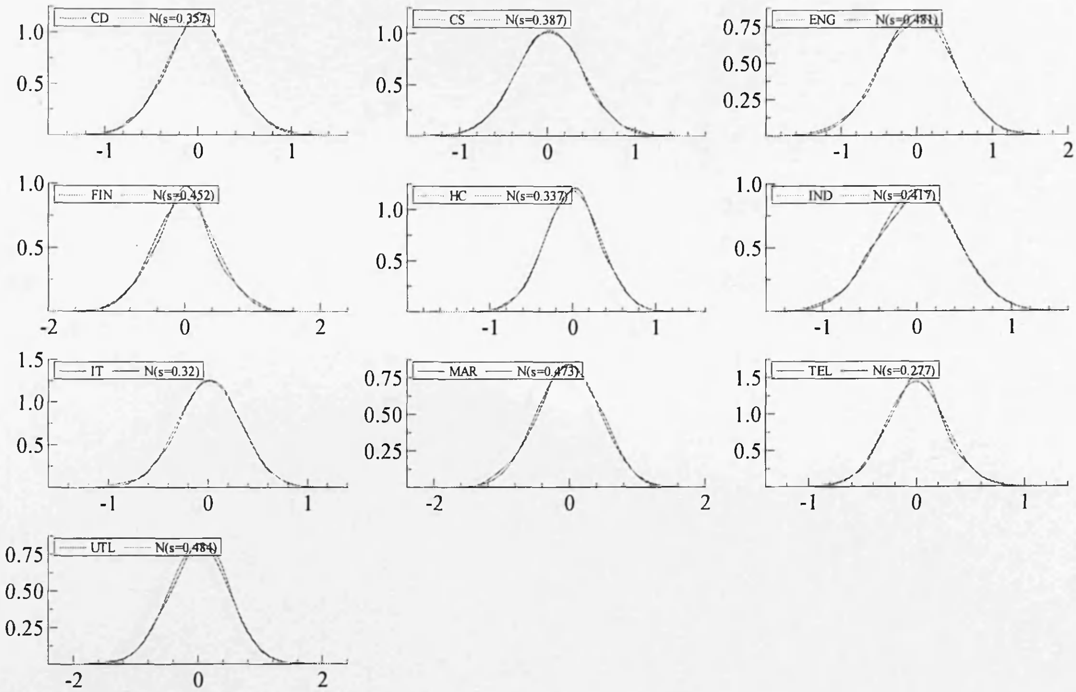


Figure 3.9a: QQ plots of sector-averages daily returns standardized by realized variance sampled at 1-second frequency

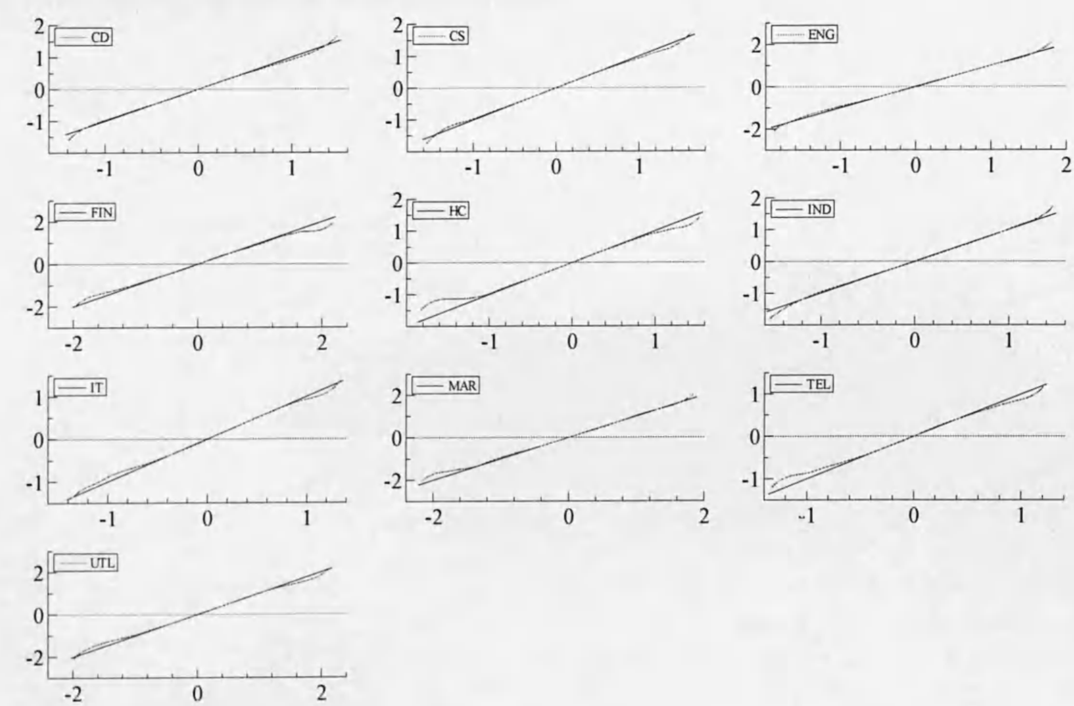


Figure 3.10a: Density plots of sector-averages daily returns standardized by realized bipower variation sampled at 1-second frequency

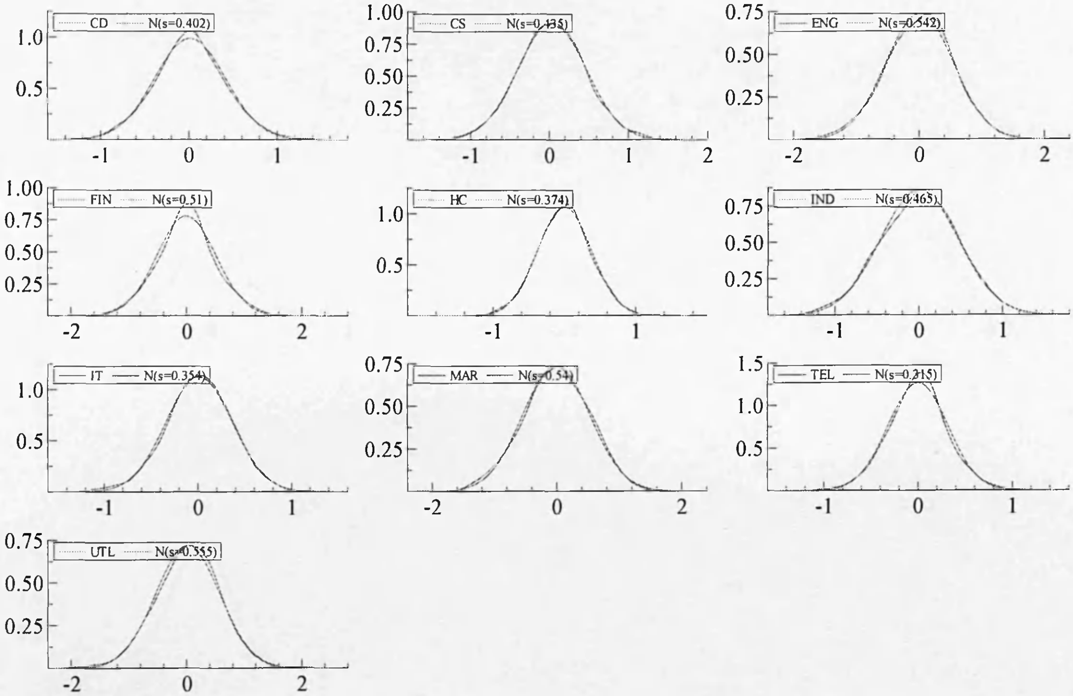


Figure 3.10b: QQ plots of sector-averages daily returns standardized by realized bipower variation sampled at 1-second frequency

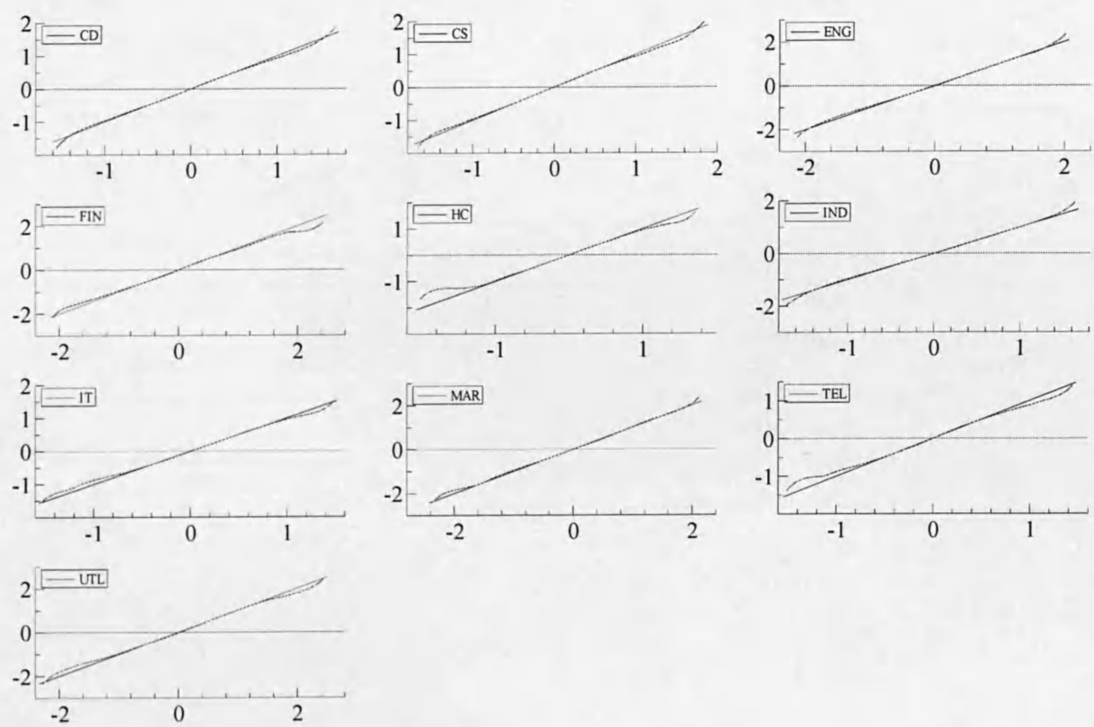


Figure 3.11a: Density plots of sector-averages daily returns standardized by realized power variation sampled at 1-second frequency

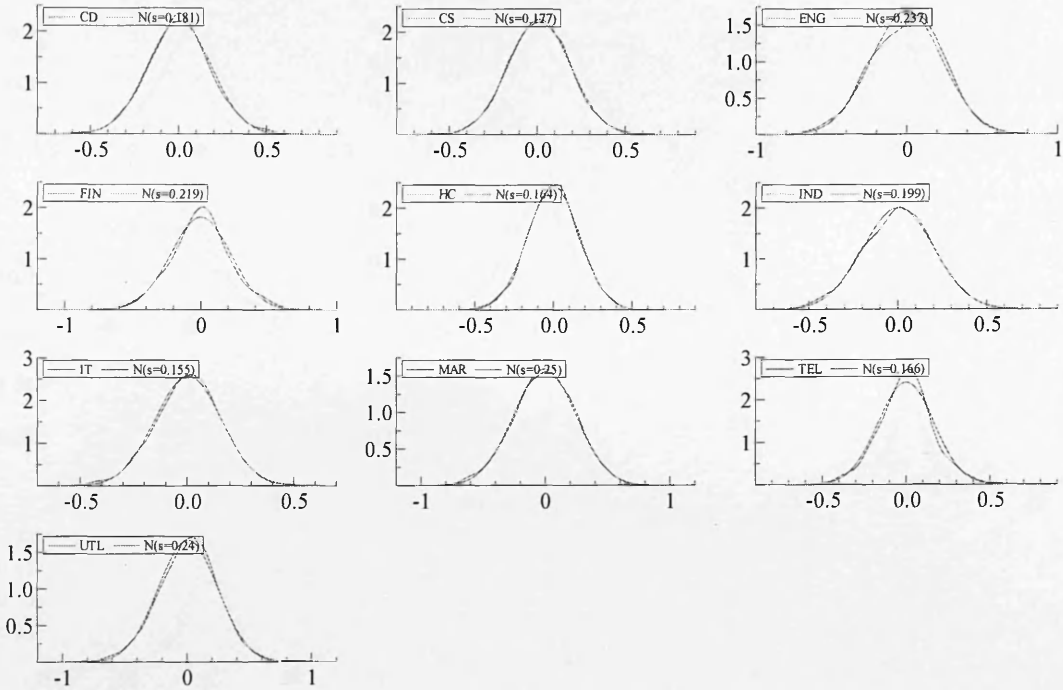


Figure 3.11b: QQ plots of sector-averages daily returns standardized by realized power variation sampled at 1-second frequency

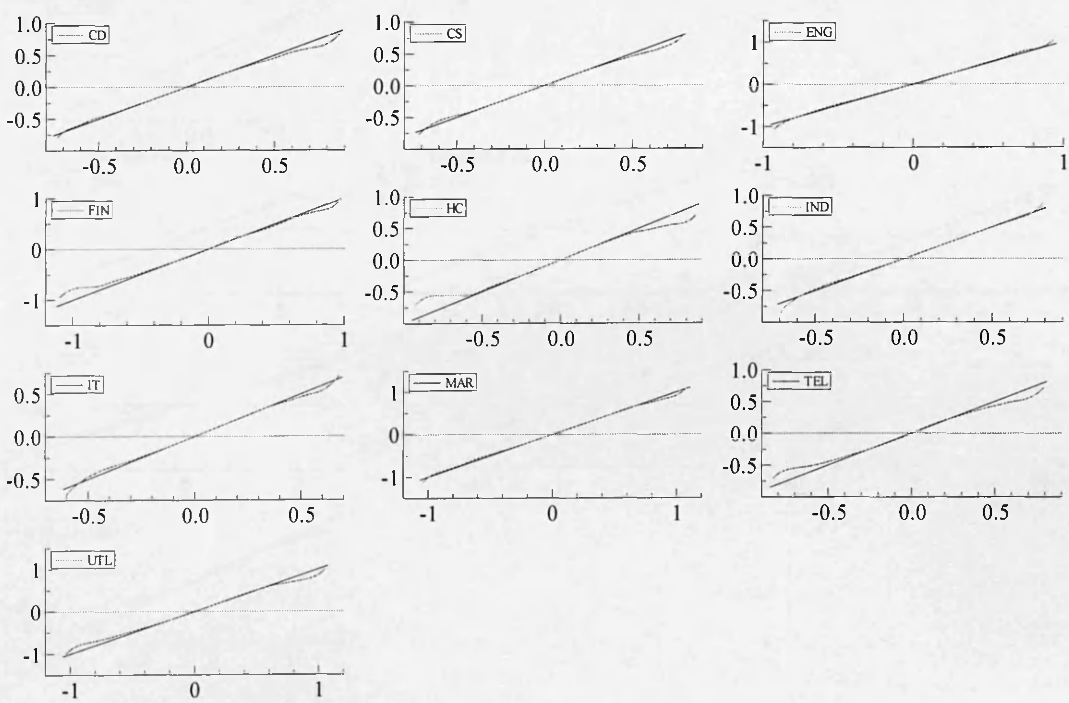


Figure 3.12a: Density plots of sector-averages daily returns standardized by realized range sampled at 1-second frequency

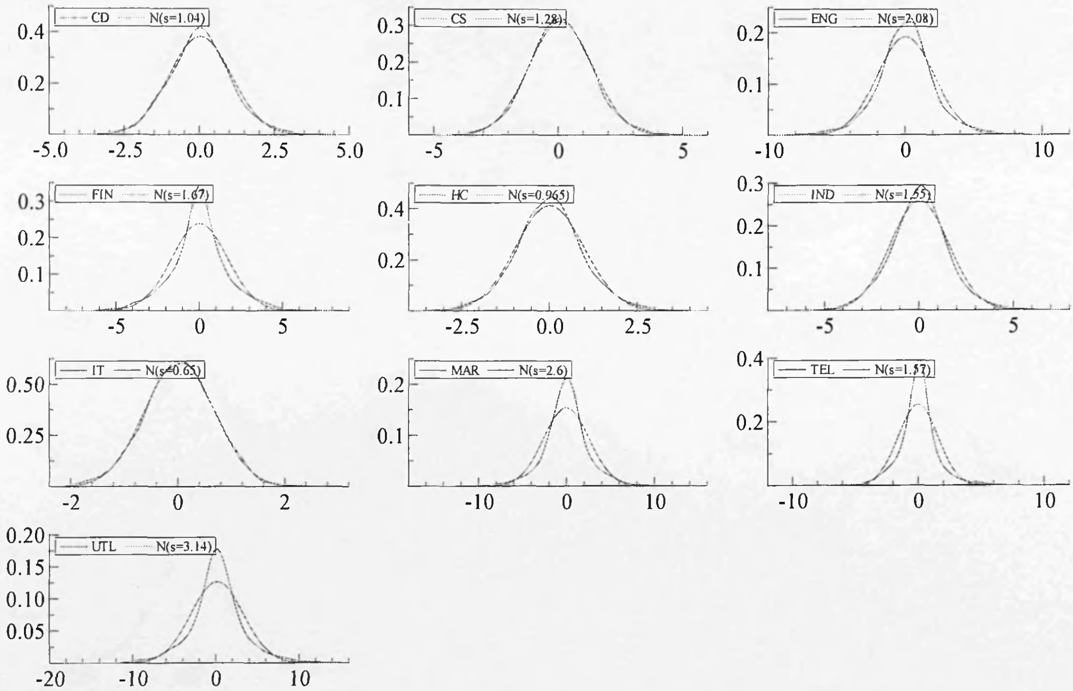


Figure 3.12a: QQ plots of sector-averages daily returns standardized by realized range sampled at 1-second frequency

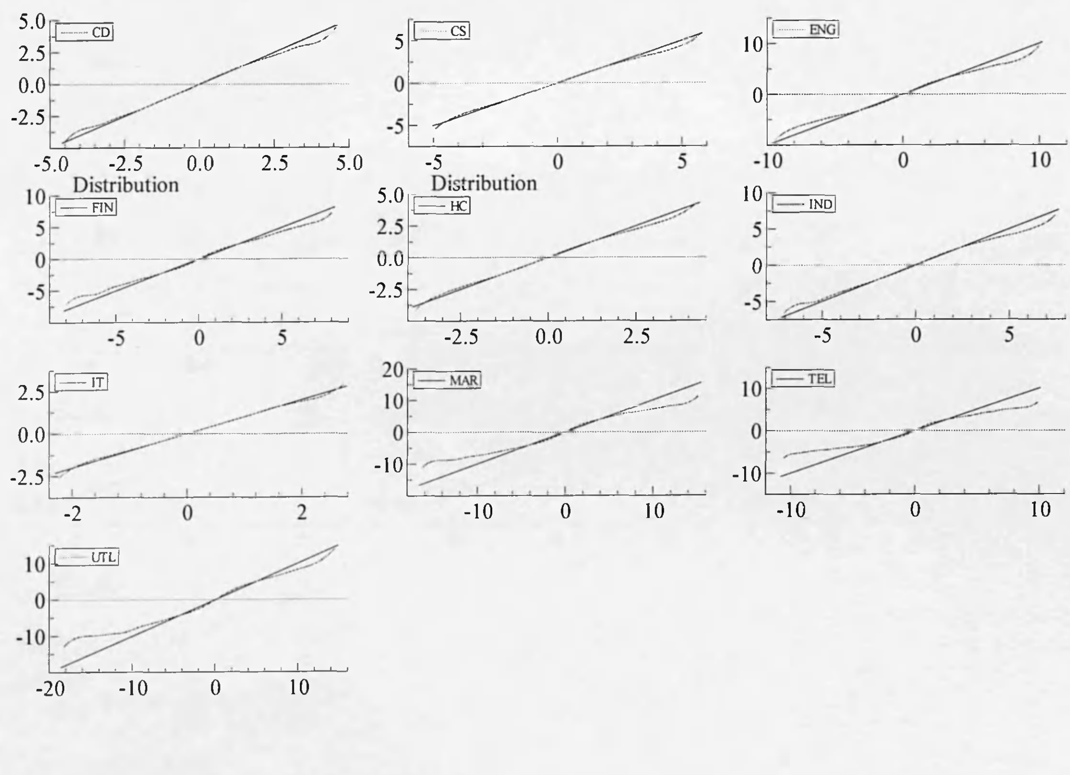


Figure 3.13a: Density plots of sector-averages daily returns standardized by realized continuous variance sampled at 5-minute frequency

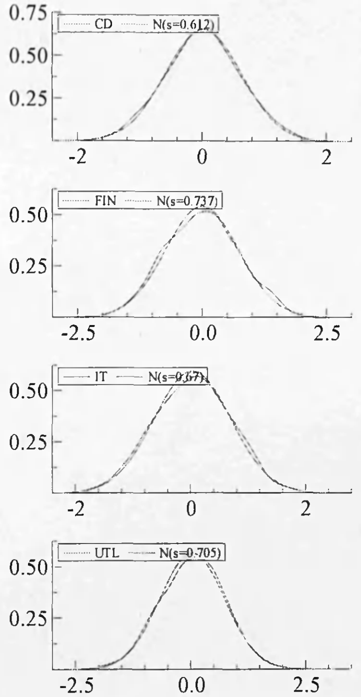


Figure 3.13b: QQ plots of sector-averages daily returns standardized by realized continuous variance sampled at 5-minute frequency

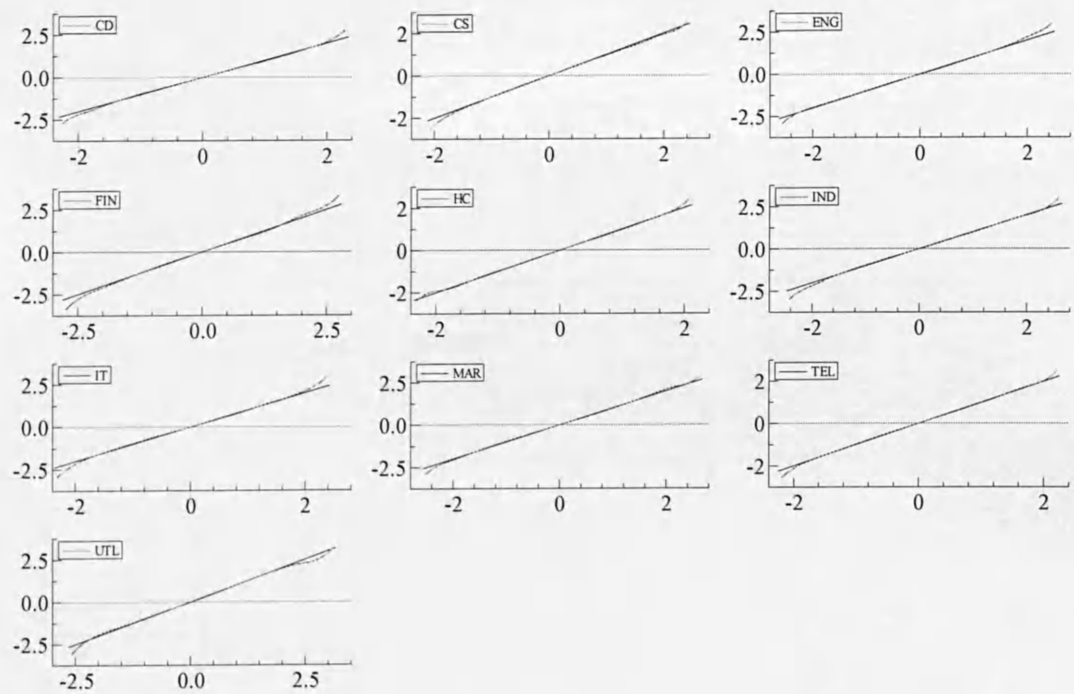


Figure 3.14a: Density plots of sector-averages daily returns standardized by realized continuous variance sampled at 1-second frequency

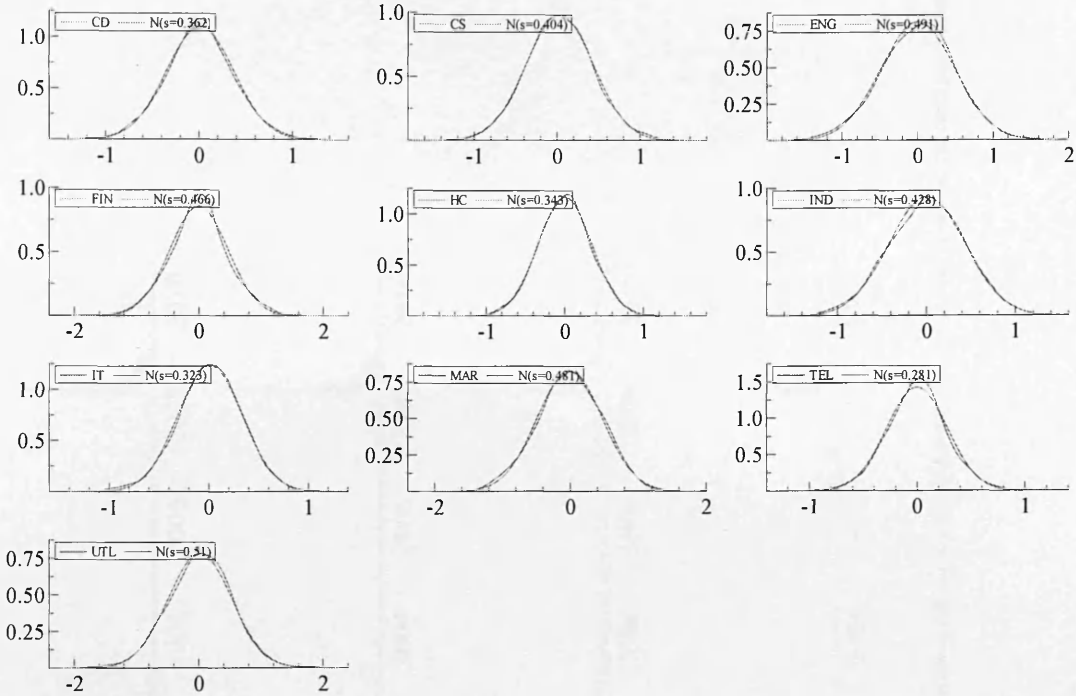


Figure 3.14b: QQ plots of sector-averages daily returns standardized by realized continuous variance sampled at 1-second frequency

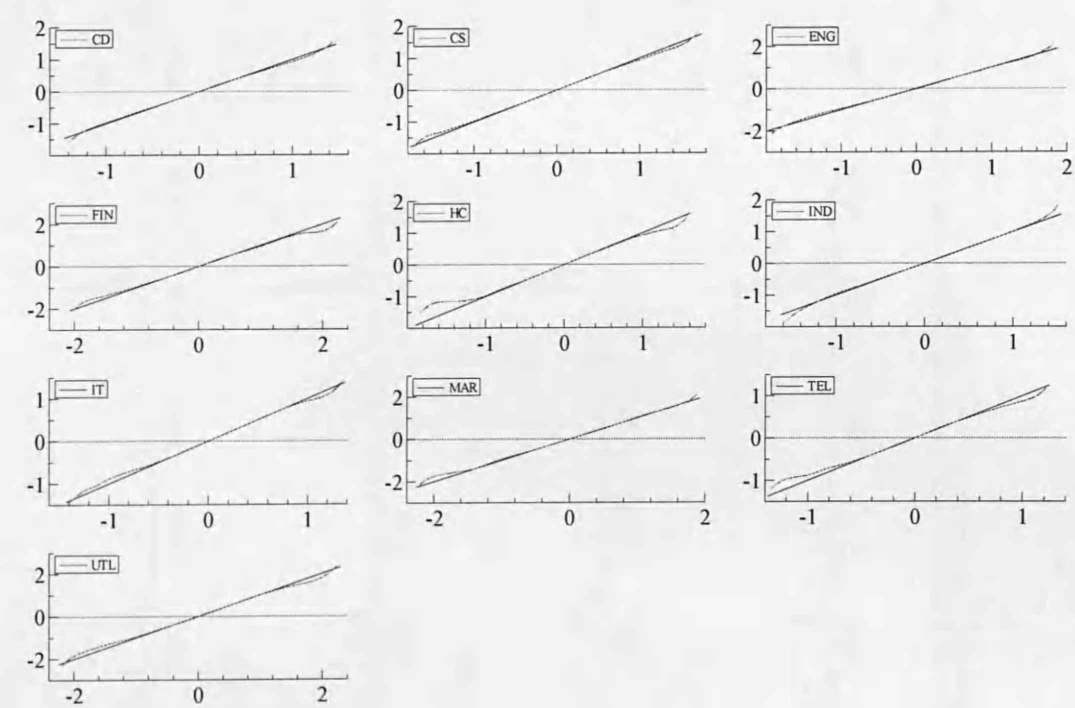
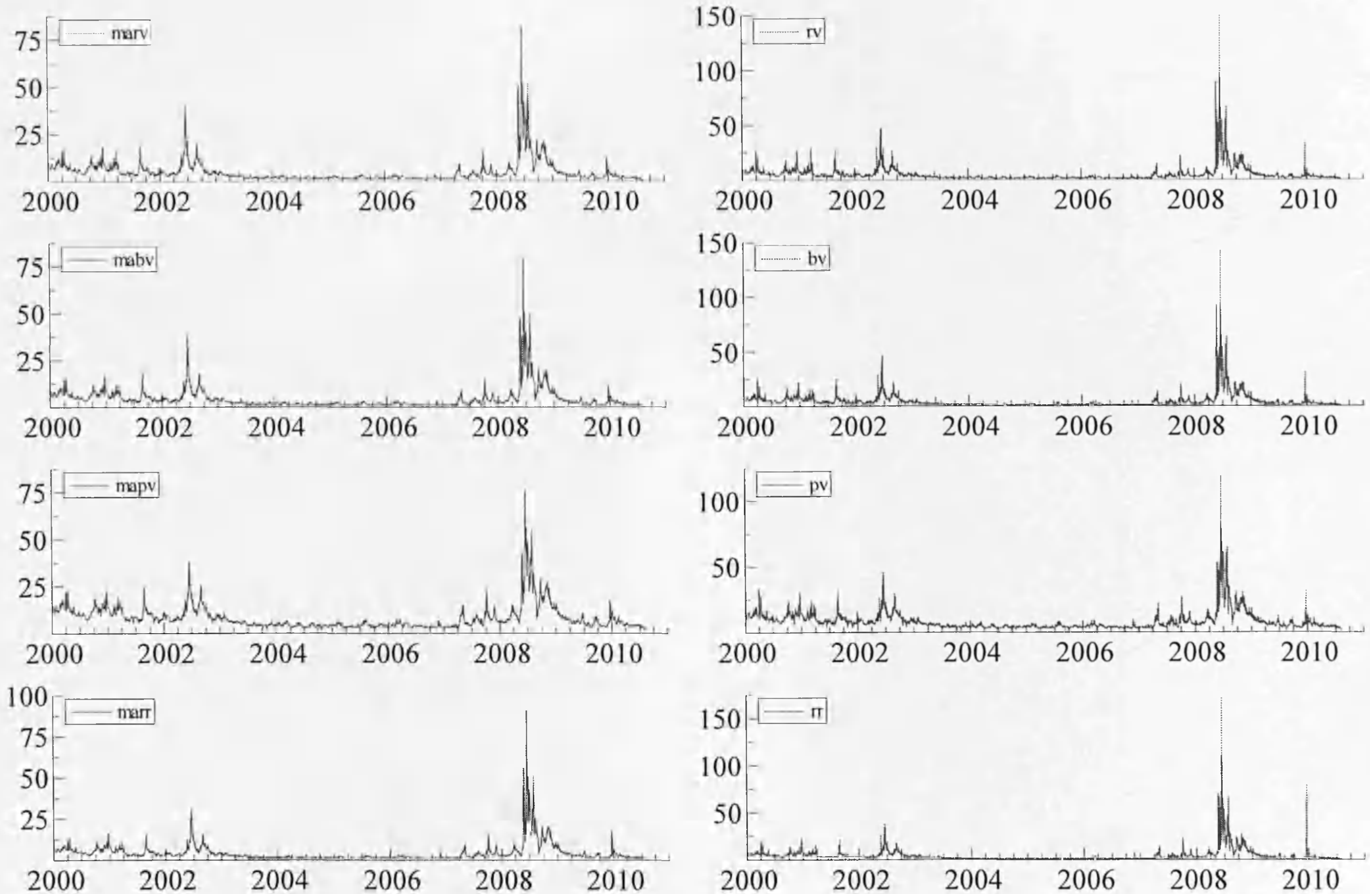


Figure 3.15: Plots of raw and EWMA-filtered realized volatility measures (5-minute frequency)



Note: The figures in the left panel are the realized volatility measures which are filtered by the exponentially weighted moving average (weighting equals 0.5). The figures on the right panel are the raw realized volatility measures. The sampling frequency is 5-minute.

Figure 3.16a: Density plots of sector-averages daily returns standardized by GARCH sampled at 5-minute frequency

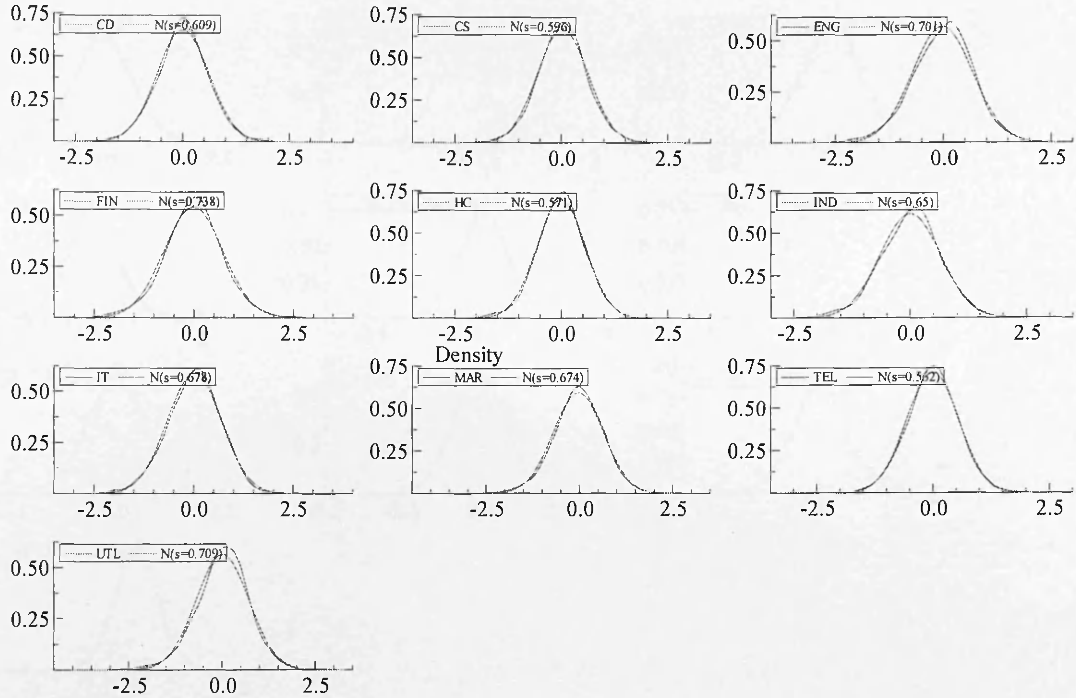


Figure 3.16b: QQ plots of sector-averages daily returns standardized by GARCH sampled at 5-minute frequency

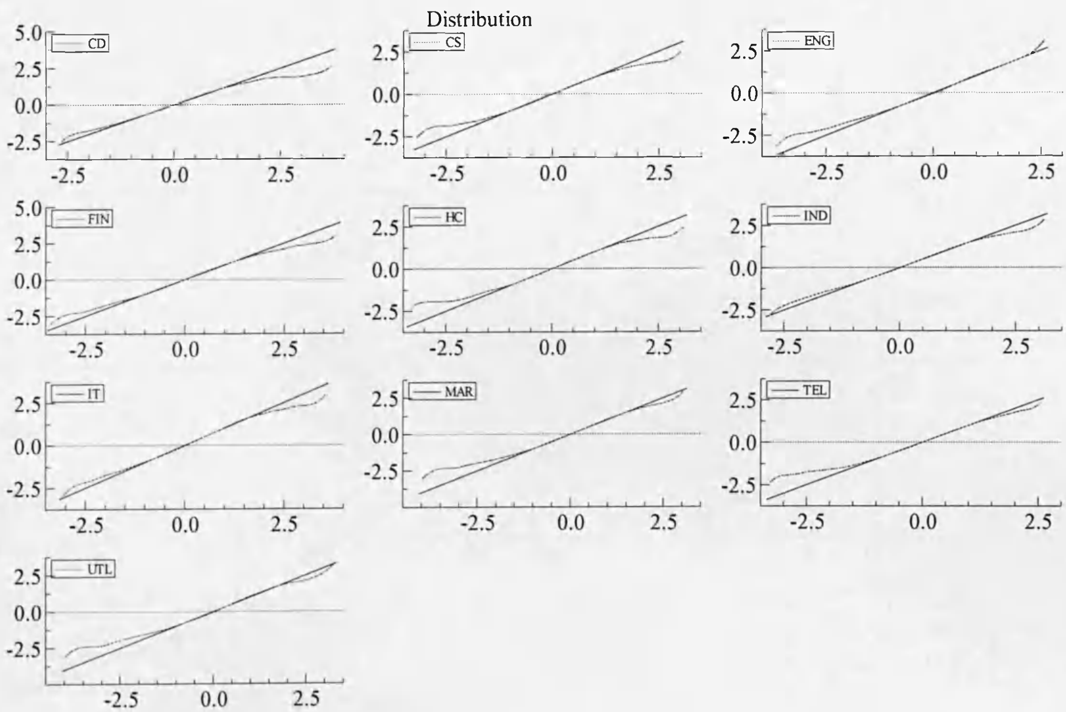


Figure 3.17a: Density plots of sector-averages daily returns standardized by GARCH sampled at 5-minute frequency

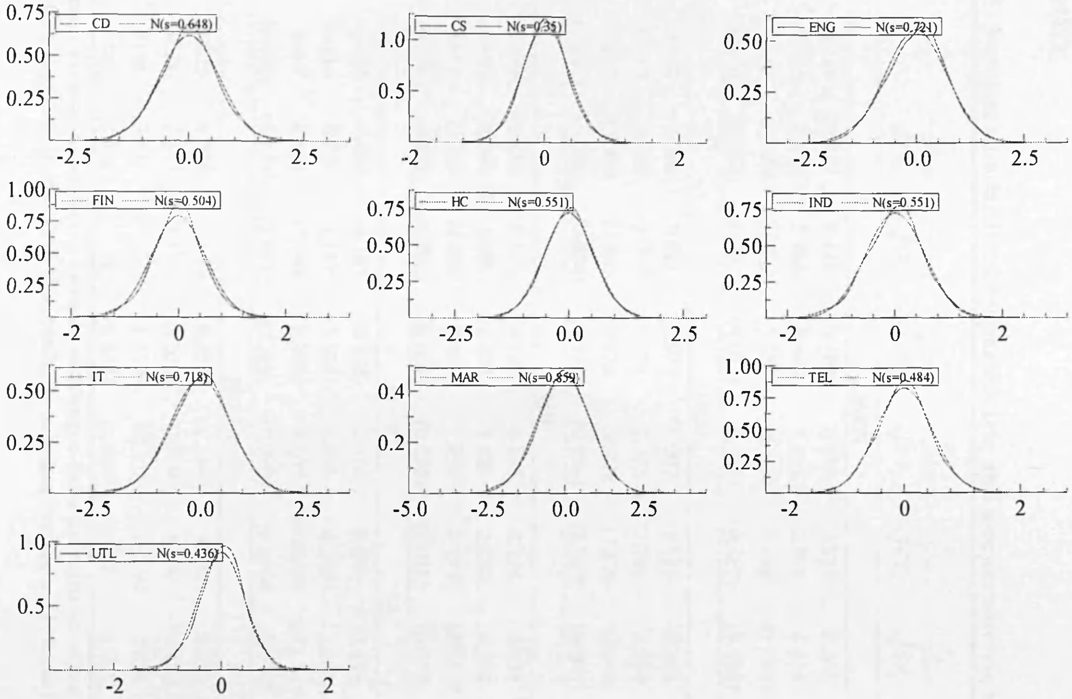
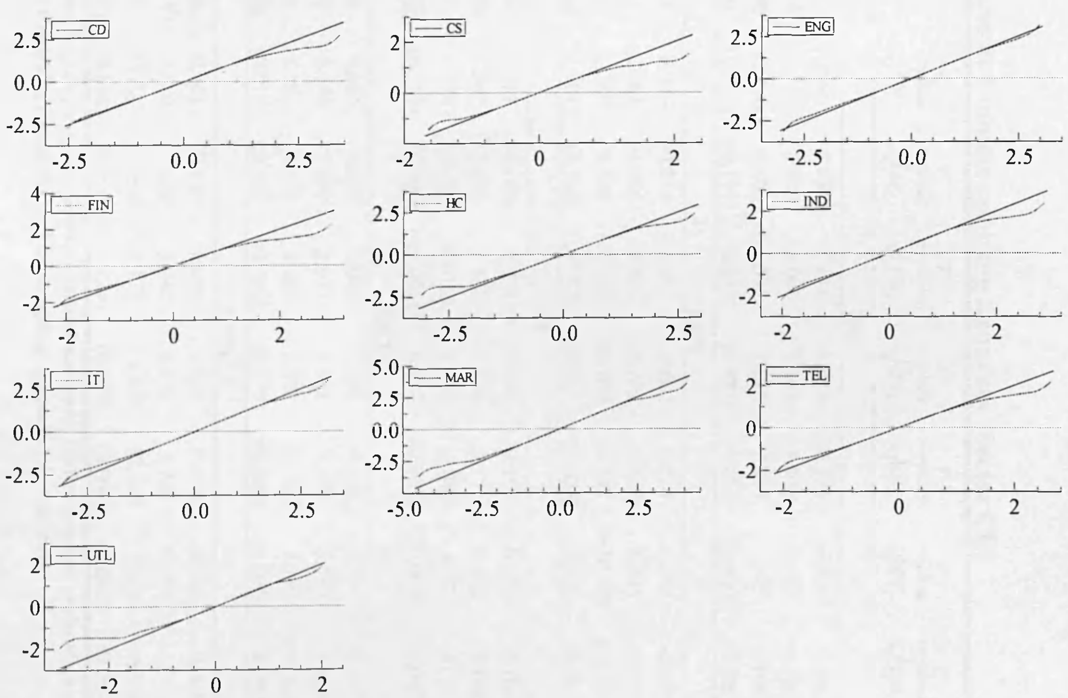


Figure 3.17b: QQ plots of sector-averages daily returns standardized by GARCH sampled at 5-minute frequency



Appendix

Table A.1: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector CD)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	AMZN								BBY							
Skewness	0.411	0.113	0.118	0.107	0.117	0.083	0.261	0.248	-1.485	0.096	0.089	0.113	0.071	0.076	-0.134	-0.085
Kurtosis	9.065	2.775	2.786	3.020	2.819	2.938	4.855	4.878	36.815	2.733	3.080	3.007	2.842	2.806	8.337	6.176
JB	4319	11.704	11.736	5.279	10.129	3.646	427.899	434.959	132846	12.486	4.405	5.891	5.187	7.044	3293	1166
p value	(0.000)	(0.003)	(0.003)	(0.071)	(0.006)	(0.162)	(0.000)	(0.000)	(0.000)	(0.002)	(0.111)	(0.053)	(0.075)	(0.030)	(0.000)	(0.000)
	DIS								GPS							
Skewness	0.224	0.078	0.081	0.074	0.097	0.118	0.113	0.029	0.238	0.005	0.011	0.004	-0.002	0.024	0.087	-0.036
Kurtosis	6.185	2.635	2.735	2.779	2.775	2.786	3.534	3.681	7.294	2.664	2.935	2.920	2.705	2.717	5.295	5.693
JB	1193	18.105	11.065	8.120	10.228	11.676	38.754	53.944	2152	13.035	0.541	0.750	10.044	9.508	610.984	836.909
p value	(0.000)	(0.000)	(0.004)	(0.017)	(0.006)	(0.003)	(0.000)	(0.000)	(0.000)	(0.001)	(0.763)	(0.687)	(0.007)	(0.009)	(0.000)	(0.000)
	HD								IPG							
Skewness	0.576	0.118	0.148	0.135	0.115	0.113	0.293	0.273	0.528	0.083	-0.004	0.074	0.090	0.150	0.131	-0.094
Kurtosis	6.883	2.759	2.880	2.895	2.876	2.870	4.124	4.626	20.576	2.902	3.496	3.325	2.959	3.171	9.121	8.984
JB	1891	13.081	11.830	9.639	7.850	7.799	185.234	339.133	35745	4.324	28.385	14.680	3.940	13.803	4327	4132
p value	(0.000)	(0.001)	(0.003)	(0.008)	(0.020)	(0.020)	(0.000)	(0.000)	(0.000)	(0.115)	(0.000)	(0.001)	(0.139)	(0.001)	(0.000)	(0.000)
	MAR								MCD							
Skewness	0.222	-0.088	-0.107	-0.048	-0.102	0.240	0.193	0.048	0.218	0.065	0.093	0.045	0.086	-0.046	0.199	0.077
Kurtosis	8.810	2.929	3.242	3.107	2.995	6.220	5.154	5.638	9.460	2.766	2.850	2.953	3.021	3.547	6.767	5.455
JB	3915	4.178	12.098	2.399	4.807	#######	552.147	803.518	4833	8.261	6.579	1.200	3.481	35.450	1655	697.810
p value	(0.000)	(0.124)	(0.002)	(0.301)	(0.090)	(0.000)	(0.000)	(0.000)	(0.000)	(0.016)	(0.037)	(0.549)	(0.175)	(0.000)	(0.000)	(0.000)
	NWSA								TWX							
Skewness	0.521	0.039	-0.017	0.058	0.044	-0.221	0.524	0.496	0.208	0.101	0.137	0.092	0.131	0.138	0.085	0.004
Kurtosis	10.781	2.843	4.535	3.028	3.729	4.809	10.078	9.820	9.195	2.670	2.869	3.046	2.779	2.730	6.596	3.917
JB	7106	3.540	271.739	1.619	62.218	396.394	5903	5476	4445	17.225	10.642	4.171	13.510	17.174	1494	96.936
p value	(0.000)	(0.170)	(0.000)	(0.445)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.124)	(0.001)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t, sv_t, ga_t denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.2: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector CS)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	AVP								BFB							
Skewness	0.220	0.071	0.075	0.088	0.078	0.473	-0.021	-0.040	0.315	0.034	0.061	0.051	0.060	0.473	0.245	0.032
Kurtosis	9.275	2.863	3.070	3.124	2.911	6.817	7.188	6.972	10.338	3.358	4.515	3.396	3.424	7.020	7.113	6.387
JB	4562	4.515	3.177	5.370	3.756	1781	2023	1820	6254	15.324	266.453	19.272	22.442	1917.604	#######	#######
p value	(0.000)	(0.105)	(0.204)	(0.068)	(0.153)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	COST								EL							
Skewness	0.336	0.130	0.101	0.091	0.135	0.193	0.142	0.081	0.148	0.086	0.061	0.079	0.063	0.135	0.366	0.287
Kurtosis	7.320	2.770	2.728	2.902	2.801	3.055	3.894	3.640	7.250	2.988	3.188	3.128	2.987	3.451	5.098	5.369
JB	2204	13.932	13.221	4.938	13.044	17.594	101.432	50.234	2093	3.419	5.818	4.805	1.822	31.801	569.407	685.139
p value	(0.000)	(0.001)	(0.001)	(0.085)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.181)	(0.055)	(0.090)	(0.402)	(0.000)	(0.000)	(0.000)
	KMB								KO							
Skewness	-0.036	0.094	0.108	0.075	0.063	43.067	-0.307	-0.345	0.266	0.072	0.098	0.108	0.112	-0.324	0.086	0.034
Kurtosis	10.763	2.904	3.212	3.231	2.974	2115.918	8.612	7.461	8.119	3.081	3.211	3.300	3.146	33.798	4.177	4.303
JB	6949	5.141	10.560	8.739	1.880	5.E+08	3674.1	2349.6	3054	3.170	9.529	15.743	8.191	108455	163.211	196.367
p value	(0.000)	(0.077)	(0.005)	(0.013)	(0.391)	(0.000)	(0.000)	(0.000)	(0.000)	(0.205)	(0.009)	(0.000)	(0.017)	(0.000)	(0.000)	(0.000)
	PEP								PG							
Skewness	0.147	0.022	0.010	0.007	0.046	0.093	0.018	-0.034	-0.229	0.029	0.034	-0.014	0.017	0.660	-0.122	-0.203
Kurtosis	10.082	2.823	3.016	3.058	2.956	4.224	4.743	4.356	10.909	2.942	3.061	3.213	3.085	8.703	5.719	5.787
JB	5793	3.837	0.075	0.414	1.204	175.786	350.502	212.463	7237	0.786	0.970	5.304	0.962	3926	858.997	914.419
p value	(0.000)	(0.147)	(0.963)	(0.813)	(0.548)	(0.000)	(0.000)	(0.000)	(0.000)	(0.675)	(0.616)	(0.071)	(0.618)	(0.000)	(0.000)	(0.000)
	UL								WMT							
Skewness	-0.058	-0.034	-0.054	-0.070	-0.040	-0.520	-0.133	-0.165	0.210	0.091	0.081	0.101	0.071	0.052	0.251	0.141
Kurtosis	8.258	2.902	3.185	3.144	2.989	11.718	6.389	4.310	7.329	2.875	2.945	3.109	2.945	3.534	4.153	3.910
JB	3189	1.639	5.286	4.651	0.745	8656	1332	210.493	2181	5.594	3.393	6.034	2.693	34.056	182.158	104.697
p value	(0.000)	(0.441)	(0.071)	(0.098)	(0.689)	(0.000)	(0.000)	(0.000)	(0.000)	(0.061)	(0.183)	(0.049)	(0.260)	(0.000)	(0.000)	(0.000)

Note: Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, $rr_t, pv_t, bv_t, cv_t, sv_t, ga_t$ denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.3: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector ENG)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	BHI								CHK							
Skewness	-0.074	0.109	0.096	0.086	0.092	0.112	0.052	0.010	-0.478	0.056	0.084	0.023	0.063	0.064	0.486	0.046
Kurtosis	6.806	2.852	2.948	2.920	2.878	2.854	3.359	3.412	15.730	2.807	2.872	3.087	2.841	2.807	7.853	4.483
JB	1673	8.033	4.525	4.144	5.597	8.264	16.075	19.605	18789	5.761	5.174	1.116	4.784	6.187	2825	254.696
p value	(0.000)	(0.018)	(0.104)	(0.126)	(0.061)	(0.016)	(0.000)	(0.000)	(0.000)	(0.056)	(0.075)	(0.572)	(0.091)	(0.045)	(0.000)	(0.000)
	CVX								DVN							
Skewness	0.172	0.037	0.059	0.017	0.067	0.090	-0.071	-0.117	-0.096	-0.026	-0.048	-0.039	-0.005	-0.038	-0.038	-0.130
Kurtosis	13.415	2.780	3.014	2.954	2.870	3.000	3.521	3.678	7.336	2.733	3.415	2.865	2.898	2.800	4.401	4.808
JB	12519	6.226	1.618	0.379	4.015	3.758	33.580	59.246	2172	8.518	20.924	2.788	1.221	5.249	226.900	384.632
p value	(0.000)	(0.044)	(0.445)	(0.827)	(0.134)	(0.153)	(0.000)	(0.000)	(0.000)	(0.014)	(0.000)	(0.248)	(0.543)	(0.072)	(0.000)	(0.000)
	HAL								OXY							
Skewness	-2.764	0.033	0.035	-0.027	0.047	0.036	-0.251	-1.555	-0.393	-0.007	-0.041	-0.062	0.014	0.051	-0.197	-0.166
Kurtosis	61.460	2.794	3.042	3.164	2.782	2.782	12.806	27.633	8.669	2.842	2.963	2.846	2.863	2.994	4.038	3.560
JB	397534	5.415	0.773	3.442	6.496	6.050	11115	71073	3776	2.897	0.932	4.489	2.260	1.225	142.009	48.930
p value	(0.000)	(0.067)	(0.679)	(0.179)	(0.039)	(0.049)	(0.000)	(0.000)	(0.000)	(0.235)	(0.628)	(0.106)	(0.323)	(0.542)	(0.000)	(0.000)
	SUN								TE							
Skewness	0.153	0.027	0.037	0.019	0.036	0.021	0.073	0.050	-1.353	-0.069	-0.082	-0.101	-0.038	0.064	0.038	-0.242
Kurtosis	8.655	2.944	3.031	3.065	2.988	2.982	4.011	4.077	26.572	2.902	3.050	3.189	3.030	2.807	5.612	5.569
JB	3698	0.689	0.751	0.655	0.630	0.236	120.324	134.967	64904	3.278	3.413	8.806	0.779	6.187	786.974	787.969
p value	(0.000)	(0.709)	(0.687)	(0.721)	(0.730)	(0.889)	(0.000)	(0.000)	(0.000)	(0.194)	(0.181)	(0.012)	(0.677)	(0.045)	(0.000)	(0.000)
	WMB								XOM							
Skewness	-2.338	-0.001	-0.018	-0.072	0.008	1.114	-0.154	-0.384	0.197	0.061	0.097	0.038	0.067	0.005	-0.068	-0.124
Kurtosis	46.957	2.743	2.811	3.215	2.802	46.400	6.544	6.115	12.204	2.721	2.924	2.893	2.823	2.733	3.413	4.404
JB	225286	7.637	4.258	7.707	4.564	215917	1459.125	1186.785	9784	10.689	4.968	1.977	5.653	8.203	21.786	234.452
p value	(0.000)	(0.022)	(0.119)	(0.021)	(0.102)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.083)	(0.372)	(0.059)	(0.017)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, $rr_t, pv_t, bv_t, cv_t, sv_t, ga_t$ denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.4: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector FIN)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	ALL								AXP							
Skewness	-0.033	0.065	0.074	0.093	0.083	1.001	0.116	0.093	0.057	0.050	0.058	0.022	0.057	2.019	-0.035	0.093
Kurtosis	21.060	2.861	2.945	3.396	2.904	15.610	8.698	5.956	7.956	2.751	2.896	2.913	2.826	33.776	4.313	5.956
JB	37603	4.177	2.890	22.029	4.274	18735	3749.848	1011.319	2834	8.272	2.831	1.091	5.001	110641	199.458	1011
p value	(0.000)	(0.124)	(0.236)	(0.000)	(0.118)	(0.000)	(0.000)	(0.000)	(0.000)	(0.016)	(0.243)	(0.580)	(0.082)	(0.000)	(0.000)	(0.000)
	BAC								BK							
Skewness	-0.508	0.016	-0.014	-0.036	0.004	0.114	-0.121	-0.414	0.348	0.095	0.045	0.123	0.073	0.102	0.141	-0.074
Kurtosis	20.028	2.648	2.841	3.356	2.727	3.080	7.486	5.504	13.236	2.930	3.202	3.202	2.960	2.940	6.407	5.296
JB	33547	14.376	2.989	15.208	8.608	6.685	2327.164	801.620	12135	4.691	5.610	11.692	2.611	5.159	1348	610.127
p value	(0.000)	(0.001)	(0.224)	(0.000)	(0.014)	(0.035)	(0.000)	(0.000)	(0.000)	(0.096)	(0.061)	(0.003)	(0.271)	(0.076)	(0.000)	(0.000)
	C								GS							
Skewness	-1.814	0.154	0.175	0.052	0.153	0.282	-0.285	0.080	-0.098	0.041	0.019	0.030	0.033	0.065	-0.174	-0.431
Kurtosis	28.052	2.731	2.908	3.158	2.830	3.608	7.908	4.448	9.978	2.691	2.950	2.888	2.809	2.822	5.222	8.174
JB	73877	19.295	15.031	4.160	14.124	79.261	2814.873	244.553	5618	11.808	0.456	1.868	4.694	5.586	583.274	3172
p value	(0.000)	(0.000)	(0.001)	(0.125)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.796)	(0.393)	(0.096)	(0.061)	(0.000)	(0.000)
	JPM								MS							
Skewness	0.529	0.122	0.100	0.151	0.109	0.145	0.410	-0.021	-0.845	0.035	0.042	0.042	0.057	0.035	0.248	0.079
Kurtosis	13.747	2.733	2.835	3.186	2.753	2.862	7.463	4.728	22.517	2.647	2.816	2.886	2.687	2.648	4.540	4.263
JB	13445	15.050	7.710	14.482	12.541	11.947	2374.111	344.404	44247	14.901	4.705	2.323	12.768	14.843	302.049	186.637
p value	(0.000)	(0.001)	(0.021)	(0.001)	(0.002)	(0.003)	(0.000)	(0.000)	(0.000)	(0.001)	(0.095)	(0.313)	(0.002)	(0.001)	(0.000)	(0.000)
	TRV								WFC							
Skewness	0.364	0.053	0.066	0.075	0.038	0.192	0.340	0.262	0.086	0.122	0.116	0.105	0.126	0.325	0.247	0.213
Kurtosis	9.708	2.755	3.113	3.017	2.833	9.468	7.265	7.460	18.265	2.887	3.013	3.624	2.929	4.561	7.136	4.782
JB	5249	8.171	3.474	2.601	3.915	4828	2150.616	2324.972	26870	8.321	6.197	50.008	7.869	327.535	2000	386.910
p value	(0.000)	(0.017)	(0.176)	(0.272)	(0.141)	(0.000)	(0.000)	(0.000)	(0.000)	(0.016)	(0.045)	(0.000)	(0.020)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t, sv_t, ga_t denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.5: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector HC)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	ABT								AMGN							
Skewness	0.001	0.028	0.065	0.026	0.043	0.139	0.074	-0.021	-0.273	0.002	-0.023	-0.046	0.007	0.022	-0.079	-0.260
Kurtosis	6.479	2.885	3.149	3.087	2.996	3.676	4.696	5.300	6.640	2.690	2.843	2.961	2.732	2.717	4.486	4.842
JB	1396	1.889	4.517	1.188	0.851	61.440	334.041	610.216	1562	11.105	3.081	1.139	8.282	9.441	257.300	422.397
p value	(0.000)	(0.389)	(0.105)	(0.552)	(0.654)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.214)	(0.566)	(0.016)	(0.009)	(0.000)	(0.000)
	BSX								GILD							
Skewness	0.119	0.121	0.132	0.118	0.258	-8.996	0.045	-0.003	0.009	0.046	0.001	0.025	0.032	0.047	-0.054	-0.066
Kurtosis	7.807	2.685	2.887	2.817	3.752	267.642	4.800	5.504	7.496	2.747	2.823	3.068	2.813	2.745	3.770	3.835
JB	2670	18.166	9.560	10.239	95.865	8108857	374.319	723.021	2328	8.353	3.615	0.806	4.496	8.491	69.686	82.266
p value	(0.000)	(0.000)	(0.008)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.015)	(0.164)	(0.668)	(0.106)	(0.014)	(0.000)	(0.000)
	HUM								JNJ							
Skewness	-0.178	0.130	0.123	0.095	0.163	0.154	0.055	-0.032	0.085	0.042	0.054	0.026	0.061	0.774	0.107	0.016
Kurtosis	7.384	2.865	2.929	2.891	2.916	2.938	5.290	5.260	8.140	2.851	2.978	3.137	2.890	34.355	4.141	4.406
JB	2231	9.855	7.591	5.543	13.039	11.398	606.091	589.423	3049	3.373	1.417	2.473	3.104	112516	155.277	227.968
p value	(0.000)	(0.007)	(0.022)	(0.063)	(0.001)	(0.003)	(0.000)	(0.000)	(0.000)	(0.185)	(0.492)	(0.290)	(0.212)	(0.000)	(0.000)	(0.000)
	MDT								MRK							
Skewness	-0.150	0.035	0.086	-0.005	0.018	0.016	0.062	0.029	0.054	0.065	0.039	0.060	0.044	0.077	-0.311	-0.541
Kurtosis	6.367	2.809	3.274	2.952	3.087	2.996	4.300	4.180	8.172	2.790	2.868	3.029	2.896	3.461	7.507	7.944
JB	1318	4.760	12.070	0.280	1.029	0.117	196.462	160.810	3086	7.076	2.711	1.750	2.157	27.164	2387	2953
p value	(0.000)	(0.093)	(0.002)	(0.870)	(0.598)	(0.943)	(0.000)	(0.000)	(0.000)	(0.029)	(0.258)	(0.417)	(0.340)	(0.000)	(0.000)	(0.000)
	PFE								UNH							
Skewness	0.177	0.052	0.080	0.071	0.063	0.017	0.143	0.100	0.453	0.017	0.039	0.016	0.017	0.046	-0.269	-0.223
Kurtosis	5.049	2.683	2.801	2.828	2.872	2.796	3.939	4.292	16.458	2.786	3.027	3.048	2.810	2.803	4.956	4.911
JB	498.350	12.802	7.459	5.729	3.736	4.907	111.103	197.091	20977	5.398	0.801	0.389	4.277	5.410	474.363	443.897
p value	(0.000)	(0.002)	(0.024)	(0.057)	(0.154)	(0.086)	(0.000)	(0.000)	(0.000)	(0.067)	(0.670)	(0.823)	(0.118)	(0.067)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t, sv_t, ga_t denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.6: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector IND)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
					BA								CAT			
Skewness	0.045	0.047	0.043	0.035	0.048	0.072	0.042	-0.025	-0.018	0.019	0.006	0.015	0.030	0.051	0.139	0.025
Kurtosis	5.264	2.763	2.810	2.812	2.763	2.797	3.748	3.708	6.160	2.698	2.995	2.811	2.872	2.782	4.325	4.049
JB	591.652	7.535	5.008	4.663	7.562	7.169	65.293	58.019	1151.600	10.698	0.022	4.215	2.298	6.673	211.365	127.092
p value	(0.000)	(0.023)	(0.082)	(0.097)	(0.023)	(0.028)	(0.000)	(0.000)	(0.000)	(0.005)	(0.989)	(0.122)	(0.317)	(0.036)	(0.000)	(0.000)
					CMI								GD			
Skewness	-0.188	0.045	-0.001	0.051	0.039	0.016	0.108	0.162	-0.198	0.031	0.051	-0.003	0.045	0.037	-0.160	-0.037
Kurtosis	7.699	2.762	3.162	2.856	2.863	2.840	4.628	4.662	6.900	2.815	3.389	2.973	3.020	5.876	5.502	4.994
JB	2562	7.502	3.013	3.595	2.858	3.065	310.761	330.429	1771.740	4.381	18.668	0.088	0.987	951.048	733.614	458.948
p value	(0.000)	(0.023)	(0.222)	(0.166)	(0.240)	(0.216)	(0.000)	(0.000)	(0.000)	(0.112)	(0.000)	(0.957)	(0.610)	(0.000)	(0.000)	(0.000)
					GE								HON			
Skewness	-0.092	0.128	0.124	0.100	0.133	0.046	0.115	0.193	-0.433	0.038	0.041	0.011	0.077	0.058	-0.511	-0.366
Kurtosis	9.363	2.673	2.724	2.873	2.685	2.880	4.927	4.706	10.093	2.842	2.965	2.996	2.931	2.932	10.998	6.112
JB	4671	19.832	15.852	6.423	19.559	2.660	434.181	352.748	5886.503	3.539	0.929	0.058	3.270	2.063	7496	1179
p value	(0.000)	(0.000)	(0.000)	(0.040)	(0.000)	(0.264)	(0.000)	(0.000)	(0.000)	(0.170)	(0.629)	(0.972)	(0.195)	(0.357)	(0.000)	(0.000)
					LUV								MMM			
Skewness	0.105	0.164	0.152	0.171	0.179	0.176	0.473	0.495	0.300	0.046	0.032	0.057	0.029	0.005	0.047	-0.244
Kurtosis	6.992	3.010	3.158	3.100	3.094	3.047	5.654	6.554	7.273	2.935	2.998	3.167	2.983	3.897	5.835	6.610
JB	1842	12.457	13.526	14.580	15.825	14.599	915.584	1569	2146.380	1.449	0.466	4.712	0.428	92.414	927.749	1530
p value	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.485)	(0.792)	(0.095)	(0.807)	(0.000)	(0.000)	(0.000)
					UPS								UTX			
Skewness	0.184	-0.020	-0.055	-0.009	-0.038	0.226	0.306	0.064	-0.234	0.049	0.037	0.031	0.074	0.090	-0.168	-0.207
Kurtosis	7.262	3.052	3.348	3.329	3.247	14.332	5.369	4.977	9.330	2.772	3.023	3.003	2.852	2.906	5.439	5.986
JB	2109	0.507	15.387	12.552	7.710	14728.120	690.414	452.429	4644.655	7.099	0.697	0.442	5.043	4.756	699.096	1048
p value	(0.000)	(0.776)	(0.000)	(0.002)	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)	(0.029)	(0.706)	(0.802)	(0.080)	(0.093)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t, sv_t, ga_t denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.6: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector IND)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	BA								CAT							
Skewness	0.045	0.047	0.043	0.035	0.048	0.072	0.042	-0.025	-0.018	0.019	0.006	0.015	0.030	0.051	0.139	0.025
Kurtosis	5.264	2.763	2.810	2.812	2.763	2.797	3.748	3.708	6.160	2.698	2.995	2.811	2.872	2.782	4.325	4.049
JB	591.652	7.535	5.008	4.663	7.562	7.169	65.293	58.019	1151.600	10.698	0.022	4.215	2.298	6.673	211.365	127.092
p value	(0.000)	(0.023)	(0.082)	(0.097)	(0.023)	(0.028)	(0.000)	(0.000)	(0.000)	(0.005)	(0.989)	(0.122)	(0.317)	(0.036)	(0.000)	(0.000)
	CMI								GD							
Skewness	-0.188	0.045	-0.001	0.051	0.039	0.016	0.108	0.162	-0.198	0.031	0.051	-0.003	0.045	0.037	-0.160	-0.037
Kurtosis	7.699	2.762	3.162	2.856	2.863	2.840	4.628	4.662	6.900	2.815	3.389	2.973	3.020	5.876	5.502	4.994
JB	2562	7.502	3.013	3.595	2.858	3.065	310.761	330.429	1771.740	4.381	18.668	0.088	0.987	951.048	733.614	458.948
p value	(0.000)	(0.023)	(0.222)	(0.166)	(0.240)	(0.216)	(0.000)	(0.000)	(0.000)	(0.112)	(0.000)	(0.957)	(0.610)	(0.000)	(0.000)	(0.000)
	GE								HON							
Skewness	-0.092	0.128	0.124	0.100	0.133	0.046	0.115	0.193	-0.433	0.038	0.041	0.011	0.077	0.058	-0.511	-0.366
Kurtosis	9.363	2.673	2.724	2.873	2.685	2.880	4.927	4.706	10.093	2.842	2.965	2.996	2.931	2.932	10.998	6.112
JB	4671	19.832	15.852	6.423	19.559	2.660	434.181	352.748	5886.503	3.539	0.929	0.058	3.270	2.063	7496	1179
p value	(0.000)	(0.000)	(0.000)	(0.040)	(0.000)	(0.264)	(0.000)	(0.000)	(0.000)	(0.170)	(0.629)	(0.972)	(0.195)	(0.357)	(0.000)	(0.000)
	LUV								MMM							
Skewness	0.105	0.164	0.152	0.171	0.179	0.176	0.473	0.495	0.300	0.046	0.032	0.057	0.029	0.005	0.047	-0.244
Kurtosis	6.992	3.010	3.158	3.100	3.094	3.047	5.654	6.554	7.273	2.935	2.998	3.167	2.983	3.897	5.835	6.610
JB	1842	12.457	13.526	14.580	15.825	14.599	915.584	1569	2146.380	1.449	0.466	4.712	0.428	92.414	927.749	1530
p value	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.485)	(0.792)	(0.095)	(0.807)	(0.000)	(0.000)	(0.000)
	UPS								UTX							
Skewness	0.184	-0.020	-0.055	-0.009	-0.038	0.226	0.306	0.064	-0.234	0.049	0.037	0.031	0.074	0.090	-0.168	-0.207
Kurtosis	7.262	3.052	3.348	3.329	3.247	14.332	5.369	4.977	9.330	2.772	3.023	3.003	2.852	2.906	5.439	5.986
JB	2109	0.507	15.387	12.552	7.710	14728	690.414	452.429	4644.655	7.099	0.697	0.442	5.043	4.756	699.096	1048
p value	(0.000)	(0.776)	(0.000)	(0.002)	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)	(0.029)	(0.706)	(0.802)	(0.080)	(0.093)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, $rr_t, pv_t, bv_t, cv_t, sv_t, ga_t$ denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.7: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector IT)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	AAPL								CSCO							
Skewness	0.037	0.124	0.130	0.092	0.142	0.153	0.042	0.019	0.391	0.051	0.066	0.048	0.044	0.053	0.186	0.051
Kurtosis	5.110	2.739	2.757	2.777	2.793	2.819	3.876	4.163	11.487	2.748	2.837	3.118	2.775	2.740	6.015	4.146
JB	513.941	15.009	14.621	9.642	14.270	14.592	89.249	156.131	8374.472	8.538	5.077	2.664	6.724	9.045	1064	152.752
p value	(0.000)	(0.001)	(0.001)	(0.008)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.014)	(0.079)	(0.264)	(0.035)	(0.011)	(0.000)	(0.000)
	DELL								EMC							
Skewness	0.181	0.048	0.038	0.075	0.070	-0.118	0.063	-0.043	0.196	0.061	0.053	0.035	0.073	0.064	0.124	0.075
Kurtosis	7.309	2.857	2.870	3.182	2.897	4.085	4.068	3.957	8.054	2.718	2.783	2.958	2.752	2.704	4.423	4.140
JB	2156	3.409	2.622	6.410	3.467	142.030	133.436	106.486	2962.432	10.866	6.715	0.767	9.511	11.976	240.511	152.458
p value	(0.000)	(0.182)	(0.270)	(0.041)	(0.177)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.035)	(0.681)	(0.009)	(0.003)	(0.000)	(0.000)
	HPQ								IBM							
Skewness	0.294	0.050	0.041	0.006	0.047	0.372	0.137	0.100	0.060	0.022	0.020	0.005	0.050	0.240	0.106	-0.089
Kurtosis	7.053	2.722	2.846	2.882	2.763	5.385	5.294	5.689	7.246	2.684	2.837	2.897	2.749	3.775	4.927	3.963
JB	1934	10.070	3.494	1.624	7.481	718.843	615.135	838.501	2080.406	11.705	3.238	1.246	8.449	95.369	433.321	110.605
p value	(0.000)	(0.007)	(0.174)	(0.444)	(0.024)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.198)	(0.536)	(0.015)	(0.000)	(0.000)	(0.000)
	INTC								MSFT							
Skewness	0.065	0.069	0.061	0.053	0.067	0.074	0.070	-0.002	0.257	0.112	0.152	0.103	0.108	-0.007	0.074	0.114
Kurtosis	5.790	2.741	2.723	2.850	2.733	2.727	3.442	3.514	6.389	2.794	2.856	2.982	2.862	3.532	3.953	3.927
JB	899.165	9.920	10.589	3.882	10.307	11.121	24.731	30.498	1354.820	10.624	13.094	4.927	7.519	32.598	107.191	105.006
p value	(0.000)	(0.007)	(0.005)	(0.144)	(0.006)	(0.004)	(0.000)	(0.000)	(0.000)	(0.005)	(0.001)	(0.085)	(0.023)	(0.000)	(0.000)	(0.000)
	ORCL								XRX							
Skewness	0.507	0.082	0.096	0.075	0.092	0.093	0.077	0.038	-0.243	0.083	0.106	0.075	0.079	0.145	-0.250	-0.240
Kurtosis	11.382	2.652	2.730	3.058	2.694	2.638	3.701	3.867	15.026	2.852	3.042	3.140	3.178	3.136	8.021	8.187
JB	8219	17.053	12.595	2.990	14.768	19.118	59.296	87.301	16700.040	5.684	5.403	4.857	6.499	11.855	2935	3129
p value	(0.000)	(0.000)	(0.002)	(0.224)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.058)	(0.067)	(0.088)	(0.039)	(0.003)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, $rr_t, pv_t, bv_t, cv_t, sv_t, ga_t$ denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.8: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector MAR)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	AA								AKS							
Skewness	-0.417	0.112	0.104	0.087	0.122	0.111	0.025	-0.069	-0.223	0.089	0.119	0.054	0.101	0.111	0.053	0.045
Kurtosis	7.920	2.706	2.854	2.836	2.739	2.711	3.642	3.536	7.490	2.868	3.174	3.052	2.995	2.956	5.347	6.254
JB	2870	15.766	7.481	6.619	14.673	15.328	47.818	35.374	2347	5.622	10.064	1.663	4.716	5.909	636.535	#######
p value	(0.000)	(0.000)	(0.024)	(0.037)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.060)	(0.007)	(0.435)	(0.095)	(0.052)	(0.000)	(0.000)
	DD								DOW							
Skewness	0.049	0.154	0.197	0.139	0.179	0.250	0.188	0.088	0.168	0.136	0.132	0.144	0.127	0.234	0.460	0.204
Kurtosis	7.482	2.884	3.070	3.078	2.955	3.629	4.690	4.536	8.834	2.892	3.065	3.073	2.898	4.099	8.051	5.573
JB	2317	12.468	18.426	9.552	14.934	74.411	345.673	275.429	3937	9.866	8.517	10.169	8.668	164.336	3039	782.327
p value	(0.000)	(0.002)	(0.000)	(0.008)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.007)	(0.014)	(0.006)	(0.013)	(0.000)	(0.000)	(0.000)
	FCX								IP							
Skewness	-0.230	0.093	0.144	0.059	0.141	0.243	0.063	-0.075	-0.148	0.006	0.018	-0.012	0.009	0.027	-0.115	-0.117
Kurtosis	6.384	2.799	3.256	2.859	2.964	3.826	4.596	4.111	10.143	2.750	2.843	2.925	2.819	2.901	4.992	4.123
JB	1345	8.696	17.091	3.888	9.290	105.979	295.333	144.768	5892	7.204	2.974	0.714	3.830	1.470	463.829	151.711
p value	(0.000)	(0.013)	(0.000)	(0.143)	(0.010)	(0.000)	(0.000)	(0.000)	(0.000)	(0.027)	(0.226)	(0.700)	(0.147)	(0.480)	(0.000)	(0.000)
	NEM								NUE							
Skewness	0.172	0.107	0.129	0.116	0.115	0.114	0.189	0.135	-0.511	0.041	0.003	-0.009	-0.002	0.034	-0.379	-0.447
Kurtosis	7.411	2.711	2.858	2.882	2.750	2.717	4.073	4.034	9.419	2.891	3.513	3.002	3.118	2.915	6.525	6.876
JB	2257	14.924	9.995	7.758	13.331	15.233	149.193	131.651	4871	2.114	30.348	0.038	1.608	1.367	1499	1824
p value	(0.000)	(0.001)	(0.007)	(0.021)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.347)	(0.000)	(0.981)	(0.447)	(0.505)	(0.000)	(0.000)
	WY								X							
Skewness	-0.334	0.104	0.111	0.080	0.114	0.111	0.111	0.059	-0.389	0.070	0.055	0.037	0.064	0.071	0.011	-0.030
Kurtosis	8.989	2.858	3.129	3.020	2.942	2.901	4.174	4.439	6.115	2.750	3.018	2.794	2.818	2.771	3.703	4.052
JB	4186	7.262	7.568	2.999	6.374	6.761	164.507	240.319	1188	9.431	1.413	5.509	5.699	8.316	56.983	127.924
p value	(0.000)	(0.026)	(0.023)	(0.223)	(0.041)	(0.034)	(0.000)	(0.000)	(0.000)	(0.009)	(0.493)	(0.064)	(0.058)	(0.016)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, $rr_t, pv_t, bv_t, cv_t, sv_t, ga_t$ denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.9: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector TEL)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
				AMT									BT			
Skewness	-0.454	0.055	0.037	0.038	0.054	0.072	0.080	-0.056	0.020	0.027	-11.571	0.041	0.041	3.027	-0.119	0.032
Kurtosis	14.652	2.664	3.028	2.919	2.817	2.715	4.106	4.683	7.956	3.020	370.827	3.301	3.008	60.011	4.703	4.171
JB	15747	14.382	0.728	1.443	5.196	11.724	143.857	328.029	2832	0.375	#######	11.191	0.771	374294	340.857	158.615
p value	(0.000)	(0.001)	(0.695)	(0.486)	(0.074)	(0.003)	(0.000)	(0.000)	(0.000)	(0.829)	(0.000)	(0.004)	(0.680)	(0.000)	(0.000)	(0.000)
				CTL									FTR			
Skewness	-0.201	-0.090	-0.075	-0.103	-0.080	-0.135	-0.113	-0.251	-0.456	0.116	0.178	0.134	0.090	0.164	0.081	-0.007
Kurtosis	8.842	2.878	3.069	3.071	2.951	3.137	6.664	9.543	14.656	2.920	3.418	3.422	3.111	3.592	4.951	4.652
JB	3954	5.433	3.139	5.431	3.243	10.550	1554	4965	15760	6.908	34.826	28.788	5.166	52.442	442.064	314.548
p value	(0.000)	(0.066)	(0.208)	(0.066)	(0.198)	(0.005)	(0.000)	(0.000)	(0.000)	(0.032)	(0.000)	(0.000)	(0.076)	(0.000)	(0.000)	(0.000)
				Q									S			
Skewness	-0.549	0.052	0.028	0.080	0.067	0.063	0.168	-0.422	-0.850	0.153	0.150	0.114	0.181	0.284	-0.154	-0.068
Kurtosis	26.816	2.781	3.170	3.179	2.872	2.801	5.626	12.621	13.893	2.802	3.214	3.026	2.901	3.690	6.091	4.920
JB	65535	6.743	3.711	6.597	3.953	6.392	808.309	10755	14013	15.357	15.709	6.097	16.191	92.171	1113	427.068
p value	(0.000)	(0.034)	(0.156)	(0.037)	(0.139)	(0.041)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.047)	(0.000)	(0.000)	(0.000)	(0.000)
				T									TEF			
Skewness	0.421	0.139	0.151	0.151	0.099	0.833	0.157	0.049	1.155	-0.052	-0.073	-0.056	-0.046	0.067	0.647	0.054
Kurtosis	6.611	2.755	3.047	2.986	2.828	11.957	3.788	3.811	23.778	2.756	2.913	3.083	2.904	5.177	12.434	5.576
JB	1585	15.832	10.792	10.559	7.984	9543	82.987	76.832	50390	8.110	3.358	2.262	2.028	540.037	10455	766.645
p value	(0.000)	(0.000)	(0.005)	(0.005)	(0.018)	(0.000)	(0.000)	(0.000)	(0.000)	(0.017)	(0.187)	(0.323)	(0.363)	(0.000)	(0.000)	(0.000)
				VOD									VZ			
Skewness	0.152	-0.004	-0.010	-0.012	-0.011	-0.209	0.191	0.149	0.460	0.100	0.123	0.139	0.119	0.205	0.101	0.088
Kurtosis	8.860	2.793	2.928	3.153	2.828	4.635	5.172	5.539	6.769	2.754	2.911	2.991	2.840	3.895	3.879	4.048
JB	3970	4.936	0.640	2.778	3.483	326.717	560.880	753.660	1735	11.561	7.917	8.927	9.490	111.531	93.801	130.122
p value	(0.000)	(0.085)	(0.726)	(0.249)	(0.175)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.019)	(0.012)	(0.009)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t, sv_t, ga_t denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table A.10: Returns standardized by parametric and nonparametric volatility measures at 5-minute sampling frequency (Sector UTL)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	$\frac{r_t}{\sqrt{sv_t}}$	$\frac{r_t}{\sqrt{ga_t}}$
	AEP								CEG							
Skewness	-0.426	0.029	-0.021	0.005	0.019	0.033	-0.080	-0.251	-5.074	-0.085	-0.067	-0.166	-0.065	0.824	-0.390	-0.615
Kurtosis	22.015	2.926	3.193	3.169	3.010	3.609	8.517	4.548	90.456	2.928	3.203	3.215	2.945	12.680	6.413	8.565
JB	41769	1.031	4.494	3.313	0.176	43.096	3512	305.407	893684	3.944	6.839	18.124	2.270	11084	1413	3744
p value	(0.000)	(0.597)	(0.106)	(0.191)	(0.916)	(0.000)	(0.000)	(0.000)	(0.000)	(0.139)	(0.033)	(0.000)	(0.321)	(0.000)	(0.000)	(0.000)
	DUK								ETR							
Skewness	0.349	0.091	0.045	0.054	0.093	2.390	-0.047	-0.126	-0.271	-0.030	-0.027	-0.054	-0.045	-0.037	-0.234	-0.320
Kurtosis	10.453	2.735	2.933	2.855	2.846	38.178	4.017	4.055	11.333	2.930	3.254	3.094	3.000	3.397	5.786	5.182
JB	6460	11.930	1.463	3.791	6.707	144727	120.297	135.661	8040	0.963	7.776	2.373	0.934	18.628	920.322	595.822
p value	(0.000)	(0.003)	(0.481)	(0.150)	(0.035)	(0.000)	(0.000)	(0.000)	(0.000)	(0.618)	(0.020)	(0.305)	(0.627)	(0.000)	(0.000)	(0.000)
	EXC								OKE							
Skewness	0.449	-0.032	-0.059	-0.031	-0.059	-5.241	0.226	-0.089	-0.061	-0.002	0.003	0.006	0.044	0.200	0.146	-0.031
Kurtosis	10.643	2.758	3.092	2.958	2.866	129.858	5.898	4.253	19.042	2.847	3.543	3.042	3.070	4.246	4.392	6.234
JB	6827	7.221	2.569	0.643	3.667	1866684	991.584	184.612	29673	2.718	34.041	0.218	1.458	196.240	233.302	1207
p value	(0.000)	(0.027)	(0.277)	(0.725)	(0.160)	(0.000)	(0.000)	(0.000)	(0.000)	(0.257)	(0.000)	(0.897)	(0.482)	(0.000)	(0.000)	(0.000)
	PCG								PEG							
Skewness	-3.057	-0.003	0.047	0.012	-0.012	0.820	-0.326	-1.038	0.124	0.049	0.010	0.028	0.053	0.054	-0.142	-0.078
Kurtosis	75.986	2.905	3.203	3.375	3.078	13.762	18.593	15.594	9.734	3.050	3.424	3.265	3.084	3.291	5.231	4.354
JB	618465	1.038	5.748	16.311	0.772	13638	28081	18785	5235	1.392	20.732	8.483	2.124	11.028	583.007	214.182
p value	(0.000)	(0.595)	(0.056)	(0.000)	(0.680)	(0.000)	(0.000)	(0.000)	(0.000)	(0.499)	(0.000)	(0.014)	(0.346)	(0.004)	(0.000)	(0.000)
	PGN								SO							
Skewness	0.390	-0.016	-0.003	-0.012	-0.030	0.038	-0.115	-0.199	0.235	0.103	0.136	0.116	0.128	0.002	0.080	0.079
Kurtosis	11.975	2.961	3.133	3.135	2.967	6.160	3.871	4.295	8.466	3.281	3.583	3.525	3.367	4.735	4.159	4.677
JB	9358	0.303	2.054	2.175	0.541	1142	93.443	211.649	3470	13.933	47.790	37.953	23.097	345.579	157.751	327.204
p value	(0.000)	(0.860)	(0.358)	(0.337)	(0.763)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, $rr_t, pv_t, bv_t, cv_t, sv_t, ga_t$ denote realized variance, realized range, realized power variation, realized bipower variation, realized continuous variance, GARCH (1,1) and stochastic volatility, respectively. The sampling frequency is 5 minute.

Table B.1: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector CD)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			AMZN						BBY			
Skewness	0.389	0.167	0.072	0.101	0.167	0.169	-1.472	0.099	-0.454	0.086	0.073	0.097
Kurtosis	8.888	3.249	2.846	2.959	3.275	3.252	36.33883	2.911472	13.86547	3.371016	2.956	2.920
JB	4067.133	19.930	5.102	4.888	21.496	20.547	129143	5.409	13706.050	19.288	2.706	5.090
p value	(0.000)	(0.000)	(0.078)	(0.087)	(0.000)	(0.000)	(0.000)	(0.067)	(0.000)	(0.000)	(0.259)	(0.078)
			DIS						GPS			
Skewness	0.274	0.146	-0.129	0.101	0.149	0.148	0.265	-0.003	-0.132	0.022	0.049	0.004
Kurtosis	6.149	3.371	3.956	3.221	3.487	3.368	7.252	3.002	7.270	3.454	3.310	3.011
JB	1178.158	25.757	113.202	10.347	37.502	25.820	2116.506	0.006	2110.573	23.948	12.205	0.022
p value	(0.000)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.997)	(0.000)	(0.000)	(0.002)	(0.989)
			HD						IPG			
Skewness	0.540	0.205	0.045	0.193	0.204	0.205	0.388	-0.029	-6.749	-0.039	-0.118	-0.134
Kurtosis	6.652	3.270	3.823	3.160	3.367	3.274	18.159	4.436	150.825	4.930	5.210	4.976
JB	1671.805	27.730	78.918	20.052	34.835	28.012	26438.790	237.074	2528435	428.011	566.595	458.324
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			MAR						MCD			
Skewness	0.199	-0.050	1.455	-0.021	-0.047	-0.724	0.221	0.102	0.600	0.120	0.206	0.127
Kurtosis	8.799	3.330	81.926	3.532	3.576	9.943	9.388	3.303	8.021	3.356	4.066	3.390
JB	3813.798	13.413	704082	32.081	38.463	5744.364	4727.773	15.412	3072.954	21.290	150.554	24.938
p value	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			NWSA						TWX			
Skewness	0.306	0.142	0.371	0.216	0.232	4.118	0.293	0.233	0.163	0.188	0.245	0.241
Kurtosis	11.769	4.233	25.113	5.217	5.879	84.133	9.963	3.786	3.233	3.571	3.937	3.846
JB	7147.082	148.118	45282.000	472.076	786.345	748998	5628.714	96.233	18.567	53.877	129.018	109.332
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.2: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector CS)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			AVP						BFB			
Skewness	0.233	0.084	0.121	0.084	0.097	0.175	0.266	0.056	-0.829	0.021	-0.107	0.981
Kurtosis	9.157	3.184	27.955	3.429	3.492	4.100	11.70605	4.115229	35.06983	4.078358	4.997	64.138
JB	4323.964	7.073	70638	24.021	31.803	153.443	7043.489	116.310	95474.130	107.820	373.548	398341
p value	(0.000)	(0.029)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			COST						EL			
Skewness	0.313	0.151	0.108	0.103	0.176	0.179	0.230	0.099	0.393	0.078	0.100	0.186
Kurtosis	7.115	3.269	3.008	2.850	3.334	3.404	7.776	3.236	59.871	3.433	3.384	4.528
JB	1997.535	18.904	5.356	7.482	27.137	33.691	2546.688	10.505	357869.500	23.484	20.778	284.377
p value	(0.000)	(0.000)	(0.069)	(0.024)	(0.000)	(0.000)	(0.000)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)
			KMB						KO			
Skewness	-0.033	0.172	0.103	0.158	0.177	0.194	0.259	0.108	0.312	0.119	0.102	0.105
Kurtosis	10.731	3.233	13.425	3.574	3.389	4.082	8.037	3.442	5.310	3.501	3.495	3.445
JB	6888.536	19.857	12530.010	49.379	31.916	152.258	2956.050	27.896	660.351	35.454	33.076	27.902
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			PEP						PG			
Skewness	0.101	0.046	1.087	0.007	0.028	0.063	-0.226	0.026	0.475	-0.024	-0.043	0.053
Kurtosis	10.070	3.341	17.034	3.454	3.650	3.421	10.719	3.266	6.399	3.455	3.674	3.319
JB	5766.861	14.351	23251.280	23.776	49.122	22.295	6892.808	8.477	1436.007	24.152	53.236	13.075
p value	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.014)	(0.000)	(0.000)	(0.000)	(0.001)
			UL						WMT			
Skewness	-0.102	-0.003	1.366	-0.030	0.013	0.185	0.273	0.138	0.125	0.123	0.112	0.139
Kurtosis	8.939	3.363	35.422	3.669	3.481	7.328	7.287	2.951	3.704	3.133	3.308	2.969
JB	3349.048	12.518	100395	42.745	22.009	2156.308	2153.341	9.031	64.349	9.034	16.763	8.994
p value	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.011)	(0.000)	(0.011)	(0.000)	(0.011)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.3: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector ENG)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			BHI						CHK			
Skewness	-0.074	0.082	-0.398	0.039	0.043	0.085	-0.559	0.030	-0.109	0.076	0.040	0.041
Kurtosis	6.791	3.005	8.582	2.994	3.018	3.013	16.356	3.463	27.663	3.859	3.407	3.505
JB	1659.764	3.082	3665.747	0.716	0.880	3.335	20044	24.309	67877.850	84.894	19.191	30.146
p value	(0.000)	(0.214)	(0.000)	(0.699)	(0.644)	(0.189)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			CVX						DVN			
Skewness	0.114	0.056	-0.243	0.034	0.064	0.078	-0.117	-0.050	10.744	-0.076	-0.005	-0.047
Kurtosis	13.788	2.963	10.142	3.304	3.105	3.134	7.307	2.975	419.537	3.761	3.648	3.150
JB	13423	1.613	5907.704	11.224	3.170	4.907	2051	1.164	19179644	66.351	46.290	3.600
p value	(0.000)	(0.446)	(0.000)	(0.004)	(0.205)	(0.086)	(0.000)	(0.559)	(0.000)	(0.000)	(0.000)	(0.165)
			HAL						OXY			
Skewness	-2.747	0.004	-0.209	-0.038	0.061	0.007	-0.391	-0.083	0.149	-0.090	-0.062	-0.069
Kurtosis	61.332	2.999	8.182	3.104	3.386	3.014	8.723	3.202	7.826	3.007	3.284	3.307
JB	395770	0.008	3115.932	1.923	18.889	0.041	3830	7.838	2683.296	3.686	11.065	13.033
p value	(0.000)	(0.996)	(0.000)	(0.382)	(0.000)	(0.979)	(0.000)	(0.020)	(0.000)	(0.158)	(0.004)	(0.001)
			SUN						TE			
Skewness	0.202	0.019	5.488	0.052	0.017	0.097	-1.359	-0.117	-0.124	-0.070	-0.067	-0.104
Kurtosis	8.611	3.324	135.961	3.163	3.388	3.839	26.144	3.241	59.874	3.761	3.449	3.266
JB	3525.105	11.822	1982362	4.150	16.884	85.463	59890	12.464	356758	66.008	24.247	13.154
p value	(0.000)	(0.003)	(0.000)	(0.126)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
			WMB						XOM			
Skewness	-2.286	-0.096	-0.185	-0.096	-0.087	-0.094	0.052	0.081	-0.086	0.004	0.064	0.111
Kurtosis	46.109	3.057	6.881	3.101	3.155	3.058	11.715	3.137	5.539	3.174	3.318	3.191
JB	216664	4.624	1752.108	5.396	6.266	4.430	8757.847	5.215	746.588	3.492	13.570	9.933
p value	(0.000)	(0.099)	(0.000)	(0.067)	(0.044)	(0.109)	(0.000)	(0.074)	(0.000)	(0.174)	(0.001)	(0.007)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.4: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector FIN)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			ALL						AXP			
Skewness	0.036	0.026	-0.156	0.044	0.010	0.043	0.060	0.070	0.084	0.022	0.073	0.094
Kurtosis	20.832	3.246	7.188	3.508	3.209	3.248	8.211	3.142	4.744	3.023	3.167	3.181
JB	36662	7.294	2033.664	30.618	5.064	7.952	3132.834	4.571	353.896	0.275	5.661	7.812
p value	(0.000)	(0.026)	(0.000)	(0.000)	(0.080)	(0.019)	(0.000)	(0.102)	(0.000)	(0.872)	(0.059)	(0.020)
			BAC						BK			
Skewness	-0.523	-0.049	-0.066	-0.073	-0.052	-0.029	0.472	0.085	-0.113	0.090	0.074	0.082
Kurtosis	20.042	3.680	7.515	4.097	3.716	3.657	13.340	3.155	10.499	3.446	3.247	3.142
JB	33610	54.474	2352.285	141.254	60.335	50.109	12429	6.120	6490	26.694	9.547	5.406
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.047)	(0.000)	(0.000)	(0.008)	(0.067)
			C						GS			
Skewness	-1.647	0.073	0.267	0.028	0.072	0.103	0.119	0.111	0.062	0.124	0.188	0.141
Kurtosis	25.030	3.990	4.049	4.096	4.459	4.018	13.842	2.747	7.063	3.202	3.069	2.952
JB	57205	115.508	159.557	138.907	247.735	124.339	13558	13.083	1905.124	11.769	16.918	9.471
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.003)	(0.000)	(0.009)
			JPM						MS			
Skewness	0.538	0.041	-0.062	0.068	0.034	0.042	-0.794	0.016	0.047	0.000	-0.004	0.021
Kurtosis	13.706	3.347	5.440	3.481	3.348	3.315	22.289	2.824	5.947	3.149	3.064	2.849
JB	13347	14.704	688.419	28.757	14.481	12.240	43185	3.671	1002.289	2.557	0.488	2.836
p value	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)	(0.160)	(0.000)	(0.278)	(0.783)	(0.242)
			TRV						WFC			
Skewness	0.164	0.037	-1.896	0.113	0.131	0.356	0.036	0.081	-0.013	-0.025	0.070	0.064
Kurtosis	13.009	3.260	65.387	4.146	3.998	5.863	18.210	3.446	5.071	4.368	3.434	3.474
JB	11279	8.225	439321	153.355	119.661	1002.010	26673	26.002	494.712	216.066	23.971	27.779
p value	(0.000)	(0.016)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.5: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector HC)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			ABT						AMGN			
Skewness	0.004	0.023	0.082	0.030	0.065	0.028	-0.278	-0.035	0.018	-0.059	-0.030	-0.024
Kurtosis	6.415	3.421	5.751	3.480	3.630	3.403	6.594	3.570	3.089	3.429	3.628	3.560
JB	1344.213	20.675	875.534	27.015	47.712	19.084	1524.757	38.027	1.065	22.782	45.857	36.458
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.587)	(0.000)	(0.000)	(0.000)
			BSX						GILD			
Skewness	0.108	0.146	-1.082	0.126	0.078	0.145	0.006	-0.020	-0.630	-0.076	-0.041	-0.021
Kurtosis	7.532	3.435	28.556	3.717	3.727	3.500	7.474	2.943	8.938	3.210	3.017	2.953
JB	2365.590	31.602	75593	66.362	63.466	38.537	2305.469	0.568	4244.129	7.736	0.813	0.452
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.753)	(0.000)	(0.021)	(0.666)	(0.798)
			HUM						JNJ			
Skewness	-0.264	0.119	36.899	0.077	0.103	0.128	0.136	0.105	0.260	0.084	0.135	0.109
Kurtosis	7.803	3.236	1718.352	2.961	3.302	3.352	8.284	3.301	5.725	3.539	3.460	3.321
JB	2638.680	12.714	3.E+08	2.867	15.108	21.790	3228.090	15.486	887.248	36.735	32.864	17.342
p value	(0.000)	(0.002)	(0.000)	(0.239)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			MDT						MRK			
Skewness	-0.168	0.033	0.095	0.003	0.150	0.044	0.035	0.049	0.177	0.040	0.040	0.053
Kurtosis	6.287	2.985	5.130	3.111	3.693	2.990	8.097	3.120	4.667	3.232	3.215	3.110
JB	1258.669	0.527	527.257	1.426	65.690	0.897	2995.517	2.780	334.717	6.959	6.078	2.701
p value	(0.000)	(0.768)	(0.000)	(0.490)	(0.000)	(0.639)	(0.000)	(0.249)	(0.000)	(0.031)	(0.048)	(0.259)
			PFE						UNH			
Skewness	0.194	0.041	0.120	0.060	0.046	0.041	0.456	0.059	-7.197	0.071	0.062	0.077
Kurtosis	4.934	3.126	4.210	3.295	3.167	3.118	16.481	3.025	210.813	3.445	3.267	3.165
JB	448.795	2.603	175.583	11.729	4.217	2.374	20773.740	1.635	4.94E+06	24.779	9.868	5.878
p value	(0.000)	(0.272)	(0.000)	(0.003)	(0.121)	(0.305)	(0.000)	(0.441)	(0.000)	(0.000)	(0.007)	(0.053)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.6: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector IND)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			BA						CAT			
Skewness	0.053	0.003	0.027	-0.006	0.001	-0.022	0.013	-0.018	0.083	0.001	-0.032	-0.016
Kurtosis	5.297	3.113	5.487	2.869	3.161	3.304	5.343	2.990	9.494	3.093	3.045	3.017
JB	609.784	1.481	713.479	1.980	2.978	10.909	633.116	0.159	4865.452	1.003	0.712	0.153
p value	(0.000)	(0.477)	(0.000)	(0.372)	(0.226)	(0.004)	(0.000)	(0.924)	(0.000)	(0.606)	(0.701)	(0.927)
			CMI						GD			
Skewness	-0.230	0.071	-9.384	0.049	0.087	0.093	-0.206	0.060	-6.039	0.005	0.031	0.068
Kurtosis	7.762	3.083	221.777	3.182	3.684	4.026	6.641	3.027	196.260	3.365	3.277	3.316
JB	2466.118	2.911	5.20E+06	4.597	53.651	125.080	1530.861	1.734	4.27E+06	15.226	9.185	13.587
p value	(0.000)	(0.233)	(0.000)	(0.100)	(0.000)	(0.000)	(0.000)	(0.420)	(0.000)	(0.000)	(0.010)	(0.001)
			GE						LUV			
Skewness	-0.115	0.082	0.106	0.085	0.078	0.085	0.090	0.091	0.136	0.101	0.100	0.041
Kurtosis	9.256	3.461	3.865	3.677	3.569	3.454	6.910	3.372	5.880	3.474	3.468	3.210
JB	4517.602	27.538	91.549	56.170	40.154	27.089	1765.023	19.745	963.931	30.519	29.831	5.837
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.054)
			MMM						UPS			
Skewness	0.498	-0.005	-0.436	0.113	-0.010	0.092	0.195	-0.003	1.351	-0.044	-0.050	-0.023
Kurtosis	8.805	2.960	10.924	3.771	3.072	3.367	7.171	3.373	64.548	3.522	3.536	3.019
JB	3999.275	0.194	7326.603	74.509	0.644	19.470	2023.329	16.073	437591	32.278	34.250	0.288
p value	(0.000)	(0.907)	(0.000)	(0.000)	(0.725)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.866)
			HON						UTX			
Skewness	-0.423	0.033	0.897	0.026	0.050	0.006	-0.234	0.049	0.037	0.031	0.074	0.108
Kurtosis	10.040	3.190	26.622	3.363	3.281	3.498	9.330	2.772	3.023	3.003	2.852	3.435
JB	5796.558	4.675	64703.750	15.495	10.255	28.614	4644.655	7.099	0.697	0.442	5.043	27.247
p value	(0.000)	(0.097)	(0.000)	(0.000)	(0.006)	(0.000)	(0.000)	(0.029)	(0.706)	(0.802)	(0.080)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.7: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector IT)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			AAPL						DELL			
Skewness	0.064	0.107	0.083	0.072	0.107	0.111	0.217	-0.003	0.002	0.004	-0.012	-0.033
Kurtosis	5.020	3.273	3.252	2.938	3.236	3.276	7.280	3.245	3.126	3.257	3.321	3.490
JB	472.348	13.856	10.507	2.818	11.678	14.469	2133.824	6.944	1.836	7.639	11.910	28.197
p value	(0.000)	(0.001)	(0.005)	(0.244)	(0.003)	(0.001)	(0.000)	(0.031)	(0.399)	(0.022)	(0.003)	(0.000)
			EMC						HPQ			
Skewness	0.197	0.038	0.161	0.020	0.043	-0.005	0.303	0.025	-0.171	-0.047	0.033	0.039
Kurtosis	8.047	3.120	3.602	3.277	3.204	3.251	7.045	3.188	4.439	3.126	3.233	3.116
JB	2954.311	2.326	53.709	9.023	5.670	7.276	1928.636	4.373	252.381	2.852	6.765	2.234
p value	(0.000)	(0.313)	(0.000)	(0.011)	(0.059)	(0.026)	(0.000)	(0.112)	(0.000)	(0.240)	(0.034)	(0.327)
			IBM						INTC			
Skewness	0.059	0.064	-0.052	-0.037	0.075	0.041	0.061	0.064	0.076	0.072	0.059	0.112
Kurtosis	7.167	3.164	3.590	2.978	3.209	3.212	5.700	3.092	3.053	3.048	3.149	3.241
JB	2003.553	4.987	41.341	0.699	7.621	5.970	842.260	2.855	2.955	2.636	4.142	12.519
p value	(0.000)	(0.083)	(0.000)	(0.705)	(0.022)	(0.051)	(0.000)	(0.240)	(0.228)	(0.268)	(0.126)	(0.002)
			MSFT						ORCL			
Skewness	0.248	0.091	0.134	0.109	0.089	0.064	0.483	0.049	0.084	0.039	0.038	0.091
Kurtosis	6.216	3.613	3.200	3.657	3.625	3.092	11.183	3.365	3.142	3.198	3.466	3.599
JB	1220.467	47.160	12.883	55.235	48.727	2.872	7827.708	16.436	5.564	5.228	25.739	45.134
p value	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.238)	(0.000)	(0.000)	(0.062)	(0.073)	(0.000)	(0.000)
			XRX						CSCO			
Skewness	-0.235	0.086	0.092	0.057	-0.007	0.049	0.373	-0.033	-0.004	-0.021	-0.042	0.085
Kurtosis	15.032	3.412	7.146	4.303	5.158	3.374	11.427	3.489	3.331	3.439	3.568	3.409
JB	16717	23.010	1985.468	197.087	536.797	17.213	8250.922	28.079	12.678	22.420	37.949	22.618
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.8: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector MAR)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			AA						AKS			
Skewness	-0.451	0.036	0.199	0.070	0.036	0.037	-0.239	0.086	-13.058	0.075	0.149	0.109
Kurtosis	8.308	3.226	8.356	3.181	3.260	3.235	7.271	3.285	427.631	3.267	3.726	3.383
JB	3342.485	6.460	3325.977	6.069	8.429	7.003	1911.751	11.448	1.87E+07	9.584	63.709	22.437
p value	(0.000)	(0.040)	(0.000)	(0.048)	(0.015)	(0.030)	(0.000)	(0.003)	(0.000)	(0.008)	(0.000)	(0.000)
			DD						DOW			
Skewness	0.035	0.123	0.478	0.157	0.181	0.142	0.113	0.058	-1.191	0.097	0.100	0.052
Kurtosis	7.523	3.169	6.912	3.362	3.461	3.183	8.894	3.007	22.514	3.331	3.304	2.999
JB	2358.693	10.224	1869.817	26.574	39.636	13.229	4009.810	1.534	44539	16.987	15.248	1.263
p value	(0.000)	(0.006)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.464)	(0.000)	(0.000)	(0.000)	(0.532)
			FCX						IP			
Skewness	-0.240	0.083	2.611	0.149	0.199	0.098	-0.165	-0.030	0.168	-0.022	-0.031	-0.015
Kurtosis	6.462	2.972	67.971	3.540	3.571	3.092	10.205	3.034	5.527	2.974	3.081	3.139
JB	1330.447	3.066	462735.100	41.436	52.874	5.392	5997.822	0.541	749.074	0.293	1.191	2.318
p value	(0.000)	(0.216)	(0.000)	(0.000)	(0.000)	(0.067)	(0.000)	(0.763)	(0.000)	(0.864)	(0.551)	(0.314)
			NEM						NUE			
Skewness	0.178	0.112	0.452	0.129	0.113	0.111	-0.456	0.028	40.338	0.000	0.001	0.053
Kurtosis	7.423	2.882	7.152	2.971	2.868	2.889	9.958	3.227	1936.654	3.728	3.993	3.725
JB	2264.693	7.312	2076.874	7.710	7.916	7.072	5459.391	6.031	4.15E+08	58.723	109.398	61.767
p value	(0.000)	(0.026)	(0.000)	(0.021)	(0.019)	(0.029)	(0.000)	(0.049)	(0.000)	(0.000)	(0.000)	(0.000)
			WY						X			
Skewness	-0.352	0.071	-0.161	0.095	0.092	0.088	-0.329	0.038	-0.903	0.027	0.099	0.052
Kurtosis	8.987	3.186	26.363	3.407	3.408	3.406	6.322	2.828	38.393	3.019	3.094	2.917
JB	4185.492	6.348	62874.110	23.245	23.062	22.613	1294.738	3.984	141819.200	0.366	5.446	2.053
p value	(0.000)	(0.042)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.136)	(0.000)	(0.833)	(0.066)	(0.358)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.9: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector TEL)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			AMT						BT			
Skewness	-0.452	0.168	2.142	0.159	0.184	0.154	-0.397	0.086	4.041	0.069	0.123	0.233
Kurtosis	15.583	2.970	71.583	3.479	3.103	3.032	9.416	3.183	89.083	3.594	3.367	3.940
JB	17746	12.630	5.27E+05	36.877	16.360	11.097	3337.033	5.013	5.97E+05	29.687	15.588	126.477
p value	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.082)	(0.000)	(0.000)	(0.000)	(0.000)
			CTL						FTR			
Skewness	-0.221	-0.069	-1.232	-0.033	-0.041	-0.143	0.062	0.221	-1.146	0.302	0.352	0.226
Kurtosis	8.868	3.181	29.650	3.407	3.548	4.925	14.951	4.791	26.458	6.271	5.853	4.671
JB	3860.873	5.744	79866	18.927	34.238	435.842	15356	365.750	59719	1189.087	928.557	345.586
p value	(0.000)	(0.057)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			Q						S			
Skewness	-0.474	-0.083	-0.291	0.000	-0.095	-0.087	-0.802	0.113	-0.364	0.124	0.231	0.122
Kurtosis	26.906	5.422	9.069	5.256	5.448	5.463	13.279	4.062	9.425	4.774	4.839	4.103
JB	65993.310	679.293	4285.882	586.546	694.985	702.816	12478.630	135.986	4820.770	369.756	414.545	147.119
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			T						TEF			
Skewness	0.391	0.064	-0.242	0.113	0.049	0.061	1.180	-0.029	-0.033	-0.032	0.141	-0.742
Kurtosis	6.604	3.557	8.291	4.003	3.676	3.558	25.244	3.022	31.204	3.617	4.328	17.115
JB	1567.516	37.633	3254.665	121.895	53.788	37.607	50848	0.388	80839	39.093	187.430	23121.530
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.824)	(0.000)	(0.000)	(0.000)	(0.000)
			VOD						VZ			
Skewness	0.153	-0.018	-0.284	-0.019	-0.021	-0.012	0.488	0.069	0.124	0.151	0.080	0.072
Kurtosis	8.869	3.370	8.083	3.654	3.427	3.329	6.640	3.283	4.887	3.451	3.389	3.276
JB	3982.464	15.950	3016.337	49.479	21.189	12.530	1637.755	11.376	417.499	33.918	20.335	11.136
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.004)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table B.10: Returns standardized by realized volatility measures at 1-second sampling frequency (Sector UTL)

	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$	r_t	$\frac{r_t}{\sqrt{rv_t}}$	$\frac{r_t}{\sqrt{rr_t}}$	$\frac{r_t}{\sqrt{pv_t}}$	$\frac{r_t}{\sqrt{bv_t}}$	$\frac{r_t}{\sqrt{cv_t}}$
			AEP						**CEG**			
Skewness	-0.415	-0.002	-8.842	-0.054	-0.196	0.084	-5.105	-0.091	0.072	-0.126	-0.125	-0.065
Kurtosis	21.837	3.093	242.070	3.463	4.338	3.538	90.319	3.438	34.539	3.657	3.841	3.711
JB	40738	1.001	6584788	25.887	222.852	36.619	859844	24.989	110665	55.157	85.616	60.161
p value	(0.000)	(0.606)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			DUK						**ETR**			
Skewness	0.428	-0.041	-1.207	-0.056	-0.059	-0.039	-0.302	-0.004	-0.769	-0.082	-0.017	2.439
Kurtosis	10.698	3.325	21.886	3.788	3.388	3.323	11.375	3.261	44.097	3.889	3.759	66.178
JB	6915.797	12.955	41792	73.072	18.995	12.695	7937.384	7.679	190418	91.926	64.990	457745
p value	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.022)	(0.000)	(0.000)	(0.000)	(0.000)
			OKE						**PCG**			
Skewness	-0.138	-0.047	8.981	-0.030	-0.039	-0.049	-3.052	0.028	-0.506	0.062	0.055	-0.025
Kurtosis	21.550	3.041	326.000	3.524	3.308	3.595	75.234	3.286	37.916	4.010	3.454	4.486
JB	33529	1.012	1.02E+07	27.092	9.813	41.868	602789	9.762	139960	118.666	25.024	254.015
p value	(0.000)	(0.603)	(0.000)	(0.000)	(0.007)	(0.000)	(0.000)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)
			PEG						**PGN**			
Skewness	0.111	0.001	2.039	-0.034	0.029	0.057	0.401	-0.004	2.271	0.019	0.015	-0.266
Kurtosis	9.763	3.369	33.306	3.872	3.825	3.330	11.903	3.130	62.485	3.583	3.267	7.361
JB	5099.590	15.125	104145	85.296	76.163	14.109	9048.454	1.915	403071	38.703	8.185	2221.207
p value	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.384)	(0.000)	(0.000)	(0.017)	(0.000)
			SO						**EXC**			
Skewness	0.210	0.143	0.125	0.143	0.134	1.933	0.426	-0.051	1.541	-0.044	-0.099	0.139
Kurtosis	8.472	3.784	6.653	4.125	3.923	32.032	10.639	3.162	21.410	3.573	3.461	3.742
JB	3470.970	80.230	1545.129	155.334	106.439	98858	6755.537	4.167	39837.540	38.461	28.779	72.388
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.124)	(0.000)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. rv_t, rr_t, pv_t, bv_t, cv_t denote realized variance, realized range, realized power variation, realized bipower variation, and realized continuous variance, respectively. The sampling frequency is 1 second.

Table C.1: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector CD)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			AMZN					BBY		
Skewness	-0.085	0.147	0.136	0.067	0.141	-1.193	0.045	0.009	-0.022	-0.020
Kurtosis	7.020	3.191	3.160	3.328	3.198	24.672	3.236	3.609	3.868	3.589
JB	1866.119	14.167	11.415	14.434	13.657	54808	7.377	42.784	87.182	40.202
p value	(0.000)	(0.001)	(0.003)	(0.001)	(0.001)	(0.000)	(0.025)	(0.000)	(0.000)	(0.000)
			DIS					GPS		
Skewness	0.198	0.021	0.041	0.032	0.053	0.154	-0.034	-0.062	-0.027	-0.047
Kurtosis	5.725	2.822	2.845	2.934	2.918	6.616	3.169	3.559	3.519	3.192
JB	874.308	3.843	3.521	0.965	2.097	1518.156	3.832	37.778	31.391	5.292
p value	(0.000)	(0.146)	(0.172)	(0.617)	(0.350)	(0.000)	(0.147)	(0.000)	(0.000)	(0.071)
			HD					IPG		
Skewness	0.514	0.109	0.127	0.110	0.106	0.082	-0.028	-0.262	-0.084	-0.015
Kurtosis	7.080	2.991	3.165	3.167	3.045	19.014	3.368	4.912	4.140	3.372
JB	2040.648	5.478	10.568	8.827	5.449	29571	15.932	452.958	153.104	16.035
p value	(0.000)	(0.065)	(0.005)	(0.012)	(0.066)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			MAR					MCD		
Skewness	-0.020	-0.122	-0.119	-0.115	-0.141	0.083	0.028	0.031	0.026	0.084
Kurtosis	7.733	3.178	3.390	3.247	3.196	7.371	2.826	2.950	3.191	3.184
JB	2582.933	10.542	24.080	13.161	13.656	2205.782	3.835	0.741	4.497	7.142
p value	(0.000)	(0.005)	(0.000)	(0.001)	(0.001)	(0.000)	(0.147)	(0.691)	(0.106)	(0.028)
			NWSA					TWX		
Skewness	0.201	-0.035	-0.032	-0.028	0.029	-0.066	0.028	0.032	-0.024	0.048
Kurtosis	8.750	2.985	4.021	3.342	4.368	7.571	2.840	3.032	3.236	2.918
JB	3830.545	0.592	120.710	13.834	216.136	2410.782	3.318	0.593	6.684	1.865
p value	(0.000)	(0.744)	(0.000)	(0.001)	(0.000)	(0.000)	(0.190)	(0.744)	(0.035)	(0.394)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.2: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector CS)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			AVP					BFB		
Skewness	-0.010	-0.042	-0.022	-0.038	-0.015	0.393	0.070	0.096	0.095	0.068
Kurtosis	8.616	3.437	3.587	3.740	3.370	10.451	3.357	4.265	3.618	3.434
JB	3636.647	22.871	40.004	63.882	15.912	6471.521	16.967	188.917	48.233	23.852
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			COST					EL		
Skewness	0.392	0.041	0.015	0.018	0.037	0.077	0.128	0.093	0.114	0.125
Kurtosis	8.471	3.004	2.972	3.259	3.007	6.622	2.985	3.176	3.159	3.030
JB	3522.429	0.766	0.192	7.858	0.622	1515.407	7.588	7.526	8.892	7.290
p value	(0.000)	(0.682)	(0.908)	(0.020)	(0.733)	(0.000)	(0.023)	(0.023)	(0.012)	(0.026)
			KMB					KO		
Skewness	-0.067	-0.005	-0.064	-0.059	-0.009	-0.143	0.008	0.013	-0.010	0.020
Kurtosis	9.920	3.173	3.686	3.590	3.222	6.839	3.042	3.157	3.221	3.063
JB	5523.493	3.445	56.086	41.701	5.736	1708.364	0.228	2.903	5.704	0.654
p value	(0.000)	(0.179)	(0.000)	(0.000)	(0.057)	(0.000)	(0.892)	(0.234)	(0.058)	(0.721)
			PEP					PG		
Skewness	-0.166	0.012	-0.031	-0.041	0.028	-0.516	-0.072	-0.079	-0.140	-0.105
Kurtosis	9.383	2.973	3.141	3.308	3.046	8.379	2.974	3.114	3.239	3.150
JB	4709.515	0.148	2.754	11.711	0.605	3458.303	2.486	4.412	15.708	7.680
p value	(0.000)	(0.929)	(0.252)	(0.003)	(0.739)	(0.000)	(0.289)	(0.110)	(0.000)	(0.021)
			UL					WMT		
Skewness	-0.330	-0.224	-0.219	-0.286	-0.221	-0.194	-0.008	-0.017	-0.029	-0.014
Kurtosis	8.466	3.038	3.220	3.346	3.073	7.720	3.041	3.123	3.303	3.099
JB	3494.419	23.327	27.702	51.436	23.045	2586.294	0.223	1.882	10.948	1.219
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.895)	(0.390)	(0.004)	(0.544)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.3: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector ENG)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			BHI					**CHK**		
Skewness	-0.309	0.055	0.001	0.006	0.064	-0.910	-0.058	-0.055	-0.125	-0.051
Kurtosis	5.500	3.055	3.217	3.115	3.108	14.140	3.066	3.062	3.389	3.053
JB	764.684	1.734	5.436	1.548	3.261	14690	2.054	1.833	24.669	1.513
p value	(0.000)	(0.420)	(0.066)	(0.461)	(0.196)	(0.000)	(0.358)	(0.400)	(0.000)	(0.469)
			CVX					**DVN**		
Skewness	-0.471	-0.075	-0.104	-0.126	-0.048	-0.282	-0.119	-0.181	-0.150	-0.110
Kurtosis	9.793	3.033	3.352	3.159	3.093	5.775	3.185	3.981	3.343	3.323
JB	5423.203	2.742	19.261	10.225	2.067	924.537	10.463	126.149	23.967	17.628
p value	(0.000)	(0.254)	(0.000)	(0.006)	(0.356)	(0.000)	(0.005)	(0.000)	(0.000)	(0.000)
			HAL					**OXY**		
Skewness	-2.651	-0.171	-0.159	-0.400	-0.131	-0.518	-0.073	-0.110	-0.141	-0.061
Kurtosis	44.867	4.030	4.110	6.258	3.802	7.473	3.068	3.249	3.064	3.072
JB	205330	135.864	153.790	1297.396	82.128	2430.618	3.021	12.769	9.669	2.290
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.221)	(0.002)	(0.008)	(0.318)
			SUN					**TE**		
Skewness	0.082	0.031	0.042	-0.014	0.031	-2.100	-0.135	-0.178	-0.250	-0.104
Kurtosis	7.624	3.043	3.249	3.085	3.044	31.315	3.241	3.477	3.867	3.338
JB	2468.502	0.642	7.980	0.922	0.672	94468	15.133	40.774	115.496	18.096
p value	(0.000)	(0.726)	(0.018)	(0.631)	(0.714)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
			WMB					**XOM**		
Skewness	-4.092	-0.114	-0.172	-0.341	-0.109	-0.233	0.007	0.047	-0.064	0.003
Kurtosis	84.144	3.328	3.417	5.078	3.377	9.203	3.064	3.215	3.117	3.079
JB	766845	18.417	33.689	551.395	21.902	4460.612	0.497	6.376	3.465	0.718
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.780)	(0.041)	(0.177)	(0.698)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.4: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector FIN)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			ALL					**AXP**		
Skewness	-1.038	-0.049	-0.034	-0.051	-0.052	0.089	-0.044	-0.051	-0.065	-0.032
Kurtosis	25.522	3.053	3.145	3.764	3.074	7.838	2.793	2.900	3.032	2.797
JB	58975.410	1.406	2.937	68.419	1.876	2702.630	5.842	2.335	2.069	5.202
p value	(0.000)	(0.495)	(0.230)	(0.000)	(0.391)	(0.000)	(0.054)	(0.311)	(0.355)	(0.074)
			BAC					**BK**		
Skewness	-1.749	-0.057	-0.096	-0.218	-0.070	0.121	0.021	-0.042	0.049	-0.007
Kurtosis	27.030	3.006	3.218	3.866	3.104	12.370	3.091	3.377	3.375	3.179
JB	67984.490	1.505	9.778	108.381	3.514	10128	1.150	17.236	17.313	3.723
p value	(0.000)	(0.471)	(0.008)	(0.000)	(0.173)	(0.000)	(0.563)	(0.000)	(0.000)	(0.155)
			C					**GS**		
Skewness	-4.154	0.059	0.080	-0.203	0.059	-0.415	-0.015	-0.043	-0.064	-0.004
Kurtosis	48.213	3.079	3.279	3.824	3.163	8.794	2.842	3.161	3.092	2.931
JB	243634	2.317	11.927	97.259	4.705	3950.383	2.959	3.821	2.884	0.558
p value	(0.000)	(0.314)	(0.003)	(0.000)	(0.095)	(0.000)	(0.228)	(0.148)	(0.236)	(0.757)
			JPM					**MS**		
Skewness	-0.317	0.006	-0.027	-0.019	-0.001	-0.923	-0.004	-0.007	-0.016	0.014
Kurtosis	13.679	2.950	3.077	3.369	2.972	19.800	2.851	3.043	3.023	2.869
JB	13193	0.303	1.032	15.902	0.092	32931	2.561	0.234	0.183	2.062
p value	(0.000)	(0.860)	(0.597)	(0.000)	(0.955)	(0.000)	(0.278)	(0.890)	(0.913)	(0.357)
			TRV					**WFC**		
Skewness	0.586	0.079	0.162	0.143	0.079	-0.696	0.110	0.103	0.033	0.110
Kurtosis	11.479	3.263	3.829	3.673	3.366	19.146	3.072	3.234	3.643	3.093
JB	8446.766	10.870	91.326	61.716	18.320	30278	6.133	11.169	48.231	6.610
p value	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.047)	(0.004)	(0.000)	(0.037)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.5: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector HC)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			ABT					**AMGN**		
Skewness	-0.102	0.016	0.054	0.009	0.043	-0.649	-0.069	-0.091	-0.171	-0.067
Kurtosis	5.670	3.075	3.227	3.251	3.162	6.551	2.955	3.075	3.229	2.997
JB	826.948	0.760	7.289	7.316	3.862	1648.085	2.430	4.475	19.544	2.079
p value	(0.000)	(0.684)	(0.026)	(0.026)	(0.145)	(0.000)	(0.297)	(0.107)	(0.000)	(0.354)
			BSX					**GILD**		
Skewness	-0.078	0.034	0.053	0.030	0.037	-0.104	-0.030	-0.037	0.002	-0.027
Kurtosis	7.823	2.947	3.135	3.151	3.032	8.268	3.073	3.181	3.504	3.125
JB	2684.533	0.847	3.384	3.061	0.733	3204.453	1.028	4.413	29.321	2.144
p value	(0.000)	(0.655)	(0.184)	(0.216)	(0.693)	(0.000)	(0.598)	(0.110)	(0.000)	(0.342)
			HUM					**JNJ**		
Skewness	-0.443	0.069	0.066	-0.002	0.086	-0.082	-0.033	-0.022	-0.093	-0.018
Kurtosis	7.472	2.980	3.109	3.179	3.010	8.532	3.027	3.201	3.389	3.072
JB	2396.356	2.225	3.393	3.688	3.459	3532.019	0.598	4.888	21.391	0.743
p value	(0.000)	(0.329)	(0.183)	(0.158)	(0.177)	(0.000)	(0.742)	(0.087)	(0.000)	(0.690)
			MDT					**MRK**		
Skewness	-0.524	-0.097	-0.060	-0.177	-0.091	0.109	-0.046	-0.070	-0.081	-0.089
Kurtosis	7.507	2.996	3.306	3.347	3.166	7.953	3.157	3.233	3.614	3.371
JB	2468.394	4.300	12.478	28.348	6.984	2833.996	3.807	8.505	46.578	19.522
p value	(0.000)	(0.116)	(0.002)	(0.000)	(0.030)	(0.000)	(0.149)	(0.014)	(0.000)	(0.000)
			PFE					**UNH**		
Skewness	0.151	0.018	0.046	0.055	0.041	-0.090	-0.099	-0.060	-0.142	-0.102
Kurtosis	4.759	2.768	2.856	2.901	2.917	13.564	3.040	3.358	3.401	3.094
JB	367.267	6.360	3.374	2.492	1.572	12871	4.706	16.434	27.816	5.794
p value	(0.000)	(0.042)	(0.185)	(0.288)	(0.456)	(0.000)	(0.095)	(0.000)	(0.000)	(0.055)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.6: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector IND)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			BA					**CAT**		
Skewness	0.023	0.019	0.011	0.023	0.018	-0.120	0.016	0.009	0.011	0.030
Kurtosis	5.273	3.153	3.254	3.282	3.178	4.960	2.850	3.022	2.970	2.929
JB	595.923	2.850	7.464	9.371	3.800	449.434	2.709	0.091	0.154	0.990
p value	(0.000)	(0.241)	(0.024)	(0.009)	(0.150)	(0.000)	(0.258)	(0.956)	(0.926)	(0.610)
			CMI					**GD**		
Skewness	-0.222	0.091	0.070	0.066	0.090	-0.469	-0.058	-0.029	-0.120	-0.057
Kurtosis	7.445	3.058	3.396	3.168	3.088	6.752	3.040	3.334	3.327	3.170
JB	2300.497	4.175	20.347	5.276	4.671	1724.268	1.726	13.277	19.013	4.868
p value	(0.000)	(0.124)	(0.000)	(0.072)	(0.097)	(0.000)	(0.422)	(0.001)	(0.000)	(0.088)
			GE					**HON**		
Skewness	-0.642	0.100	0.104	0.035	0.106	-0.870	-0.085	-0.139	-0.206	-0.073
Kurtosis	9.912	2.885	2.976	3.132	2.896	9.514	3.037	3.252	3.481	3.063
JB	5698.019	6.117	5.100	2.583	6.446	5240.337	3.478	16.203	46.153	2.931
p value	(0.000)	(0.047)	(0.078)	(0.275)	(0.040)	(0.000)	(0.176)	(0.000)	(0.000)	(0.231)
			LUV					**MMM**		
Skewness	-0.158	0.134	0.140	0.116	0.136	0.213	0.073	0.058	0.082	0.063
Kurtosis	6.596	3.076	3.172	3.199	3.108	7.010	3.146	3.327	3.402	3.193
JB	1502.603	8.927	12.374	10.740	9.860	1874.683	4.915	13.878	21.723	6.148
p value	(0.000)	(0.012)	(0.002)	(0.005)	(0.007)	(0.000)	(0.086)	(0.001)	(0.000)	(0.046)
			UPS					**UTX**		
Skewness	-0.057	-0.071	-0.089	-0.092	-0.067	-0.380	-0.083	-0.140	-0.143	-0.071
Kurtosis	6.588	3.222	3.514	3.551	3.298	8.946	3.094	3.480	3.529	3.113
JB	1485.757	8.024	34.082	38.909	12.315	4142.965	4.163	35.590	41.652	3.803
p value	(0.000)	(0.018)	(0.000)	(0.000)	(0.002)	(0.000)	(0.125)	(0.000)	(0.000)	(0.149)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.7: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector IT)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			AAPL					**CSCO**		
Skewness	-0.077	0.135	0.156	0.088	0.150	-0.161	0.019	0.040	-0.026	0.028
Kurtosis	5.175	2.996	3.020	3.047	2.999	7.548	2.882	2.963	3.069	2.886
JB	548.286	8.440	11.294	3.859	10.414	2396.376	1.780	0.899	0.854	1.873
p value	(0.000)	(0.015)	(0.004)	(0.145)	(0.005)	(0.000)	(0.411)	(0.638)	(0.652)	(0.392)
			DELL					**EMC**		
Skewness	0.072	-0.037	-0.050	-0.015	-0.031	-0.057	-0.008	-0.012	-0.045	0.015
Kurtosis	6.300	2.941	2.916	3.180	2.969	6.417	2.876	2.950	3.163	2.905
JB	1258.141	1.017	1.949	3.847	0.560	1347.722	1.805	0.354	4.016	1.156
p value	(0.000)	(0.602)	(0.377)	(0.146)	(0.756)	(0.000)	(0.406)	(0.838)	(0.134)	(0.561)
			HPQ					**IBM**		
Skewness	0.008	0.051	0.009	-0.030	0.046	-0.229	-0.014	-0.029	-0.059	0.017
Kurtosis	5.968	3.070	3.097	3.250	3.071	6.164	2.948	3.026	3.132	2.964
JB	1015.846	1.787	1.129	7.630	1.533	1178.095	0.406	0.479	3.619	0.289
p value	(0.000)	(0.409)	(0.569)	(0.022)	(0.465)	(0.000)	(0.816)	(0.787)	(0.164)	(0.865)
			INTC					**MSFT**		
Skewness	-0.236	0.029	0.025	-0.012	0.041	-0.096	0.097	0.110	0.041	0.085
Kurtosis	5.444	2.881	2.885	2.940	2.880	5.631	2.905	2.933	3.010	2.931
JB	714.355	2.025	1.800	0.486	2.433	802.083	5.374	6.085	0.792	3.860
p value	(0.000)	(0.363)	(0.407)	(0.784)	(0.296)	(0.000)	(0.068)	(0.048)	(0.673)	(0.145)
			ORCL					**XRX**		
Skewness	-0.291	-0.017	0.022	-0.111	-0.008	-0.693	0.016	0.061	-0.061	-0.010
Kurtosis	8.038	2.909	2.967	3.304	2.932	13.795	3.138	3.323	3.745	3.527
JB	2965.802	1.085	0.346	16.375	0.563	13656.5	2.328	13.708	65.669	32.074
p value	(0.000)	(0.581)	(0.841)	(0.000)	(0.755)	(0.000)	(0.312)	(0.001)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.8: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector MAR)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			AA					AKS		
Skewness	-0.844	0.046	0.025	0.001	0.059	-0.495	0.063	0.076	-0.006	0.062
Kurtosis	9.224	2.924	3.031	3.081	2.955	7.437	3.292	3.511	3.513	3.345
JB	4794.013	1.643	0.404	0.760	1.853	2382.961	11.690	32.734	30.324	15.510
p value	(0.000)	(0.440)	(0.817)	(0.684)	(0.396)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)
			DD					DOW		
Skewness	-0.002	0.133	0.164	0.114	0.156	-0.114	0.090	0.116	0.074	0.085
Kurtosis	6.578	2.962	3.135	3.105	3.028	7.507	3.055	3.189	3.353	3.124
JB	1475.566	8.297	14.537	7.307	11.310	2348.300	4.049	10.269	16.855	5.120
p value	(0.000)	(0.016)	(0.001)	(0.026)	(0.004)	(0.000)	(0.132)	(0.006)	(0.000)	(0.077)
			FCX					IP		
Skewness	-0.478	0.025	0.068	-0.029	0.075	-0.508	-0.079	-0.082	-0.112	-0.095
Kurtosis	6.090	3.095	3.652	3.127	3.212	9.835	2.985	3.092	3.158	3.079
JB	1206.029	1.325	51.088	2.243	7.754	5504.859	2.901	4.067	8.640	4.932
p value	(0.000)	(0.516)	(0.000)	(0.326)	(0.021)	(0.000)	(0.234)	(0.131)	(0.013)	(0.085)
			NEM					NUE		
Skewness	0.114	0.092	0.132	0.118	0.112	-0.437	-0.073	-0.172	-0.164	-0.117
Kurtosis	6.248	2.963	3.116	3.086	3.016	8.466	3.431	4.524	3.709	3.751
JB	1222.135	4.080	9.634	7.332	5.769	3532.643	23.841	281.503	70.368	71.347
p value	(0.000)	(0.130)	(0.008)	(0.026)	(0.056)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			WY					X		
Skewness	-0.562	0.050	0.048	-0.001	0.061	-0.519	0.067	0.065	-0.009	0.053
Kurtosis	9.198	3.061	3.222	3.200	3.103	6.436	3.105	3.346	3.147	3.151
JB	4575.020	1.558	6.753	4.620	2.924	1485.7	3.315	15.741	2.534	3.940
p value	(0.000)	(0.459)	(0.034)	(0.099)	(0.232)	(0.000)	(0.191)	(0.000)	(0.282)	(0.139)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.9: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector TEL)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			AMT					BT		
Skewness	-1.196	-0.052	-0.122	-0.157	-0.070	-0.120	0.012	-0.136	-0.019	0.039
Kurtosis	18.028	2.851	3.256	3.353	2.974	7.590	2.997	4.984	3.293	2.995
JB	26696.410	3.820	14.454	25.699	2.367	2435.569	0.069	462.347	10.048	0.690
p value	(0.000)	(0.148)	(0.001)	(0.000)	(0.306)	(0.000)	(0.966)	(0.000)	(0.007)	(0.708)
			CTL					FTR		
Skewness	-0.548	-0.181	-0.181	-0.230	-0.178	-0.788	0.006	0.078	-0.001	-0.014
Kurtosis	8.579	3.049	3.266	3.445	3.175	13.994	3.034	3.428	3.414	3.255
JB	3726.632	15.349	23.351	47.219	18.112	14222.140	0.146	23.879	19.778	7.561
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.930)	(0.000)	(0.000)	(0.023)
			Q					S		
Skewness	-0.756	-0.005	-0.069	-0.012	0.008	-1.239	0.061	0.062	-0.023	0.078
Kurtosis	20.522	3.143	3.537	3.623	3.188	16.147	2.968	3.287	3.434	3.056
JB	35660.170	2.357	35.413	44.782	4.096	20635.750	1.808	11.279	22.006	3.194
p value	(0.000)	(0.308)	(0.000)	(0.000)	(0.129)	(0.000)	(0.405)	(0.004)	(0.000)	(0.202)
			T					TEF		
Skewness	0.147	0.064	0.096	0.067	0.048	0.091	-0.220	-0.244	-0.277	-0.233
Kurtosis	5.744	2.938	3.082	3.111	3.002	11.158	3.026	3.241	3.434	3.255
JB	877.776	2.326	5.048	3.480	1.059	7676.652	22.446	34.059	57.187	32.598
p value	(0.000)	(0.312)	(0.080)	(0.175)	(0.589)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			VOD					VZ		
Skewness	0.056	-0.029	-0.049	-0.044	-0.029	0.406	0.108	0.120	0.124	0.106
Kurtosis	7.939	3.116	3.196	3.463	3.103	7.181	3.017	3.079	3.258	3.037
JB	2813.932	1.943	5.535	25.662	1.622	2091.3	5.377	7.363	14.791	5.357
p value	(0.000)	(0.378)	(0.063)	(0.000)	(0.444)	(0.000)	(0.068)	(0.025)	(0.001)	(0.069)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Table C.10: Returns standardized by EWMA filtered realized volatility measures at 5-sec sampling frequency (Sector UTL)

	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$	$r_{t,MA}$	$\frac{r_{t,MA}}{\sqrt{rv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{rr_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{pv_{t,MA}}}$	$\frac{r_{t,MA}}{\sqrt{bv_{t,MA}}}$
			AEP					CEG		
Skewness	-1.440	-0.049	-0.114	-0.123	-0.058	-6.838	-0.181	-0.156	-0.354	-0.148
Kurtosis	24.940	3.394	3.836	3.895	3.501	119.829	3.491	3.762	4.252	3.456
JB	56452	19.022	86.613	99.436	30.483	1595188	42.901	78.233	238.440	34.003
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			DUK					ETR		
Skewness	-0.042	0.112	0.078	0.065	0.110	-0.491	-0.174	-0.196	-0.239	-0.194
Kurtosis	7.685	3.276	3.506	3.500	3.331	8.668	3.223	3.455	3.464	3.246
JB	2531.504	14.538	32.366	30.740	18.161	3815.704	19.665	41.524	51.153	24.258
p value	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			EXC					OKE		
Skewness	-0.038	-0.168	-0.148	-0.195	-0.181	-0.789	-0.058	-0.057	-0.111	-0.037
Kurtosis	8.330	3.146	3.416	3.370	3.253	11.534	2.923	3.348	3.121	3.116
JB	3276.259	15.429	30.090	33.313	22.501	8684.577	2.236	15.429	7.335	2.186
p value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.327)	(0.000)	(0.026)	(0.335)
			PCG					PEG		
Skewness	-2.907	-0.063	-0.065	-0.215	-0.067	-0.273	-0.008	-0.037	-0.067	0.023
Kurtosis	55.815	3.287	3.642	4.881	3.573	8.931	3.507	4.081	3.865	3.626
JB	325493	11.302	49.430	429.204	39.868	4090.587	29.702	135.461	88.361	45.473
p value	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
			PGN					SO		
Skewness	-0.027	-0.083	-0.069	-0.080	-0.078	0.101	0.105	0.080	0.092	0.116
Kurtosis	9.022	3.265	3.560	3.493	3.244	7.598	3.483	3.728	3.828	3.546
JB	4181.547	11.322	38.320	30.971	9.717	2441.8	31.927	64.023	82.869	40.538
p value	(0.000)	(0.003)	(0.000)	(0.000)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note: This table reports the skewness, kurtosis and Jarque-Bera test statistics of normality. $rv_{t,MA}$, $rr_{t,MA}$, $pv_{t,MA}$, $bv_{t,MA}$, denote EWMA filtered realized variance, realized range, realized power variation, and realized bipower variation, respectively. The sampling frequency is 5 minute.

Concluding Remarks

This thesis compares the degree of validity of stylized properties of four extensively studied realized measures of volatility using a series of econometric models. In this context, the main research topics examined include: a) Optimal sampling frequency; b) Impact of jumps; c) Correlation relationship; d) Leverage and volatility feedback effect; e) Volatility regimes; f) Volatility-volume relation; g) The distributional properties of realized volatility measures and returns; and h) Recovering return normality.

The thesis uses a unique data set that corrects the bias and errors presented in the widely used TAQ. We choose 100 stocks traded in the US equity market and segment them into 10 market sectors. An 11-year sample period which covers two crises is considered. Using this extensive data, we aim to find out which realized volatility measure(s) is the best proxy of the true integrated variance. Furthermore, the analysis based on sector segmentation also has implications for practitioners on the aspects of risk management and other financial investments.

Chapter 1 provides a general discussion of the above topics. Overall, we confirm many of the findings of the existing literature: 5-minute is the optimal sampling frequency for at least realized variance and realized bipower variation. Realized volatility measures exhibit similar properties. They are highly leptokurtic and are best described by Lognormal and Inverse Gaussian distributions. Both distributions provide almost indistinguishable empirical fits. Returns standardized by realized volatility measures are (nearly) Gaussian. Realized volatility measures are highly positively correlated with each other. On the other hand, the regime characteristics vary by both regime type (high-low) and by sector according to the Markov Regime Switching model. The presence of jumps in the realized variance has little impact on the properties of realized variance. Jumps have distinctively different properties to realized measures of volatility. Jumps are small, short-memory, and do not have the leverage and feedback effects that are found in realized volatility measures. At sector level, the proportion of detected significant jumps does not vary much across sectors. However, the sectors which are highly actively traded contain fewer jumps and are more likely to have lower jump contribution to total volatility. The less actively traded sectors are detected with more jumps and reports higher jump contribution to total volatility.

The stylized facts being addressed tend to vary widely across sectors. For example, the most actively traded sector, IT, shows the highest persistence in realized volatility measures according to the Robinson's "*d*" long memory test, suggesting a potential better performance in forecasting than the least actively traded sector UTL, which reports the lowest persistence. The volatility-volume relation also holds stronger in those actively traded sectors such as IT and FIN, and has the highest model rejection in the least traded sector, UTL.

Among the four realized volatility measures, realized power variation shows the best performance across sectors and outperforms the other realized volatility measures. It has the lowest standard deviation, well-defined distributional properties, and the highest degree of persistence. It is the most capable of recovering normality and also robust to jumps. Our findings here are in line with Ghysels *et al* (2006) where the superiority of the power variation has also been highlighted. Realized bipower variation, another jump-robust volatility measure, also shows consistent performance among all tests. Realized range, the only volatility measure in this thesis that is constructed from intraday highest and lowest prices, shows the greatest diversity of all four realized volatility measures. Finally, the most extensively discussed realized variance provides the least satisfactory results: For instance, it provides lowest estimation results in the volatility-volume relation and shows the highest rejection of normality test when used as a standardization factor.

The following two chapters extend the analysis of Chapter 1. **Chapter 2**, which discusses the volatility- volume relation under the framework of Mixture of Distributions Hypothesis (MDH), contributes to the literature as follows: First, we identify strong evidence that volatility-volume relation holds under the various realized volatility measures. The test results are generally stronger than in many recent empirical studies which also look at this relationship using realized volatility. Second, the results indicate that the number of trades is a superior market information proxy to trading volume. In particular, the number of trades that takes place either at higher or lower prices is the most capable of explaining the realized volatility measures.

In **Chapter 3**, we analyze the distributional properties of volatility standardized daily returns. The motivation of this chapter is drawn from the well-documented stylized fact in the literature that realized volatility measures standardized returns are (nearly) Gaussian. A great many researchers also record the violation of the Gaussianity of standardized returns. Possible explanations are mainly founded upon the existence of jumps and/or market microstructure noise. To identify which are more significant reasons driving the standardized returns from standard normal, we construct the realized volatility measures from two sampling frequencies and find that the presence of jumps could be a reason to reject the normality, but only at ultra-high sampling frequency. At the optimal sampling frequency (5 minute), jumps have little impact on recovering the return normality. Instead, market microstructure noise is likely to be a main reason. To smooth the microstructure noise, we propose a moving average filter on realized volatility measures, the microstructure noise in the realized variance as well as realized bipower variation was effectively filtered. Filtered realized volatility measures have stronger power to achieve the stock normality in some cases. However, the moving average filter falls to enhance the power of realized power variation on recovering returns normality. Moving average filtration does not apply for the ultra-high sampled realized volatility measures. Our results also provide evidence that more actively traded stocks (by volume) contain more market microstructure noise and fewer jumps. Filtration on realized volatility measures works better on stocks which contain a higher degree of market microstructure noise.

Bibliography

Aït –Sahalia, Y., Mykland, P.A. & Zhang, L. (2005). "How often to sample a continuous-time process in the presence of market microstructure noise", *Review of Financial Studies*, 18, 351–416.

Aït-Sahalia, Y. & Jacod, J. (2009a). "Estimating the degree of activity of jumps in high-frequency data", *The Annals of Statistics*, 37, 2202-2244.

Aït-Sahalia, Y. & Jacod, J. (2009b). "Testing whether jumps have finite or infinite activity", *The Annals of Statistics*, 39, 1689-1719.

Aït-Sahalia, Y. & Jacod, J. (2010). "Is Brownian motion necessary to model high-frequency data?" *The Annals of Statistics*, 38, 3093-3128.

Aït-Sahalia, Y. & Yu, J. (2009). "High frequency market microstructure noise estimates and liquidity measures", *The Annals of Applied Statistics*, 3, 422-457.

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., Bollerslev, T. & Diebold, F.X. (2007a). "Roughing it up: Including jump components in the measurement, modeling and forecasting", *The Review of Economics and Statistics,* 89(4), 701-720.

Andersen, T.G., Bollerslev, T., Dobrev, D. (2007b). "No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: theory and testable distributional implications", *Journal of Econometrics*, 138,125–180.

Andersen, T.G., Bollerslev, T., Frederiksen, P. & Nielsen, M.Ø. (2010). "Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns", *Journal of Applied Econometrics,* 25(2), 233-261.

Andersen, T.G, Bollerslev, T., Diebold, F.X. & Labys, P. (1999). "Realized volatility and correlation", *LN Stern School of Finance Department Working Paper*.

Andersen T.G., Bollerslev, T., Diebold, F.X. & Labys, P. (2000). "Exchange rate returns standardized by realized volatility are (nearly) Gaussian", *Multinational Finance Journal,* 4,159–179.

Andersen, T.G., Bollerslev, T., Diebold, F.X. & Labys, P. (2001). "The distribution of exchange rate volatility", *Journal of American Statistical Association*, 96, 42-55.

Andersen, T.G, Bollerslev, T., Diebold, F.X. & Ebens, H. (2001). "The distribution of realized stock return volatility", *Journal of Financial Econometrics,* 61, 43-76.

Andersen, T.G., Bollerslev, T., Diebold, F.X. & Labys, P. (2003). "Modelling and forecasting realized volatility", *Econometrica,* 71(2), 579-625.

Andersen, T.G., Bollerslev, T. & Huang, X. (2011)." A reduced form framework for modeling volatility of speculative prices based on realized variation measures", *Journal of Econometrics* , 160, 176-189.

Ané, T. & Geman, H. (2000). "Order flow, transaction clock, and normality of asset returns", *The Journal of Finance*, 55(5), 2259-2284.

Ané, T. & Metáis, C. (2010). "Jumps distribution characteristics: evidence from European stock markets", *International Journal of Business and Economics*.9, 1-22.

Areal, N. & Taylor, S.J. (2002). "The realized volatility of FTSE-100 futures prices", *Journal of Futures Markets*, 22(7), 627-648.

Bajgrowicz, P. & Scaillet, O. (2011). "Jumps in high-frequency data: spurious detections, dynamics, and news". *Swiss Finance Institute Occasional Paper No.11-36.*

Baillie, R., Bollerslev, T. & Mikkelsen, M. (1996). "Fractionally integrated generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 74, 3-30.

Bandi, F. & Perron, M. (2006). "Long memory and the relation between implied and realized volatility", *Journal of Financial Econometrics*, 4, 4, 636-670.

Bandi, F. & Russell, J. (2006). "Separating microstructure noise from volatility", *Journal of Financial Economics*, 79(3), 655-692.

Bandi, F. & Russell, J. (2008). "Microstructure noise, realized variance, and optimal sampling", *Review of Economic Studies*,75,339–369.

Barndorff-Nielsen, O.E. (2001). "Modelling by Lévy Preocesses", *Lecture Notes-Monograph Seires*, 25-31.

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

Barndorff -Nielsen, O.E. & Shephard, N. (2003). "Realized power variation and stochastic Volatility", *Bernoulli,* 9, 243-265.

Barndorff-Nielsen, O.E. & Shephard, N. (2004). "Power and bipower variation with stochastic volatility and jumps (with discussion)", Journal *of Financial Econometrics*, 2(1), 1-37.

Barndorff-Nielsen, O.E. & Shephard, N. (2006). "Econometrics of testing for jumps in financial economics using bipower variation", *Journal of Financial Econometrics*, 4, 1-30.

Barndorff-Nielsen, O.E., Hansen, P, Lunde, A. & Shephard, N. (2008). "Designing realized kernels to measure the *ex-post* variation of equity prices in the presence of noise", *Econometrica*, 76, 1481-1536.

Bessembinder, H. & Seguin, P. (1993). "Volatility, trading volume and market depth; Evidence from the futures markets", *Journal of Financial and Quantitative Analysis*, 29, 21-39.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

Bollerslev, T., Kretschmer, U., Pigorsch, C. & Tauchen, G. (2009). "A discrete-time model for daily S & P500 returns and realized variations: Jumps and leverage effects", *Journal of Econometrics* , 150, 151-166.

Bollerslev, T., Litvinova, L. &Tauchen, G. (2006). "Leverage and Volatility Feedback Effects in High-Frequency Data", *Journal of Financial Econometrics*, 4, 353–384.

Boudt, K., Croux, C. & Laurent, S. (2008). "Robust estimation of intraweek periodicity in volatility and jump detection", *Working Paper.*

Brooks, C. (1998). "Predicting stock index volatility: Can market volume help?", *Journal of Forecasting*, 17, 59-80.

Brownlees, C. & Gallo, G. (2006). "Financial econometric analysis at ultra-high frequency: Data handling concerns", *Computational Statistics & Data Analysis*, 51(4), 2232-2245.

Brownlees, C. & Gallo, G. (2010). "Comparison of volatility measures: a risk management perspective", *Journal of Financial Econometrics*, 8(1), 29-56.

Chan, C. & Fong, W.M. (2006). "Realized volatility and transactions", *Journal of Banking & Finance*, 30, 2063–2085.

Chan, K. & Fong, W.M. (2000). "Trade size, order imbalance, and the volatility–volume relation", *Journal of Financial Economics*, 57, 247–273.

Chen, G., Firth, M. & Rui, O. (2001). "The dynamic relations between stock returns, trading volume and volatility", *Financial Review*, 38,153–174.

Chiang, T. C., Qiao, Z. & Wong, W.K. (2010). "New evidence on the relation between return volatility and trading volume", *Journal of Forecasting*, 29, 202-515.

Chevallier, J. & Sevi, B. (2011). "On the realized volatility of the EXC CO2 emissions 2008 futures contract: distribution, dynamics and forecasting", *Annals of Finance*, 7, 1-29.

Christensen, K. & Podolskij, M. (2007). "Realized range-based estimation of integrated variance", *Journal of Econometrics,* 141, 323-349.

Clark, P. K. (1973). "A subordinated atochastic orocess model with finite variance for speculative prices", *Econometrica,* 411, 135-155.

Conti, R. (2000). "Empirical properties of asset returns: stylized facts and stastical issues", *Quantitative Finance*, 1, 223-236.

Corsi, F., Audrino, F. & Reno, R. (2012). "HAR modeling for realized volatility forecasting", *Handbook of Volatility Models and Their Applications*, Wiley.

Darrat, A., Rahman, F. & Zhong, M. (2003). "Intraday trading volume and return volatility of the DJIA stocks: a note", *Journal of Banking and Finance*, 27, 2035–2043.

Dickey, D. & Fuller, W. (1979). "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association*, 84, 427-431.

Ding, Z., Granger, C. & Engle, R. F. (1993). "A long memory property of stock returns and a new model", Journal *of Empirical Finance,* 1, 83-106.

Dobrev, D. & Szerszen, P. (2010). "The information content of high-frequency data for estimating equity return models and fForecasting risk", *Working Paper.*

Doornik, J. A. & Hendry, D. F. (2009). *Econometric Modelling - PCGive 13: Volume II*, Timerlake Consultants.

Donaldson, G. & Kamstra, M. (2004). "Volatility forecasts, trading volume, and the ARCH versus option-implied volatility trade-off", *Working Paper 2004-6.*

Dufour, J., Garcia, R. & Taamouti, A. (2011). "Measuring high-frequency causality between returns, realized volatility and implied volatility", *Working paper.*

Easley, D. & O'Hara, M. (1987). "Price, trade size, and information in securities markets", *Journal of Financial Economics*, 19, 69–90.

Eberlein, E. & Keller, U. (1995). " Hypobolic distributions in finance", *Bernoulli*, 1, 281-299.

Engle, R.F. (1982). "Autoregressive conditional heteroskedasticity with estimates of the Variance of U.K. Inflation", *Econometrica*, 50, 987-1008.

Eraker, B., Johannes , M. & Polson, N. (2003). "The impact of jumps in volatility and returns", *Journal of Finance*, 58, 1269-1300.

Epps T.W. & Epps M.L. (1976). "The stochastic dependence of security price changes and transaction volumes: implications for the mixture of distributions hypothesis", *Econometrica* 44,305–321.

Fama, E. (1965). "The behaviour of stock market prices", *Journal of Business*, 36, 420-429.

Forsberg, L. & Bollerslev, T. (2002). " Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the EURO): the GARCH-Normal Inverse Gaussian model", *Journal of Applied Econometrics*, 17, 535-548.

Fleming, J. & Paye, B. (2007). "The impact of microstructure noise on the distributional properties of daily stock returns standardized by realized volatility", *Woking Paper.*

Fleming, J. & Paye, B. (2011). "High-frequency returns, jumps and the mixture of normals hypothesis", *Journal of Econometrics*, 160, 119-128.

Fuertes, A.M., Izzeldin, M. & Kalotychou, E. (2009). "On forecasting daily stock volatility: the role of intraday information and market conditions", *International Journal of Forecasting,* 25, 259-281.

Gallant, A., Hsieh, D. & Tauchen, G. (1992). "Stock prices and volume", *Review of Financial Studies*, 5,199–242.

Geweke, J. & Porter-Hudak, S. (1983). "The estimation and application of long memory time series models", *Journal of Time Series Analysis*, 4, 221–238.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2006). "Predicting volatility: Getting the most out of return data sampled at different frequencies", *Journal of Econometrics*, 131, 59-95.

Giot, P. & Laurent, S. (2004). "Modelling daily Value-at-Risk using realized volatility and ARCH type models", *Journal of Empirical Finance*, 11, 379-398.

Giot, P., Laurent, S. & Petijean, M. (2010). "Trading activity, realized volatility and jumps", *Journal of Empirical Finance*, 17, 1, 168-175.

Giraitis, L., Kokoszka, P., Leipus, R. & Teyssiere, G. (2003). "Rescaled variance and related tests for long memory in volatility and levels", *Journal of econometrics*, 112. 265-294.

Granger, C. & Ding, Z. (1995). "Some properties of absolute return: An alternative measure of risk", *Annales d'Economie et de Statistique*, 40, 67-9.

Hamilton, J. (1994). "*Time Series Analysis*", Princeton University Press.

Harris, L. (1987). "Transaction data test of the mixture of distributions hypothesis", *Journal of Financial and Quantitative Analysis*, 22,127–141.

Hansen, P. & Lunde, A. (2004). "An unbiased measure of realized variance", *Working Paper*.

Hansen, P. & Lunde, A. (2006). "Realized variance and market microstructure noise", *Journal of Business and Economic Statistics.* 24, 127-161.

Hansen, P., Large, J. & Lunde, A. (2006). "Moving average-based estimators of integrated variance", *Working Paper.*

Huang, X. & Tauchen, G. (2005). "The relative contribution of jumps to total price variation", *Journal of Financial Econometrics*, 3, 456–499.

Hull, J. & White, A. (1987). "The pricing of options on assets with stochastic volatilities", *Journal of Finance*, 42, 281-300.

Izzeldin, M. (2007). "Trading volume and the number of trades: A comparative Study using high frequency data", *Working Paper.*

Izzeldin, M. & Murphy, A. (2010). "Recovering the moments of information flow and normality of asset returns", *Applied Financial Economics*, 20, 761-769.

Jones, C., Kaul, G. & Lipson, M. (1994). "Transactions, volume and volatility", *Review of Financial Studies*, 7, 631–651.

Karpoff, J. (1987). "The relation between price changes and trading volume: A survey", *Journal of Financial and Quantitative Analysis,* 22, 109-126.

Khalifa, A., Miao, H. & Ramchander, S. (2011). "Return distributions and volatility forecasting in metal futures markets: Evidence from gold, silver and copper", *Journal of Futures Markets*, 21, 55-80.

Kwiatkowski, D., Phillips, P., Schmidt, P. & Shin, P. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics*, 54, 159–178.

Lamoureux, C. & Lastrapes, W. (1990). "Heteroskedasticity in stock return data: Volume versus GARCH effects", *Journal of Finance*, 45, 221-229.

Lee, S.S. & Mykland, P. (2008). "Jumps in financial markets: A new nonparametric test and jump dynamics", Review *of Financial Studies,* 21,6, 2535-2563.

Liesenfeld, R. (2001). "A generalized bivariate mixture model for stock price volatility and trading volume", *Journal of Econometrics*, 104, 141-178.

Liu, C. & Maheu, J. (2005). "Modeling and forecasting realized volatility: the role of power variation", *Working Paper.*

Lo, A.W. (1991). "Long term memory in stock market prices", *Econometrica*, 59, 1279–1313.

Louhchi, W. (2011). "What drives the volume-volatility relationship on Euronext Paris?" *International Review of Financial Analysis*, 20, 200-206.

Luu, J. & Martens, M. (2003). "Testing the mixture-of-distribution hypothesis using 'realized' volatility", *Journal of Futures Market*, 23, 661-679.

Malmsten, M. & Terasvirta, T. (2004). "Stylized facts of financial time series and three popular models of volatility", *Working Paper.*

Mandelbrot, B. (1963). "The variation of certain speculative prices", *The Journal of Business*, 36, 394-419.

Mandelbrot, B. & Taylor, H. (1967). "On the distribution of stock price difference", *Operations Research*, 15, 1057-1062.

Martens, M. & Dijk, D. (2007). "Measuring volatility with the realized range", *Journal of Econometrics,* 138, 181-207.

Maheu, J. & McCurdy, T. (2002). "Nonlinear features of realized FX volatility", *Review of Economics and Statistics.* 84, 668-681.

McAleer, M. & Medeiros, M. (2008). "Realized volatility: A review", *Econometric Reviews*, 27, 10-45.

Merton. R. (1976). "Option pricing when underlying stock returns are discontinuous", *Journal of Econometrics.* 3, 125-144.

Najand, M. & Yung, K. (1991). "A GARCH estimation of the relationship between volume and price variability in futures market", *Journal of futures markets*, 11, 613-621.

Nelsen, K. (2010). *Financial Econometrics Notes*, Oxford University.

Osborne, M. F. (1959). "Brownian Motion in the Stock Market", *Operations Research*, 7, 145-173.

Owens, J. P. & Steigerwald, D.S. (2006). "Noise reduced realized volatility: a Kalman filter approach", *Advances in Econometrics*, 20, 211-227.

Praetz, P. (1972). "The distribution of share price changes", *Journal of Business*, 45, 49-55.

Poon, S. & Granger, C. (2003). "Forecasting volatility in financial markets: A review", *Journal of Economic Letters*, 41, 478-539.

Richardson, M. & Smith, T. (1994). "A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information", *Journal of Financial and Quantitative Analysis*, 29, 101-116.

Robinson, P. M. (1995). "Gaussian semiparametric estimation of long range dependence", *Annals of Statistics,* 23, 1630-1661.

Robinson, P.M. & Henry, M. (1999). "Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels", *Econometric Theory*, 15, 299–336.

Rogers, L. & Zhang, L. (2011). "Understanding asset returns", *Cambridge Working Paper.*

Schwert, G.W. (1990). "Stock volatility and the crash of 87", *Review of Financial Studies*, 3, 77-102.

Stentoft, L. (2008). "Option pricing using realized volatility", *Working Paper.*

Tauchen, G. & Pitts M. (1983). "The price variability-volume relationship on speculative markets", *Econometrica*, 51, 485-505.

Tauchen, G. & Zhou, H. (2011). "Realized jumps on financial markets and predicting credit spreads", *Journal of Econometrics* , 102-118.

Taylor, S.J. (1986). *Modeling Financial Time Series,* Chichester: John Wiley and Sons.

Terasvirta, T. (2006). Univariate nonlinear time series models, In Mills, T. and Patterson, K. (eds.), *Palgrave Handbook of Econometrics,* 396-424. Basingstoke: Palgrave MacMillan.

Tsay, R.S. (2005). *Analysis of Financial Time Series*, 2nd Edition, John Wiley & Sons, Inc., Publication.

Wang, T. & Huang, Z. (2012). "The relationship between volatility and trading volume in the Chinese stock market: A volatility decomposition perspective", *Annals of Economics and Finance*, 13, 211-236.

Westerfield, R. (1977). "The distribution of common stock price changes", *Journal of Quantative Analysis*, 12, 743-765.

Ying, C. (1966). "Stock market prices and volume of sales", *Econometrica,* 34, 676-686.

Zhang, L. (2006). "Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach", Bernoulli, 12, 1019–1043.

Zhang, L., Mykland, P.A. & ït-Sahalia, Y. (2005). "A tale of two time scales: determining integrated volatility with noisy high-frequency data", *Journal of the American Statistical Association*, 100, 1394–1411.

Zhou, B. (1996). "High-frequency data and volatility in foreign exchange rates", *Journal of Business and Economic Statistics*, 14, 45-52.

Zivot, E. & Andrews, D. (1992). "Further evidence of great crash, the oil price shock and unit root hypothesis", *Journal of Business and Economic Statistics*, 10, 251-2