New Considerations for Modeling Financial Volatility

CHU, Chun Fai Carlin

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Systems Engineering and Engineering Management

The Chinese University of Hong Kong
April 2011

Thesis / Assessment Committee

Professor Yu Xu, Jeffrey (Chair)

Professor Lam, Kai Pui (Thesis Supervisor)

Professor So Man-Cho, Anthony (Committee member)

Professor David Allen (External Examiner)

Abstract

This research study investigates three new considerations for improving the performance of volatility modeling of financial returns. Two of them are related to the intraday volatility modeling and the other one is about the use of overnight information for daily volatility modeling.

About the intraday volatility modeling, the limitations and potential problems of using Andersen & Bollerslev's approach are addressed and distinct modifications are proposed to tackle the corresponding issues. The first suggestion is about the utilization of the interaction between the intraday periodicity and the heteroskedasticity while the second is about the modified normalization for the estimation of the intraday periodicity.

The proposed modifications are tested with different ARCH structures, including GARCH(1,1), FIGARCH(1,d,1) and HYGARCH(1,d,1), by using simulated data and market data. Apart from studying the 1-step-ahead out-of-sample performance, several multiple-step-ahead forecasting results are also addressed. Under the same level of model flexibility (parameterized portions), our proposed modifications always outperform the original method in both in-sample fitness and out-of-sample performance on various forecasting horizons.

On the other hand, the third suggestion is about the inclusion of overnight information for the estimation of daily volatility. This study explores the possibility of incorporating the overnight variance indirectly through the use of linearly combined daily volatility estimators. The empirical results demonstrate that the inclusion of overnight variance can produce substantial influence when the minimum-variance constraints are relaxed. Besides, the influence is revealed to be not monotonic as an increase of the overnight proportion does not necessarily produce a larger influence.

Furthermore, it is demonstrated that the inclusion of overnight variance can improve the prediction accuracy of the Chicago Board of options Exchange (CBOE) volatility indexes (VIX and VXD) under specific weight combinations. The findings contradict the common perception that overnight return does not contain useful information for daily volatility modeling.

摘要

本研究提出三個新的考慮以改善金融波動模型的性能。其中兩個都涉及到日內的波動模型，另一個是關於使用整夜的金融信息以改善每日的波動模型。

本研究是透過探討 Andersen & Bollerslev 的日內波動模型的局限性和潛在的問題，從而提出相應的解決方法。第一個建議是關於善用日內的週期性和異方差之間的相互作用，而第二個是關於對估算日內的週期性之修改。

本研究使用模擬數據和市場數據對所提出之建議進行測試，當中之建議是經過不同的 ARCH 結構對之進行測試，包括 GARCH(1,1)，FIGARCH(1,d,1) 和 HYGARCH(1,d,1)。此外，本研究對所提出之建議進行不同預測範圍的性能分析，包括 1 步之預測及多步之預測。在相同的模型靈活性水平下（具備有相同的參數之模型），無論在樣本內和樣本外之性能都顯示本文所提出的建議是優於原本的方法。

另一方面，第三個建議是關於使用整夜的金融信息以改善每日的波動模型。本研究探討使用線性組合的方法作爲合併整夜金融信息的可能性。當最小方差約束被解除後，實證結果證明了合併之整夜金融信息是可以產生明顯的影響。而且，本研究還發現整夜的金融信息之影響並不是跟從線性的比例增加，較大比重之整夜信息並不一定產生較大的影響。

另外，分析結果還表明了經合併後之線性組合可以對芝加哥期權交易所（CBOE）波動指數（VIX 指數和 VXD）作出更準確的預測。此研究結果顯示了整夜的金融信息是可以對每日波動模型提供有富有意義的資訊。

Contents

1 Introduction 1
- 1.1 Thesis content . . . 1
- 1.2 Contributions . . . 8
- 1.3 Organization of the thesis . . . 10

2 Literature Review 13
- 2.1 Modeling the volatility of intraday returns . . . 14
  - 2.1.1 Characteristics of intraday returns . . . 14
  - 2.1.2 Models for intraday volatility modeling . . . 17
    - 2.1.2.1 Andersen & Bollerslev's sequential estimation model . . . 17
    - 2.1.2.2 Periodic Autoregressive Conditional Heteroskedasticity (P-GARCH) model . . . 21
    - 2.1.2.3 Engle's Multiplicative Component model . . . 22
- 2.2 Models for measuring daily volatility . . . 24
  - 2.2.1 Daily volatility estimators . . . 24
    - 2.2.1.1 Garman & Klass's daily volatility estimators . 25
    - 2.2.1.2 Hansen & Lunde's volatility estimators . . . 25
  - 2.2.2 Engle's Multiplicative Error Model (MEM) . . . 27
  - 2.2.3 Chicago Board of Exchange (CBOE) Volatility index . 30
- 2.3 Supplementary note . . . 31
  - 2.3.1 Flexible Fourier functional form for the approximation of periodic dynamics . . . 31

3 Integrated framework approach for intraday volatility modeling 33
- 3.1 Introduction and the research objective . . . 35
- 3.2 Andersen & Bollerslev's sequential intraday volatility model and its limitation . . . 38
- 3.3 Proposed Integrated framework approach . . . 41
- 3.4 Model specifications and Evaluation criteria . . . 45

3.4.1 Specifications of filtration process and ARCH structures 45
3.4.2 Evaluation criteria 48
3.5 Simulation study on the effectiveness of the integrated framework 52
3.6 Empirical investigation 57
3.6.1 In-sample fitness 59
3.6.2 Out-of-sample intraday forecasting performance 66
3.6.3 Out-of-sample 1-day-ahead forecasting performance . . 76
3.7 Concluding remarks 80
3.8 Supplementary note 82
3.8.1 Limitation of the integration framework approach for EGARCH(1,1) structure 82
3.8.2 Proposed integrated framework approach with the use of GARCH(1,1) structure - Maximum likelihood estimation method 85
3.8.3 Graphical representations on the out-of-sample forecast performance 90
3.8.4 Properties of the realized variance estimator 92

4 Day-varying structure for modeling intraday periodicity 98
4.1 Introduction and the research objective 100
4.2 Andersen & Bollerslev's intraday periodicity estimation and its limitation 103
4.3 Proposed normalization procedure for estimating intraday periodicity 108
4.4 Characteristics of the residuals of the approximation function . 109
4.4.1 Data description 109
4.4.2 Specification of the approximation function 110
4.4.3 Heteroskedasticity of the residuals 111
4.5 Simulation study on the robustness of the normalization procedures 113
4.6 Application on intraday volatility modeling 120
4.6.1 Data description and sampling method 120
4.6.2 Specification of Engle's Multiplicative Component ARCH model 121
4.6.3 Evaluation criteria 123
4.6.4 In-sample fitness 125
4.6.5 Out-of-sample forecasting performance 127
4.7 Concluding remarks 133
4.8 Supplementary note 135

4.8.1 Equivalent modeling performance by scaled intraday periodicity series . . . 135
4.8.2 Characteristics of the two normalization procedures for evaluating day-invariant periodicity . . . 136

5 Impact of the overnight information on daily volatility modeling 139
5.1 Introduction and the research objective . . . 141
5.2 Specifications of daily volatility estimators . . . 144
5.2.1 Garman & Klass's volatility estimators . . . 144
5.2.2 Realized variance and Hansen & Lunde's whole-day-based variance estimator . . . 145
5.3 Specification of Engle's Multiplicative Error Model . . . 148
5.4 Correlation analysis among models with and without overnight information . . . 150
5.4.1 Generalized whole-day-based variance formulation . . . 151
5.4.2 Preliminary study on the empirical data . . . 153
5.4.3 Results from models under minimum-variance assumptions . . . 155
5.4.4 Results from models under the generalized whole-day-based formulation . . . 158
5.5 Modeling CBOE volatility index using MEM volatility . . . 162
5.5.1 Regression models for VIX/VXD prediction . . . 162
5.5.2 Results from models with and without overnight information . . . 164
5.6 Concluding Remarks . . . 167
5.7 Supplementary note . . . 168
5.7.1 Analysis of the ratio of between On_t^2 and $\hat{\sigma}_4^2$ . . . 168
5.7.2 Analysis of the ratio of between On_t^2 and $\hat{\sigma}_{NW,t}^2$ . . . 169

6 Concluding remarks and further work 171

A Nomenclature 176

Bibliography 180

List of Figures

2.1 Correlogram of 10-minute NASDAQ returns 15
2.2 Correlogram of 10-minute NASDAQ absolute returns 16

3.1 Correlogram of the absolute intraday returns, 5-day period - (a) NASDAQ, (b) S&P 500 58
3.2 Block diagram of the two approaches 90
3.3 Improvement on R-square by using the integrated framework approach 91
3.4 Improvement on MSE by using the integrated framework approach 91
3.5 Improvement on MAE by using the integrated framework approach 92

4.1 Shape of the time series $x_{t,n}^{shape}$ 115
4.2 Expected values: Error-to-signal ratio=0.3, Mean=-2.6 118

5.1 Illustration of the scope of four estimators 144
5.2 Autocorrelation of the 10 minute intraday data - S&P 500 and DJIA (with 95% confidence level bounds) 154
5.3 Autocorrelation of the 30 minute intraday data - S&P 500 and DJIA (with 95% confidence level bounds) 155
5.4 ARCH(2) model fit measure under the generalized weighted approach 158
5.5 Ljung-Box(12) model fit measure under the generalized weighted approach 159
5.6 Comparison of correlation coefficient under the generalized weighted approach 160
5.7 Correlation coefficient under the generalized weighted approach ($\rho \geq 0.9$) 161
5.8 Adjusted R-square among the regression models 165
5.9 F statistics for testing zero coefficient 165
5.10 F statistics for testing join zero coefficients 166

List of Tables

3.1 Simulation results (10,000 repetitions for each situation) . . . 56
3.2 Descriptive statistics of the intraday returns (multiplied by 100) 59
3.3 In-sample fit - Deseasonalized by simple averaging 62
3.4 In-sample fit - Deseasonalized by FFF method 63
3.5 Diebold-Mariano test, In-sample fit for various ARCH structure - Deseasonalized by simple averaging 65
3.6 Diebold-Mariano test, In-sample fit for various ARCH structure - Deseasonalized by the FFF method 65
3.7 Out-of-sample fit, GARCH(1,1) - Deseasonalized by simple averaging 68
3.8 Out-of-sample fit, FIGARCH(1,d,1) - Deseasonalized by simple averaging 69
3.9 Out-of-sample fit, HYGARCH(1,d,1) - Deseasonalized by simple averaging 70
3.10 Out-of-sample fit, GARCH(1,1)-Deseasonalized by FFF method 71
3.11 Out-of-sample fit, FIGARCH(1,d,1) - Deseasonalized by FFF method 72
3.12 Out-of-sample fit, HYGARCH(1,d,1) - Deseasonalized by FFF method 73
3.13 Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by simple averaging 74
3.14 Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by FFF method 75
3.15 Out-of-sample fit, 1-step-ahead daily forecast - Deseasonalized by simple averaging 77
3.16 Out-of-sample fit, 1-step-ahead daily forecast - Deseasonalized by FFF method 78
3.17 Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by simple averaging 79

3.18 Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by the FFF method 79

4.1 Descriptive statistics of the intraday returns (multiplied by 100) . 110
4.2 Results of Bartlett's test and Brown-Forsythe's test 112
4.3 Heteroskedastic error - Correlation between the true series and the expected series 117
4.4 In-sample fit for GARCH(1,1) 126
4.5 In-sample fit for FIGARCH(1,d,1) 126
4.6 Out-of-sample fit for GARCH(1,1) - NASDAQ 128
4.7 Out-of-sample fit for GARCH(1,1) - S&P 500 129
4.8 Out-of-sample fit for FIGARCH(1,d,1) - NASDAQ 131
4.9 Out-of-sample fit for FIGARCH(1,d,1) - S&P 500 132

5.1 Descriptive statistics of squared overnight return (On_t^2), Squared Sigma 4 ($\hat{\sigma}_4^2$) and Newey-West realized variance ($\hat{\sigma}_{NW}^2$) 156
5.2 Diagnostics information for the MEM models 156
5.3 Combinations (λ) that give larger adjusted R-square and with F statistics larger than 95% confidence critical value 167

A.1 Nomenclature for Chapter 3 (Integrated framework approach) 177
A.2 Nomenclature for Chapter 4 (Day-varying structure) 178
A.3 Nomenclature for Chapter 5 (Impact of the overnight information) 179

Chapter 1

Introduction

1.1 Thesis content

This research study investigates three new considerations for improving the performance of volatility modeling of financial returns. The content of this study is classified into two parts, the first part studies the possible improvements on the modeling accuracy of an intraday volatility process while the second part investigates the potential usage of a commonly neglected intraday information source, overnight return, for modeling the market-expected volatility indicator (Chicago Board of Options Exchange volatility index).

The first proposed modification on intraday volatility modeling is about the utilization of the interaction between the intraday periodicity and the heteroskedasticity of intraday returns. To model the volatility of intraday returns, one approach is to handle the two properties, periodicity and heteroskedasticity, simultaneously and the other alternative is to capture each distinct property sequentially. Bollerslev and Ghysels take the first approach and they propose the Periodic ARCH (P-ARCH) model [21] by using a set of periodically varying autoregressive coefficients to represent the composite effect of the two properties. The model handles the interaction between the two properties implicitly to provide the best fit for the underlying series. On the other hand, Andersen and Bollerslev work on the alternative approach and they suggest a sequential method to capture the two properties individually. The periodicity is explicitly estimated at the first place and the heteroskedasticity is modeled afterwards. Their method makes use of a filtration process (deseasonalization), dividing the original series by the estimated periodicity, to separate out the periodicity and models the subsequently filtered series with appropriate ARCH models. The overall volatility of the underlying raw series is recovered as the product of the estimated periodicity times the volatility of the filtered series [2, 3].

Due to the complexity and the requirement of the computationally expensive parameter estimation of P-ARCH model, its practicality is limited in the literature. On the contrary, the use of the sequential method becomes a common approach in recent years because of its clarity and simplicity [4, 9, 17, 24, 42, 43, 53, 55, 57, 59, 69]. However, it is questionable whether the sequentially estimated ARCH parameters give the optimal performance when the objective is to model the volatility of the raw series. Under the sequential setting, the estimation of the ARCH parameters is solely based on the characteristics of the filtered series and therefore, the estimated model can only assure an optimal fit for the corresponding filtered series. There is no indication that the recovered intraday volatility fits the raw series optimally. On the contrary, it can be shown that the sequential estimation approach may not achieve the optimal result under a general situation.

Our proposed framework improves the subsequent ARCH structure in the sequential method by integrating the filtration process and the ARCH process in a united setting and optimizing the model parameters for the raw series instead of the filtered series. The presence of the estimated periodicity and the ARCH volatility simultaneously enables the consideration of their interaction while the use of the raw series assures the optimal fit to the target series. The integrated framework can be written as a modified P-ARCH structure where the periodically varying autoregressive coefficients are represented as the product of the estimated periodicity times the ARCH parameters. On the other hand, the computational requirement of our framework tends to be lesser than the P-ARCH structure as our framework only employs one additional variable, parameterized mean value of the intraday return, compared to the sequential approach does. The parameterized mean is demonstrated to give little influence empirically and may be restricted to its unconditional expected value to further reduce the computational requirement.

The proposed framework is tested with GARCH(1,1), FIGARCH(1,d,1) and HYGARCH(1,d,1) structures by using 10-minute returns of the NASDAQ index and the S&P 500 index. Moreover, both normally distributed and t-distributed innovations are considered in the investigation. The performance measures include regression R^2, mean squared error, mean absolute error and Diebold-Mariano hypothesis tests [30] on squared error and absolute error. Apart from studying the 1-step-ahead out-of-sample performance, several multiple-step-ahead forecasting results (up to one-day-ahead forecasts) are also addressed. Under the same level of model flexibility (parameterized portions), the proposed framework outperforms the sequential estimation method in both in-sample fitness and out-of-sample performance on all the forecasting horizons. The effectiveness of applying the integrated framework to different ARCH structure is also discussed in research work.

Another modification on the intraday volatility modeling is about the estimation method of the intraday periodicity. The magnitude of intraday periodicity is always restricted to follow an identically repeated U-shape structure across days [4, 8, 17, 24, 37, 53, 54, 55]. We speculate that this restriction may hinder the periodicity to represent the underlying dynamics effectively.

Andersen & Bollerslev provide a method to estimate the intraday periodicity with the allowance of day-variability [3]. The periodicity is modeled in two steps. The dynamics of an intraday return series is firstly approximated by a smoothing function (Flexible Fourier Form) and, secondly, the periodicity is recovered by a normalization procedure with the use of the approximation results. Their method is capable of defining the estimated periodicity to be either day-invariant or day-variant with proper adjustments. However, when the method is applied for day-variant situation, it can be shown that the resultant periodicity violates the implicit constraint, which is derived from the initial modeling assumption, in some situations. As a result, the correctness of the periodicity cannot be assured at all times.

Furthermore, the original normalization procedure is shown to be susceptible to heteroskedastic errors. It is demonstrated that the time series of the expected periodicity is deviated from its true value when the approximated series is contaminated with heteroskedastic errors. The conjecture of the presence of heteroskedastic errors is supported empirically. The approximation errors of the smoothing function are demonstrated to have non-constant variances. The Bartlett's Test [16] and the Brown-Forsythe's Test [23] significantly rejected the null hypothesis of constant error variance at 5% significance. As the approximated results are used as the substitute of the true underlying dynamics, it is reasonable to assume the approximated series to be deviated from its true value with heteroskedastic errors instead of homoskedastic errors.

A modified normalization procedure that ensures the fulfillment of the implicit constraint is proposed in this research work. The new procedure adjusts the magnitude of the periodicity with reference to the size of its corresponding daily variance. When the proposed procedure is applied for the day-variant situation, the results are turned out to be more robust to heteroskedastic errors under numeric simulations. For day-invariant situation, our procedure is proved to give the same performance as the original normalization does mathematically.

The modified normalization procedure is tested with GARCH(1,1) and FIGARCH(1,d,1) structure by using 10-minute returns of the NASDAQ index and the S&P 500 index. The modified procedure is demonstrated to outperform the original procedure in both in-sample fitness and out-of-sample performance on various the forecasting horizons for both normally distributed and t-distributed innovations.

The second part of this study explores the potential benefits of incorporating overnight information for modeling an financial instrument, the Chicago Board of options Exchange (CBOE) volatility indexes. The CBOE indexes indicate the market expectation of future volatilities and their values are compiled by averaging the weighted prices of put and call options [75]. As the option prices reflect the values of volatilities for both trading and non-trading (overnight) period, we speculate that the overnight information may help to improve the modeling accuracy of the CBOE indexes.

Due to the difference in sampling interval, overnight information is always limited to work as an exogenous variable in typical ARCH type models. This approach is shown to be ineffective as the involved coefficient is always insignificantly different from zero. Our work explores the possibility of incorporating the overnight variance indirectly through the use of linearly combined daily volatility estimators [41, 45, 46]. Unlike common time series data, the values involved in our work are all positive in nature. The specialized model for positive disturbances, Multiplicative Error Model (MEM) [36], is used to model the underlying conditional volatility process.

The empirical results demonstrate that the inclusion of overnight variance gives unnoticeable influence on the underlying conditional volatility process under the minimum-variance constraints. However, it can produce substantial influence when the constraints are relaxed. Besides, the influence is revealed to be not monotonic as an increase of the overnight proportion does not necessarily produce a larger influence.

Furthermore, it is demonstrated that the inclusion of overnight variance can improve the prediction accuracy of the CBOE volatility indexes (VIX and VXD). The conditional volatilities resulted from the linearly combined overnight variance and realized variance are shown to improve the adjusted R-square of the AR(1) regression models under specific weight combinations. The findings contradict the common perception that overnight return does not contain useful information for daily volatility modeling.

Overall, three new considerations for modeling the volatility of financial returns are proposed in this research work. They are 1) Utilization of the interaction effect between the intraday periodicity and the heteroskedasticity; 2) Modified normalization for the estimation of the intraday periodicity; and 3) Inclusion of overnight information for the estimation of daily volatility. Details of these propositions are documented in the following chapters.

1.2 Contributions

This research work addresses the possible improvements on volatility modeling of financial return with the use of three new considerations stated in the previous subsection. The key contributions from the corresponding research works can be summarized as:

1) Utilization of the interaction between the intraday periodicity and the heteroskedasticity (covered in Chapter 3)

- Demonstration of the mathematical limitations of Andersen & Bollerslev's sequential estimation approach for modeling intraday volatility process
- Proposition of a new framework for modeling intraday volatility
- Discussion of the effectiveness of the proposed framework on different ARCH structure theoretically
- Performance demonstration of the proposed method through simulation study and empirical analysis

2) Modified normalization for the estimation of the intraday periodicity (covered in Chapter 4)

- Demonstration of the potential contradiction in Andersen & Bollerslev's normalization method when it is applied to estimate day-varying intraday periodicity
- Empirical illustration of the heteroskedasticity of the residuals for calculating the intraday periodicity

- Simulation study of the robustness of Andersen & Bollerslev's method for heteroskedastic residuals

- Proposition of a new normalization for estimating intraday periodicity

- Mathematical proof of the equivalent modeling performance between Andersen & Bollerslev's method and the proposed method for estimating day-invariant intraday periodicity.

- Performance demonstration of the proposed method through simulation study and empirical analysis

3) Inclusion of overnight information for the estimation of daily volatility (covered in Chapter 5)

- Proposition of a linearly combined daily volatility estimator with the inclusion of overnight information

- Empirical analyses on the proposed linearly combined estimator on MEM structure and its prediction performance on CBOE volatility indexes (VIX and VXD)

1.3 Organization of the thesis

The main contents of this work is divided into four chapters. Chapter 2 gives a brief literature review on modeling the intraday volatility and daily volatility respectively. The commonly used Andersen & Bollerslev's method together with the P-ARCH structure will be discussed as they are two different fundamental concepts for intraday volatility modeling. Understanding of the two methods will make it easier to comprehend our proposed modifications on intraday volatility modeling in the subsequent chapters. On the other hand, for daily volatility modeling, literatures of some well-known daily volatility estimators, including Garman & Klass's versions and Hansen & Lunde's versions, will be addressed. Furthermore, a brief review on the Multiplicative Error Model (MEM) will be included. The MEM has a unique characteristics as it works for data with non-negative disturbances, which is different from typical time series data. The linkage between the Chicago Board of Exchange (CBOE) volatility indexes and the model-based volatility estimates will also be discussed.

Chapter 3 focus on our first proposition on intraday volatility modeling, namely integrated framework approach. The mathematical limitations of Andersen & Bollerslev's sequential approach will be addressed. It is shown that the utilization of the interaction between the intraday periodicity and the ARCH process can be a potential improvement of the model accuracy and therefore, a revised version is proposed. The new approach improves the subsequent ARCH structure in the sequential method by integrating the filtration (deseasonalization) process and the ARCH process in a united setting and optimizing the model parameters for the raw series instead of the filtered series. Our proposed method is tested through simulation study and empirical analysis. Preferences on using our approach for various forecasting horizons are strongly supported. Besides, a discussion of the effectiveness of the proposed framework on different ARCH structure is provided in this chapter.

On the other hand, a modified method for the estimation of intraday periodicity is discussed in Chapter 4. To tackle the potential contradiction in Andersen & Bollerslev's normalization method for estimating day-varying intraday periodicity, the original method is modified to fulfill the implicit constraint for the construction of daily variances from their corresponding intraday variances. For the situation that the periodicity is modeled to be day-variant, the proposed method is shown to be less susceptible to heteroskedastic errors through numerical simulations. For day-invariant periodicity, our method is proven to give the same performance as Andersen and Bollerslev's method does mathematically. Preference on using the proposed method is supported empirically.

Apart from the investigation of modeling an intraday volatility process, this research work also covers the area of daily volatility modeling. In Chapter 5, we explores the possibility of incorporating the overnight information indirectly through the use of linearly combined daily volatility estimators. Our empirical results demonstrate that the inclusion of overnight variance can produce substantial influence when the minimum-variance constraints are relaxed. Moreover, the influence is revealed to be not monotonic as an increase of the overnight proportion does not necessarily produce a larger influence. Furthermore, it is demonstrated that the new estimators can improve the prediction accuracy of the Chicago Board of options Exchange (CBOE) volatility indexes (VIX and VXD) under specific weight combinations. The concluding remarks and future directions appear in the last part, Chapter 6.

□ **End of chapter.**

Chapter 2

Literature Review

This chapter provides a brief review on modeling the intraday volatility and daily volatility respectively. Before presenting the reviews on the intraday volatility modeling methods, the key characteristics of intraday returns will be firstly addressed. After that, there will be reviews on the commonly used Andersen & Bollerslev's sequential estimation method, the P-ARCH model and the Engle's Multiplicative Component model.

For daily volatility modeling, literatures of some well-known daily volatility estimators, including Garman & Klass's versions and Hansen & Lunde's versions, will be addressed in the section 2.2. A brief review on the specialized model for non-negative disturbances, Multiplicative Error Model (MEM), will be included. Furthermore, the linkage between the Chicago Board of Exchange (CBOE) volatility indexes and the model-based volatility estimates will also be discussed.

2.1 Modeling the volatility of intraday returns

2.1.1 Characteristics of intraday returns

There is a long stem of empirical studies on intraday return. In 1985, Wood et al. [72] presented their investigation on the 1-minute returns of public stocks listed in New York Stock Exchange(NYSE). The mean of the return tended to be around zero but the mean of absolute return tended to change in a noticeable pattern. The mean of absolute return was large at the market start and dropped to a small value in the middle of the day, then it rose to a larger value when the market was near to close. This distinct pattern is called as a 'U-shape' pattern.

Some empirical works were done for different sampling frequencies in recent years, including Admati & Pfleiderer [1] worked on 1-hour stock returns; Beltratti & Claudio [17] studied 30-minute exchange rate returns; Giot [43] used 15-minute and 30-minute stock returns and Andersen & Bollerslev [3] investigate the return of 5-minute exchange rate and stock respectively. Overall, it is commonly assumed that the mean of intraday return is nearly zero and tends to be a true zero when the sampling frequency increases. The mean of absolute returns and the square of the returns gives a U-shape pattern when they are plotted against time for all sampling frequencies. And, the kurtosis increases as the sampling interval becomes smaller. There is not any clear relationship between skewness and sampling frequency.

Apart from distributional characteristics, one of the key ingredients for time series modeling is the properties of autocorrelation. Studies worked on either the returns of exchange rate or stock [72] [1] [3] [17] [43] [67] in different sampling frequency gave similar results. The sample autocorrelations of intraday returns oscillate around zero and are insignificantly different from zero. However, the autocorrelations of absolute returns give a remarkably regular pattern within a day shift and across days. The autocorrelation is high at the beginning, declines in the middle and rises at the end when it is plotted against time within a day shift. The pattern appears within a day shift is replicated with a little decay in the magnitude when the autocorrelations are plotted across a longer period (e.g. a few days shift). The remarkably repeated pattern in the autocorrelations sheds the possibility of the existence of regularity of the variance of intraday returns.

The correlograms (autocorrelation plot) of raw returns and absolute returns for 10-minute NASDAQ index data are plotted in Figure 2.1 and Figure 2.2 respectively.

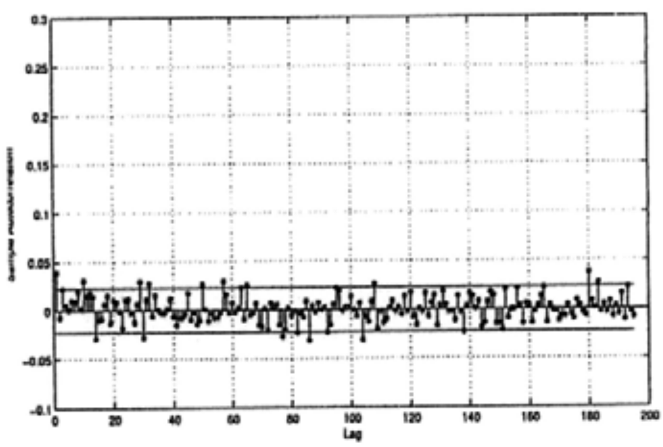


Figure 2.1: Correlogram of 10-minute NASDAQ returns

The existence of the periodic U-shape patterns in Figure 2.2 reveals that typical ARCH type models (e.g. ARCH [31], GARCH [19], EGARCH [61], FIGARCH [13], FIEGARCH [22], HYGARCH [29]) cannot be applied to

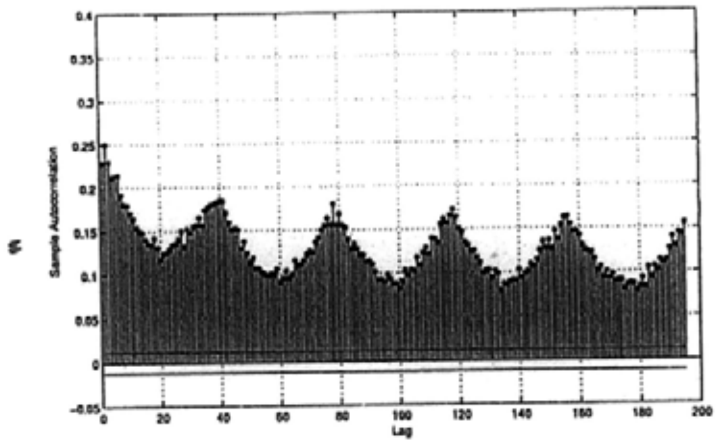


Figure 2.2: Correlogram of 10-minute NASDAQ absolute returns

model the volatility of the series directly. It is because the autocorrelations of the absolute series of typical ARCH models are specified to follow a strictly decaying pattern under their theoretical assumptions.

On the other hand, there is a specialized model for handling the above periodic U-shape patterns. Bollerslev and Ghysels proposed a model called P-GARCH [21] to capture the repetitive U-shape variations with periodically changing coefficients in the conditional variance equation. Different from typical ARCH type models where the conditional variances are formulated by a ARMA structure, the P-GARCH model expresses the conditional variances as a periodic ARMA structure instead. The use of a periodic ARMA structure enables the capability to capture the repetitive periodic patterns theoretically. However, the applications of P-GARCH on high-frequency financial data are limited in the literature due to its complexity and computationally expensive parameter estimation requirement.

2.1.2 Models for intraday volatility modeling

The summary of three different approaches for modeling an intraday volatility process is presented in this section. The elaborations on Andersen & Bollerslev's sequential estimation model and the P-GARCH model may help the reader to better understand on the content in Chapter 3. The subsections, Engle's Multiplicative Component model, can be used as the background information for Chapter 4.

2.1.2.1 Andersen & Bollerslev's sequential estimation model

The seminal work by Andersen & Bollerslev establishes a commonly used method for modeling the intraday volatility in a sequential way [3]. They notice a prominent U-shape periodic pattern in an absolute intraday return series and discuss the inappropriateness to employ a ARCH type model on an intraday return series directly. To solve the problem, they propose the periodicity should be filtered out before the ARCH model is applied. They propose to decompose the intraday volatility into two components, representing the periodicity and heteroskedasticity respectively. The periodicity is estimated at the first place and the heteroskedasticity is modeled afterwards. Their method makes use of a filtration process, dividing the series by the estimated periodicity, to separate out the periodicity and models the subsequently filtered series with appropriate ARCH models. The overall volatility of the underlying series is recovered as the product of the estimated periodicity times the volatility of the filtered series.

Estimation of the intraday periodicity

The estimation of periodicity is carried out in two steps. The dynamics of an intraday return series is firstly approximated by a smoothing function (Flexible Fourier Form) and, secondly, the periodicity is recovered by a normalization procedure with the use of the smoothed results.

The estimation of the periodicity is based on an assumption on an intraday return process:

$$\begin{aligned} r_{t,n} &= E(r_{t,n}) + s_{t,n} \cdot \frac{\sigma_t}{\sqrt{N}} \cdot z_{t,n} \\ z_{t,n} &\sim N(0,1) \end{aligned} \tag{2.1}$$

where $r_{t,n}$ is an intraday return of day t in n-th intraday interval and $E(r_{t,n})$ is the unconditional expectation of $r_{t,n}$. The $r_{t,n}$ is calculated as $log(P_{t,n}/P_{t,n-1})$. $P_{t,n}$ is the index value and $P_{t,0}$ is the value at the market open. The intraday periodicity is denoted as $s_{t,n}$ and it reflects the U-shape regularity across days. σ_t is a daily volatility (standard deviation) of the return in day t and N is the number of intraday interval per day. $z_{t,n}$ is a i.i.d. standard normal random term. The above formula provides a simplified expression of an intraday return process and makes the estimation of the intraday periodicity feasible.

There are a number of estimators available to evaluate the value of σ_t, including ARCH type conditional volatility [33], realized variance [6] and range-based estimators [41, 64, 66, 68, 73]. The GARCH(1,1) conditional volatility is used as the substitute of σ_t throughout this research work.

By squaring and taking logs on both sides on the equation 2.1, we get:

$$\begin{aligned} s_{t,n}^2 * z_{t,n}^2 * \sigma_t^2 &= N * (r_{t,n} - E(r_t))^2 \\ log(s_{t,n}^2) + log(z_{t,n}^2) + log(\sigma_t^2) &= log(N) + 2log(|r_{t,n} - E(r_t)|) \\ log(s_{t,n}^2) + log(z_{t,n}^2) &= 2log(|r_{t,n} - E(r_t)|) + log(N) \\ &\quad -log(\sigma_t^2) \end{aligned} \tag{2.2}$$

Therefore, the $s_{t,n}$ can be expressed in terms of an intermediate proxy variable for the intraday dynamics $x_{t,n}$ under the following settings:

$$\begin{aligned} x_{t,n} &\equiv 2log[|r_{t,n} - E(r_{t,n})|] - log(\sigma_t^2) + log(N) \\ x_{t,n} &= f(\Theta|\sigma_t, n) + u_{t,n} \\ \hat{x}_{t,n} &\equiv f(\hat{\Theta}|\hat{\sigma}_t, n) \end{aligned} \tag{2.3}$$

The function $f(\Theta|\sigma_t, n)$ is a flexible Fourier form[39] which is defined as below[1]:

$$f(\Theta|\sigma_t, n) = \sum_{j=0}^{J} \sigma_t^j [\sum_{k=0}^{K} \nu_{kj} \cdot n^k + \sum_{p=1}^{P} (\gamma_{pj} cos \frac{pn2\pi}{N} + \delta_{pj} sin \frac{pn2\pi}{N})] \tag{2.4}$$

where Θ represents the parameter set $\{\nu_{kj}, \gamma_{pj}, \delta_{pj}\}$. Variables J, K and P control the order of expansion. The parameter set Θ is estimated by setting $f(\Theta|\sigma_t, n)$ as a regressor (independent variable) for a dummy variable $x_{t,n}$ of a linear regression with $u_{t,n}$ as a zero mean i.i.d. error term. $\hat{x}_{t,n}$ is an approximation of the underlying periodic dynamic. The $\hat{x}_{t,n}$ is allowed to

[1] Additional elaboration on the flexible Fourier form is presented in the Supplementary note 2.3.1.

vary across days by setting $J > 0$. Reader can refer to page 152-153 in the paper [3] for detailed explanation on the validity of this method.

After obtaining the approximated value of the intraday dynamics $\hat{x}_{t,n}$, a normalization step is used to retrieve the intraday periodicity $s_{t,n}$ accordingly.

$$s_{t,n} = \frac{S \cdot exp(\frac{\hat{x}_{t,n}}{2})}{\sum_{i=1}^{S/N} \sum_{j=1}^{N} exp(\frac{\hat{x}_{i,j}}{2})} \tag{2.5}$$

where $s_{t,n}$ denotes the intraday periodicity and S is the sample size of the intraday return series.

On the other hand, Andersen & Bollerslev also suggest a heuristic method to estimate the intraday periodicity without imposing any model assumption. They suggest that the periodicity can be estimated as the mean absolute value for the corresponding interval [2].

$$s_{t,n} = \frac{\sum_{i=T_1}^{T_2} |r_{i,n}|}{T} \tag{2.6}$$

where $r_{i,n}$ is the intraday return of the n-th intraday interval on day i. T is the total number of days in the sample, T_1 and T_2 are the indexes of the start day and the end day respectively.

Filtration process (Deseasonalization)

The filtration (deseasonalization) process is defined as:

$$\tilde{r}_{t,n} = \frac{r_{t,n} - E(r_{t,n})}{s_{t,n}} \tag{2.7}$$

The $\tilde{r}_{t,n}$ denotes a filtered series which is free from the problematic (periodic) pattern in the autocorrelations of the absolute intraday return series.

$s_{t,n}$ represents the intraday periodicity which is required to be estimated based on the observed intraday returns.

Composition of an intraday volatility

After the filtration process, the volatility of the filtered series can be modeled by any ARCH type models. Let the conditional volatility of the filtered series be $\tilde{\sigma}_{t,n}$, the intraday volatility $\sigma_{t,n}$ for the underlying observed intraday return series will be calculated as:

$$\sigma_{t,n} = s_{t,n} \cdot \tilde{\sigma}_{t,n} \tag{2.8}$$

Overall, it is shown that the intraday volatility can be expressed as the result of the volatility of the filtered series times the intraday periodicity.

2.1.2.2 Periodic Autoregressive Conditional Heteroskedasticity (P-GARCH) model

To model the volatility of intraday returns, another approach is to use a P-GARCH model [21]. It is designed to capture the repetitive periodic time variation in the second-order moment (variance) of a series. The model makes use of periodically varying autoregressive coefficients to represent 'the periodicity in conditional heteroskedasticity' (Quoted from the paper [21]). For example, a P-GARCH(1,1) structure can be applied on ϵ_t directly to model its underlying volatility.

P-GARCH(1,1)

$$\begin{aligned} \sigma_t^2 &= \omega_{s(t)} + \alpha_{s(t)}\epsilon_{t-1}^2 + \beta_{s(t)}\sigma_{t-1}^2 \\ \epsilon_t &\sim D(0, \sigma_t^2) \end{aligned} \tag{2.9}$$

where σ_t is the volatility of the intraday innovation ϵ_t. The innovations are assumed to follow a zero-mean i.i.d. distribution with varying variances σ_t^2. The subscripts of the variables denote the corresponding time index. $\omega_{s(t)}$, $\alpha_{s(t)}$ and $\beta_{s(t)}$ are periodically varying coefficients with particular value for each period $s(t)$. For the situation where there are L stages for a single periodic cycle, the number of parameters for the above P-GARCH(1,1) structure will be $3L$.

Due to the complexity of the model and the computationally expensive requirement for estimating the parameters, the P-GARCH structure is not a popular approach for intraday volatility modeling.

2.1.2.3 Engle's Multiplicative Component model

Apart from decomposing the intraday volatility into two distinct components (eqt. 2.8), Engle et al [37] suggest a refined version to further decompose the volatility into three multiplicative components. The model is called as the Multiplicative Component model.

$$\sigma_{t,n} = s_{t,n} \cdot \sigma_t \cdot \sqrt{q}_{t,n} \tag{2.10}$$

where $\sigma_{t,n}$ is an intraday volatility, $s_{t,n}$ and σ_t are the intraday periodicity and the daily conditional volatility respectively. The $q_{t,n}$ is a ARCH conditional volatility.

For instance, the overall framework for modeling the intraday volatility with GARCH(1,1) [19] structures with normally distributed innovation can be described as:

$$\begin{aligned} \xi_{t,n} &= \frac{r_{t,n} - E(r_{t,n})}{s_{t,n} \cdot \sigma_t} - \kappa \\ q_{t,n} &= \omega + \alpha \cdot \xi^2_{t,n-1} + \beta \cdot q_{t,n-1} \\ r_{t,n} &\sim N(0, [s_{t,n}\sigma_t\sqrt{q}_{t,n}]^2) \end{aligned} \tag{2.11}$$

where $\xi_{t,n}$ denotes a series with heteroskedasticity and κ is its conditional mean. ω, α and β are the GARCH(1,1) parameters. The parameters κ, ω, α and β are estimated with the maximum likelihood method.

2.2 Models for measuring daily volatility

There are different methods to measure the daily volatility of a financial instrument. The following paragraphs give a brief overview on the possible ways used in this research, namely the point estimator, model-based (ARCH models) conditional volatility and market-based volatility indexes.

2.2.1 Daily volatility estimators

Estimating volatility has long been a major issue in the financial literature. The classical estimation method assumes the underlying stock price follows simple Brownian motion and takes the squared log daily return as a unbiased variance estimator. However, this simple estimator cannot satisfactorily capture the underlying dynamics and was shown to be not efficient [41]. In 1980, Parkinson formulated a way to use high and low price to better capture the underlying dynamics [64]. In the same year, Garman and Klass proposed a minimum-variance unbiased variance estimator for simple Brownian motion [41]. Ten years later, an estimator for drifted Brownian motion was proposed by Rogers and Satchell. It was a drift-independent variance estimator that was proven to be unbiased for Brownian motion with drift [66]. Besides, Yang and Zhang formulated a multiple-day-averaged drift independent minimum-variance unbiased variance estimator in 2000 [73]. These volatility estimators are called range-based estimators as they are calculated by using the open, high, low and close information.

The linearly combined estimator used in this research (third modification) is based on the seminal work by Garman & Klass.

2.2.1.1 Garman & Klass's daily volatility estimators

Garman and Klass proposed a number of estimators in their seminal paper in 1980 [41]. Among the few estimators, there is one for estimating the volatility in the active trading period and one for whole day period. The estimator for active trading period is formulated as follows.

$$\hat{\sigma}_4^2 = 0.511(u-d)^2 - 0.019[c(u+d) - 2ud] - 0.383c^2 \tag{2.12}$$

where C_t and O_t are the log values of the closing price and the opening price, H_t and L_t are the log values of the highest price and the lowest price for day t respectively. The normalized high, low and close are expressed as u, d and c individually, where $u = H_t - O_t$, $d = L_t - O_t$ and $c = C_t - O_t$.

<u>Garman & Klass's minimum-variance estimator</u>

The whole day version is formulated as :

$$\hat{\sigma}_6^2 \;=\; \frac{a}{f}(O_t - C_{t-1})^2 + \frac{1-a}{1-f}\hat{\sigma}_4^2 \tag{2.13}$$

where f is the fraction of that day that trading is closed and it is set to 1050/1440 in our study. a is a weight parameter and it is set to 0.12 to achieve minimum-variance property regardless of the value of f.

2.2.1.2 Hansen & Lunde's volatility estimators

Another stream of volatility estimators are called high-frequency estimators as they are based on the high-frequency intraday return information. Realized variance is the most well-known high frequency estimator to measure the volatility in active trading period. It is formulated as the sum of squared high-frequency intraday returns. Its formulation for day t can be expressed as:

$$\hat{\sigma}^2_{RV} = \sum_{i=1}^{m}\{(p(x_i) - p(x_i - \Delta)\}^2 \tag{2.14}$$

where x_i denotes the time, $p(x_i)$ is the price at time x_i, Δ is the sampling interval, m is the total number of intraday price recorded (excluding the opening price) in a day and $p(x_0)$ is the opening price of day t.

An important concern about the estimation of realized variance is whether the return series is autocorrelated or not [63]. The standard realized variance becomes biased when the returns are autocorrelated and the result should then be adjusted accordingly. The autocorrelation phenomenon always happen in high frequency data and there are various ways to offset the bias. Hansen & Lunde suggested an estimation method to handle the bias and regarded it as Newey-West modified realized variance [45]. This method is based on Bartlett kernel and is guaranteed to be nonnegative. The Newey-West modified variance for day t is defined as:

$$\begin{aligned} \hat{\sigma}^2_{NW} &= \sum_{i=1}^{m} y_i^2 + 2\sum_{h=1}^{q}(1 - \frac{h}{q+1})\sum_{i=1}^{m-h} y_i y_{i+h} \\ y_i &= p(x_i) - p(x_i - \Delta) \end{aligned} \tag{2.15}$$

the variable q represents the lag-length and it is set to $q = ceil(\frac{mw}{d})$ where w is the desired length of lag window and d is the total length of sampling period(trading period) in minutes[45]

Hansen and Lunde's minimum-variance estimator

For the whole day period, Hansen and Lunde defined an optimally combined whole day variance estimator in 2005 [46]. It is a minimum-variance estimate constructed by weighted squared overnight return and Newey-West modified realized variance. The following equations show the settings for Hansen and Lunde's minimum-variance whole-day-based estimator ($\hat{\sigma}^2_{wholeRV}$).

$$\begin{aligned}
\hat{\sigma}^2_{wholeRV} &= \omega_1^* r_{1,t}^2 + \omega_2^* \hat{\sigma}^2_{NW,t} \\
\omega_1^* &= (1-\varphi)\frac{\mu_0}{\mu_1} \\
\omega_2^* &= \varphi\frac{\mu_0}{\mu_2} \\
\varphi &= \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}}
\end{aligned} \tag{2.16}$$

where ω_1^* and ω_2^* are the optimal weight for the min-variance estimator. $r_{1,t}^2$ is the squared overnight return. μ_0, μ_1 and μ_2 are the expected value of integrated variance, overnight variance and Newey-West realized variance respectively. η_1, η_2 and η_{12} are the variance of overnight variance and Newey-West realized variance and their covariance respectively.

2.2.2 Engle's Multiplicative Error Model (MEM)

Most of the popular volatility prediction models such as variants of ARCH models [19, 36, 35, 47] and RiskMetrics [60] are not suitable to handle non-negative time series as the information source. To properly handle the non-negativity, Engle proposed to model the error in the series as the multiple of the conditional mean estimates, adopting it as the mean equation in the

GARCH framework [36]. This model is known as Multiplicative Error Model (MEM) and it can provide consistent results for various distributions of error terms under its quasi maximum likelihood estimation method, making it robust to ambiguous error assumptions [52, 38].

The way to employ volatility estimators as input series for GARCH type models is different from those for treating returns as input series. Due to the non-negative nature of the volatility estimators, it is difficult to use traditional GARCH type models, which are based on linear formulations on the return process, to estimate the model parameters as the variance and higher moments of error distribution are unlikely to be constant [36]. Engle proposed an efficient way to model non-negative series in a GARCH framework by treating the series as a composition of its conditional mean multiplied by a unit-mean error term. This multiplicative error structure is able to provide consistent results for error terms belong to a family of gamma distribution as the corresponding first order optimality conditions on the log-likelihood functions is the same.

The MEM (1,1) model is defined by the following two equations.

$$x_t = \mu_t \epsilon_t \qquad \text{Mean eqt.} \tag{2.17}$$

$$\mu_t = \omega + \alpha x_{t-1} + \beta \mu_{t-1} + c' z_{t-1} \qquad \text{Variance eqt.} \tag{2.18}$$

In the mean equation, x_t is the non-negative time series, μ_t is the conditional mean estimates and ϵ_t represents a unit-mean gamma-distributed i.i.d. error process. The variance equation is similar as that in the GARCH framework by replacing the error squared term with x_t in the ARCH term. Furthermore, exogenous variables are treated by including z_t in the variance equation.

With the restriction of unit-mean on the distribution of ϵ_t, the corresponding log-likelihood function for the model is defined as $L(\theta)$.

$$L(\theta) = constant - a\sum_{t=1}^{T}[-log(\mu_t(\theta)) - \frac{x_t}{\mu_t(\theta)}] \tag{2.19}$$

where a controls the shape of the gamma distribution, θ is the parameter set $\{\alpha$, β and $c'\}$ to be estimated and T is the size of training sample. The first order optimality condition for maximum likelihood estimation is:

$$\sum_{t=1}^{T}[-log(\mu_t(\theta)) - \frac{x_t}{\mu_t(\theta)}] \tag{2.20}$$

Noticing that the shape variable a does not affect the first order condition and its value is irrelevant and does not have any influence on θ and their standard errors [36, 38].

Without the justification of the underlying error distribution, it is suggested to derive the parameters in equation 2.18 by traditional GARCH framework as the maximizer of equation 2.19, which is a quasi maximum likelihood estimator. Lee & Hansen demonstrated the Gaussian likelihood estimation method for GARCH(1,1) can provide consistently estimated parameters for input series which is neither Gaussian nor independent [52]. Referring to Engle's procedure, the parameters can be obtained by taking the positive square root of the non-negative variable of interest $\sqrt{x_t}$ as the dependent variable and setting the mean value to zero with the assumption of normally distributed errors [36, 38]. This framework has been successfully modeled the dynamics of non-negative volatility series, including range-based and high-frequency estimates, in some applications [38, 51, 50] and has been extended to multivariate cases recently [25].

2.2.3 Chicago Board of Exchange (CBOE) Volatility index

Besides using econometric models to measure future volatility, there is another way to indicate the level of fluctuation in the future. The Chicago Board of Options Exchange (CBOE), the world's largest options exchange, has compiled volatility indexes (e.g. VIX and VXD. VIX aims to capture the volatility of S&P 500 index while VXD aims at Dow Jones Industrial Average (DJIA)) by averaging the weighted prices of the corresponding put and call options to measure the market expectation on future volatility[75]. The indexes are used to represent the market's expectation of future fluctuation on the next 30 calendar days (22 trading days) and hence they are called market-based volatility [65].

The CBOE market-based volatility index has been related to the model based conditional volatility in recent literatures. Blair et al. investigated the information content of the VIX for the prediction of GARCH volatility and found out the current VIX value contains the richest information content for 1-step-ahead predicted realized volatility [18]. Besides, Engle & Gallo studied the possibility of using MEM volatilities to improve the prediction of VIX in 2006. They demonstrated multi-step average volatilities can be incorporated as statistically significant regressors in the auto-regression of the VIX [38].

2.3 Supplementary note

2.3.1 Flexible Fourier functional form for the approximation of periodic dynamics

The Fourier series is always considered as a nature choice to model a periodic series. Any periodic series can be decomposed and expressed as a sum of simple trigonometric functions, sines and cosines, and the approximation accuracy is controlled by the total number of terms employed accordingly. A closed-form periodic function $p(x)$ of period 2π, it can be represented as:

$$\begin{aligned} p(x) &= a_0 + \sum_{n=1}^{\infty}(a_n cos(nx) + b_n sin(nx)) \quad (2.21)\\ a_0 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} p(x)dx \\ a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} p(x)cos(nx)dx \qquad n = 1, 2, \cdots \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} p(x)sin(nx)dx \qquad n = 1, 2, \cdots \end{aligned}$$

The coefficients a_0, a_n and b_n are evaluated based on Euler formulas and the R.H.S. of equation 2.21 converges to the L.H.S. as n goes to infinity.

However, the above closed-form solution cannot be applied to many practical situations as the periodic function $p(x)$ is either unknown or cannot be expressed as a closed-form function. Nevertheless, the idea to decompose a series into a sum of trigonometric functions turned to be very powerful and became one of the popular methods to handle periodic series in both non-financial and financial applications [70]. For financial applications, Gallant [39] and Andersen & Bollerslev [3] modified the original Fourier series to incorporate the possible existence of a linear trend, quadratic trend and heteroskedasticity in the stock market. The revised form is called the modified flexible Fourier form (FFF) $f(\theta;\sigma_t,n)$ and it is defined as follows:

$$
\begin{aligned}
f(\theta;\sigma_t,n) &= \sum_{j=0}^{J}\sigma_t^j[\mu_{0j}+\mu_{1j}\frac{n}{N_1}+\mu_{2j}\frac{n^2}{N_2}+\sum_{i=1}^{D}\lambda_{ij}I_{n=d_i} \\
&\quad +\sum_{i=1}^{P}(\gamma_{ij}cos\frac{pn2\pi}{N}+\delta_{ij}sin\frac{pn2\pi}{N})] \qquad (2.22)
\end{aligned}
$$

where N is the total number of subintervals for a single period. For our study, N is the number of intraday intervals per day. $N_1=\frac{N+1}{2}$ and $N_2=\frac{(N+1)(N+2)}{6}$. θ represents the parameter set $\{\mu_{0j},\mu_{1j},\mu_{2j},\lambda_{ij},\gamma_{ij},\delta_{ij}\}$; σ_t is the daily volatility; n is the index for specific intraday interval; $I_{n=d_i}$ is a indicator variable which equals to 1 for specific interval d_i and 0 otherwise. Variables J and P controls the order of expansion.

For the cases that a time series is not directly observable, the modified flexible Fourier form (FFF) can be employed to model the underlying dynamics with very few restrictions on the possible shape of the series. However, the selection of the order of expansion is still an open issue and there is not any clear consensus on the way to balance in-sample fitness and out-of-sample prediction accuracy. [70] [39] [3]

Other flexible functional forms can be used for modeling periodic patterns but the discussion of their suitabilities are out of the scope of this study. Readers may refer to the paper by Thompson [70] for additional information of other functional forms.

□ **End of chapter.**

Chapter 3

Integrated framework approach for intraday volatility modeling

Summary

This chapter discusses the limitations of Andersen & Bollerslev's sequential estimation approach for modeling an intraday volatility process. Under the sequential setting, the estimation of the ARCH parameters is solely based on the characteristics of the filtered series and therefore, the estimated model can only assure an optimal fit for the corresponding filtered series. There is no indication that the recovered intraday volatility fits the raw series optimally. On the contrary, it can be shown that the sequential estimation approach may not achieve the optimal result under a general situation.

A new approach that utilizes the interaction effect between the periodicity and the heteroskedasticity is proposed. Our method improves the subsequent ARCH structure in the sequential method by integrating the filtration (deseasonalization) process and the ARCH process in a united setting and optimizing the model parameters for the raw series instead of the filtered series.

The proposed approach is tested with GARCH(1,1), FI-GARCH(1,d,1) and HYGARCH(1,d,1) structures by using 10-minute returns of the NASDAQ and S&P 500 indexes. Preferences on using our approach for various forecasting horizons are supported empirically.

3.1 Introduction and the research objective

The existence of periodicity and heteroskedasticity in intraday returns complicates the way to model its underlying volatility process. One attempt is to handle the two properties simultaneously while the other alternative is to capture each distinct property sequentially. Bollerslev and Ghysels propose to employ a set of periodically varying autoregressive coefficients to capture the composite effect of the two properties simultaneously through the Periodic ARCH (P-ARCH) model [21]. The model handles the interaction between the two properties implicitly to provide the best fit for the underlying series. Alternatively, Andersen and Bollerslev suggest a sequential method to capture the two properties individually. The periodicity is explicitly estimated in the first place and the heteroskedasticity is modeled afterwards. Their method makes use of a filtration process, dividing the series by the estimated periodicity, to separate out the periodicity and models the subsequently filtered series with appropriate ARCH models. The overall volatility of the underlying raw series is recovered as the product of the estimated periodicity times the volatility of the filtered series [2, 3].

Due to the complexity and the computationally expensive parameter estimation requirement of P-ARCH model, its practicality is limited in the literature for modeling high frequency intraday returns. On the other hand, the use of the sequential method becomes a common approach in recent years because of its clarity and simplicity [4, 8, 9, 17, 24, 42, 43, 53, 54, 55, 56, 57, 59, 69]. However, it is questionable whether the sequentially estimated ARCH parameters give the optimal performance when the objective is to model the volatility of the raw series. Under the sequential setting, the estimation of the ARCH parameters is solely based on the characteristics of the filtered series and therefore, the estimated model can only assure an optimal fit for the corresponding filtered series. There is no indication that the recovered intraday volatility fits the raw series optimally. On the contrary, it can be shown that the sequential estimation approach may not achieve the optimal result under a general situation.

An integrated framework approach that utilizes the interaction effect between the periodicity and the heteroskedasticity is proposed. The proposed framework improves the subsequent ARCH structure in the sequential method by integrating the filtration process and the ARCH process in a united setting and optimizing the model parameters for the raw series instead of the filtered series. The simultaneous presence of the estimated periodicity and the ARCH volatility enables the consideration of their interaction while the use of the raw series assures the optimal fit to the target series. The integrated framework can be written as a modified P-ARCH structure where the periodically varying autoregressive coefficients are represented as the product of the estimated periodicity times the ARCH parameters. Under the same level of model flexibility (parameterized portions), the optimal model parameters of the integrated framework is demonstrated to achieve a better fit for the underlying series empirically.

The remainder of this chapter is organized as follows: Section 3.2 addresses the limitations of Andersen & Bollerslev's sequential estimation method. The formulation of our proposed framework together with the rationales behind our modifications are stated in Section 3.3. The effectiveness of applying the proposed framework to different ARCH structures is also discussed in this section. Section 3.4 states the model specifications and the evaluation criteria. Apart from the consideration of the two filtration approaches proposed by Andersen & Bollerslev, our framework is tested with three ARCH structures including GARCH(1,1), FIGARCH(1,d,1) and HYGARCH(1,d,1) under the assumption of either normally distributed or t-distributed innovations. Furthermore, beside the investigation of one-step-ahead out-of-sample forecasts, the performances of different forecasting horizons are also studied to illustrate the capability of the proposed framework over multiple-step-ahead periods. Section 3.6 states the details of our empirical analyses. The results are based on the use of 10-minute intraday returns of the NASDAQ index (3-year period) and the S&P 500 index (2.5-year period). Section 3.7 contains the concluding remarks.

3.2 Andersen & Bollerslev's sequential intraday volatility model and its limitation

The most commonly used method to model an intraday volatility process is Andersen & Bollerslev's sequential estimation approach [2, 3]. They propose to decompose the intraday volatility into two components, representing the periodicity and heteroskedasticity respectively. The periodicity is estimated at the first place and the heteroskedasticity is modeled afterwards. Their method makes use of a filtration process, dividing the series by the estimated periodicity, to separate out the periodicity and models the subsequently filtered series with appropriate ARCH models. The overall volatility of the underlying series is recovered as the product of the estimated periodicity times the volatility of the filtered series.

For the situation that a time series of i.i.d. zero-mean intraday innovations is given, the filtration process will then be:

$$\tilde{r}_{t,n} = \frac{\epsilon_{t,n}}{s_{t,n}} \tag{3.1}$$

where $\epsilon_{t,n}$ denotes the intraday innovation, $\tilde{r}_{t,n}$ is the filtered series and $s_{t,n}$ is the pre-estimated intraday periodicity. The subscripts denote the time index for the variable, where t, n means the n-th intraday interval on day t.

Afterwards, the volatility of the filtered series is modeled by a ARCH model accordingly. For the case that GARCH(1,1) is employed to model the filtered series, the corresponding formulation will be:

$$\begin{aligned} \tilde{\epsilon}_{t,n} &= \tilde{r}_{t,n} - \kappa \\ \tilde{\sigma}^2_{t,n} &= \omega + \alpha\tilde{\epsilon}^2_{t,n-1} + \beta\tilde{\sigma}^2_{t,n-1} \\ \tilde{\epsilon}_{t,n} &\sim D(0, \tilde{\sigma}^2_{t,n}) \end{aligned} \tag{3.2}$$

where $\tilde{\epsilon}_{t,n}$ is the innovation of the filtered series and it follows a i.i.d. zero-mean distribution with variance equals to $\tilde{\sigma}^2_{t,n}$. κ denotes the mean value of the filtered series; ω, α and β are the parameters for a GARCH(1,1) structure. The volatility of the intraday innovation is recovered as the product of the pre-estimated periodicity times the volatility of the filtered series (i.e. $\sqrt{Var(\epsilon_{t,n})} = s_{t,n}\tilde{\epsilon}_{t,n}$).

From another perspective, the distributional property of the intraday innovation ϵ_t can be inferred as:

$$\begin{aligned} \tilde{\epsilon}_{t,n} &\sim D(0, \tilde{\sigma}^2_{t,n}) \\ \Rightarrow\ s_{t,n}(\tilde{\epsilon}_{t,n} + \kappa) &\sim D(s_{t,n}\kappa, s^2_{t,n}\tilde{\sigma}^2_{t,n}) \\ \Rightarrow\ \epsilon_{t,n} &\sim D(s_{t,n}\kappa, s^2_{t,n}\tilde{\sigma}^2_{t,n}) \end{aligned} \tag{3.3}$$

It is argued that the recovered intraday volatility $s_{t,n}\tilde{\sigma}_{t,n}$ from the sequential estimation approach may not achieve the best result for modeling the volatility of $\epsilon_{t,n}$ because:

1. The distributional property implied from the sequential approach may be different from the initial assumption, which formulates the intraday innovation series $\epsilon_{t,n}$ as zero-mean.

The result from the sequential approach implies $\epsilon_{t,n} \sim D(s_{t,n}\kappa, s_{t,n}^2\bar{\sigma}_{t,n}^2)$ and therefore the mean of $\epsilon_{t,n}$ is modeled as a time varying value $s_{t,n}\kappa$. Although $\epsilon_{t,n}$ is zero-mean, there is no guarantee that the conditional mean of the filtered series κ is zero as the filtration is a nonlinear transformation process. For the case that κ is non-zero, the mean of $\epsilon_{t,n}$ implied from the sequential approach will be a time varying non-zero value.

2. The subsequently estimated ARCH parameters may not work with the estimated periodicity $s_{t,n}$ in the best possible way to represent the volatility of $\epsilon_{t,n}$.

   As the estimation of periodicity $s_{t,n}$ cannot be error-free in practice, the filtered series will always be contaminated. The contamination of the filtered series will in turn affect the correctness of the subsequently estimated ARCH parameters. As the ARCH parameters are optimized solely for the erroneous filtered series, there is no guarantee that the composite term $s_{t,n}\bar{\epsilon}_{t,n}$ provides the best fit for $\epsilon_{t,n}$. It is possible that the estimated $s_{t,n}$ can work with another parameter set to give a better representation for the raw series under the same specification of ARCH structure.

To search for an optimal fit for the intraday innovation series, an integrated framework that directly operates on $\epsilon_{t,n}$ with the consideration of the interaction effect between the periodicity and heteroskedasticity is proposed.

3.3 Proposed Integrated framework approach

Our proposed framework improves the subsequent ARCH structure in the sequential method by integrating the filtration process and the ARCH process in a united setting while optimizing the model parameters for the raw series (target) instead of the filtered series. The filtration process stated in the equation 3.1 is embedded into the ARCH framework specified by the equation set 3.2 with the change of the distributional specification on the innovation series. The simultaneous presence of the estimated periodicity and the ARCH volatility enables the consideration of their interaction while the use of the raw innovations assure the optimal fit of the target series. With a given series of pre-estimated periodicity, the volatility of the raw series is formulated as:

$$\begin{aligned}
\tilde{\epsilon}_{t,n} &= \frac{\epsilon_{t,n}}{s_{t,n}} - \kappa \\
\tilde{\sigma}^2_{t,n} &= \omega + \alpha\tilde{\epsilon}^2_{t,n-1} + \beta\tilde{\sigma}^2_{t,n-1} \\
\epsilon_{t,n} &\sim D(0, s^2_{t,n}\tilde{\sigma}^2_{t,n})
\end{aligned} \tag{3.4}$$

The key advantage of the integrated framework relies on its ability to adjust the model parameters appropriately. The simultaneous presence of the intraday periodicity $s_{t,n}$ and the ARCH conditional volatility $\tilde{\sigma}_{t,n}$ explicitly reveals the magnitude of the composite term $s_{t,n}\tilde{\sigma}_{t,n}$, which represents the volatility of the target series. As a result, the model parameters can be adjusted in a way such that $\tilde{\sigma}_{t,n}$ works complementarily with $s_{t,n}$ to provide the best fit of $\epsilon_{t,n}$. In addition, the direct specification of $\epsilon_{t,n}$ helps to assure its zero-mean property.

On the other hand, the above integrated framework can be written as a variant of Periodic ARCH type (P-ARCH) structure [21] where the periodically varying autoregressive coefficients are represented as the product of the estimated periodicity times the ARCH parameters.

Consider the multiplication of both sides of the variance equation in the equation set 3.4 with $s_{t,n}^2$:

$$\begin{aligned} s_{t,n}^2\bar{\sigma}_{t,n}^2 &= s_{t,n}^2\omega + s_{t,n}^2\alpha\bar{\epsilon}_{t,n-1}^2 + s_{t,n}^2\beta\bar{\sigma}_{t,n-1}^2 \\ &= s_{t,n}^2\omega + \frac{\alpha s_{t,n}^2}{s_{t,n-1}^2}s_{t,n-1}^2\bar{\epsilon}_{t,n-1}^2 + \frac{\beta s_{t,n}^2}{s_{t,n-1}^2}s_{t,n-1}^2\bar{\sigma}_{t,n-1}^2 \end{aligned} \tag{3.5}$$

Since $s_{t,n}\bar{\epsilon}_{t,n} = \epsilon_t - s_{t,n}\kappa$ and $Var(\epsilon_{t,n}) = s_{t,n}^2\bar{\sigma}_{t,n}^2$, the above equation can be re-written as below by letting $\sigma_{t,n}^2 \equiv Var(\epsilon_{t,n})$:

$$\begin{aligned} \sigma_{t,n}^2 &= s_{t,n}^2\omega + \frac{\alpha s_{t,n}^2}{s_{n-1}^2}(\epsilon_{t,n-1} - s_{t,n-1}\kappa)^2 + \frac{\beta s_{t,n}^2}{s_{t,n-1}^2}\sigma_{t,n-1}^2 \\ &= s_{t,n}^2\omega + \frac{\alpha s_{t,n}^2}{s_{t,n-1}^2}(\epsilon_{t,n-1}^2 - 2s_{t,n-1}\kappa\epsilon_{t,n-1} + s_{t,n-1}^2\kappa^2) + \frac{\beta s_{t,n}^2}{s_{t,n-1}^2}\sigma_{t,n-1}^2 \\ &= (s_{t,n}^2\omega + \alpha s_{t,n}^2\kappa^2) + \frac{\alpha s_{t,n}^2}{s_{t,n-1}^2}\epsilon_{t,n-1}^2 + \frac{\beta s_{t,n}^2}{s_{t,n-1}^2}\sigma_{t,n-1}^2 \\ &\quad - \frac{\alpha s_{t,n}^2}{s_{t,n-1}^2}2s_{t,n-1}\kappa\epsilon_{t,n-1} \end{aligned} \tag{3.6}$$

Substituting $\omega_{s(t,n)} = s_{t,n}^2\omega + \alpha s_{t,n}^2\kappa^2$, $\alpha_{s(t,n)} = \alpha s_{t,n}^2/s_{t,n-1}^2$ and $\beta_{s(t,n)} = \beta s_{t,n}^2/s_{t,n-1}^2$, the above equation becomes:

$$\sigma_{t,n}^2 = \omega_{s(t,n)} + \alpha_{s(t,n)}\epsilon_{t,n-1}^2 + \beta_{s(t,n)}\sigma_{t,n-1}^2 - 2\alpha_{s(t,n)}s_{t,n-1}\kappa\epsilon_{t,n-1} \tag{3.7}$$

The coefficients $\omega_{s(t,n)}$, $\alpha_{s(t,n)}$ and $\beta_{s(t,n)}$ vary with the change of the periodicity $s_{t,n}$ along the time index (t, n). The equation 3.7 can be considered as a modified version of the P-GARCH structure with an additional term, $-2\alpha_{s(t)}s_{n-1}\kappa\epsilon_{t-1}$, influences the change of σ_t^2.

On the other hand, by letting $v_{t,n} \equiv \epsilon_{t,n}^2 - \sigma_{t,n}^2$ and assuming $v_{t,n}$ to be i.i.d. zero-mean normally distributed. The above equation can be re-written as:

$$\begin{aligned}
\epsilon_{t,n}^2 - v_{t,n} &= \omega_{s(t,n)} + \alpha_{s(t,n)}\epsilon_{t,n-1}^2 + \beta_{s(t,n)}(\epsilon_{t,n-1}^2 - v_{t,n-1}) \\
&\quad -2\alpha_{s(t,n)}s_{t,n-1}\kappa\epsilon_{t,n-1} \\
\epsilon_{t,n}^2 &= \omega_{s(t,n)} + (\alpha_{s(t,n)} + \beta_{s(t,n)})\epsilon_{t,n-1}^2 - \beta_{s(t,n)}v_{t,n-1} + v_{t,n} \\
&\quad -2\alpha_{s(t,n)}s_{t,n-1}\kappa\epsilon_{t,n-1} \qquad (3.8)
\end{aligned}$$

It can be observed that the squared intraday innovations are modeled by a periodic ARMA structure with the additional term. The periodic ARMA structure provides the capability to capture the repetitive periodic autocorrelation pattern which is prevalent in most intraday innovation series. Our integrated framework aims to search for the best parameter set under the periodic structure.

The above elaboration can be extended for the situations where the innovations of the filtered series are modeled by an ARMA structure. In other words, the above argument can be applied to the situations of using the ARCH [31], the GARCH [19], the IGARCH [32], the FIGARCH [13] and the HYGARCH [29] structures to model the filtered series. However, it is noteworthy that the benefit of applying the integrated framework is limited for some variants of ARCH structures. For example, the EGARCH [61] and the

FIEGARCH [22] structures handle the log of the conditional variances with an ARMA structure instead. The integration of the filtration process can only facilitate part of the parameters in the corresponding variance equation to be periodically varying. As a result, it is not recommended to apply the integrated framework on those structures.[1]

[1]The limitation of the integrated framework for the EGARCH(1,1) structure is provided in the supplementary note 3.8.1. In addition, the proposed framework does not give a significant performance gain over Andersen & Bollerslev's sequentially estimation method in our preliminary empirical investigation when either the EGARCH(1,1) or the FIGARCH(1,d,1) structure is used.

3.4 Model specifications and Evaluation criteria

It is speculated that the proposed integrated framework would give a more favorable performance than the traditional sequential estimation method do, for the situations that the filtered series is modeled as an ARMA structure. This speculation will be checked with the use of two distinct intraday periodicities together with three different ARCH structures. The following subsections detail the corresponding specifications and the evaluation criteria.

3.4.1 Specifications of filtration process and ARCH structures

Andersen & Bollerslev have proposed two ways to estimate the intraday periodicities for the removal of the periodic effect in an intraday innovation series. The first way is to estimate the intraday periodicity as the mean absolute value for the corresponding interval [2].

$$s_{t,n} = \frac{\sum_{i=T_1}^{T_2} |r_{i,n}|}{T} \tag{3.9}$$

where $r_{i,n}$ is the intraday return of the n-th intraday interval on day i. T is the total number of days in the sample, T_1 and T_2 are the indexes of the start day and the end day respectively.

Apart from the above approach, the intraday periodicity can be estimated with appropriate assumptions on the underlying intraday dynamics [3]. It is assumed that the intraday return follows:

$$r_{t,n} = E(r_{t,n}) + s_{t,n} \cdot \frac{\sigma_t}{\sqrt{N}} \cdot z_{t,n} \tag{3.10}$$

where $z_{t,n}$ is an i.i.d. zero-mean unit-variance variable and N denotes the number of intraday interval per day. The method makes use of the modified Flexible Fourier Form (FFF) to estimate $s_{t,n}$ through an intermediate proxy variable $\hat{x}_{t,n}$ under the following settings:

$$\begin{aligned} x_{t,n} &= 2log[|r_{t,n} - E(r_{t,n})|] - log(\sigma_t^2) + log(N) \\ \hat{x}_{t,n} &\equiv f(\hat{\Theta}|n) \end{aligned} \tag{3.11}$$

where $\hat{x}_{t,n}$ is an approximation of the underlying periodic dynamic. The function $f(\Theta|n)$ is a modified version of the flexible Fourier form[39] which is defined as :

$$f(\Theta|n) = \sum_{j=0}^{J} \nu_j \cdot n^j + \sum_{p=1}^{P}(\gamma_p cos\frac{pn2\pi}{N} + \delta_p sin\frac{pn2\pi}{N}) \tag{3.12}$$

where Θ represents the parameter set $\{\nu_j, \gamma_p, \delta_p\}$. Variables J and P control the order of expansion. The parameter set Θ is estimated by setting $f(\Theta|n)$ as a regressor (independent variable) for the dummy variable $x_{t,n}$ through a linear regression with zero mean i.i.d. error terms.[2].

After obtaining the approximated value of the intraday dynamics $\hat{x}_{t,n}$, a

[2]The order of expansion for the flexible Fourier form $f(\Theta|n)$ in the equation 3.12 is set to $J = 2$ and $P = 4$. Expansion beyond this order gives insignificant coefficients for the additional ν_j, γ_p and δ_p under our empirical investigation. Furthermore, the GARCH(1,1) conditional volatility is considered as the substitute of σ_t in this paper (the same substitute as Andersen & Bollerslev's original work [3])

normalization step is used to retrieve the intraday periodicity $s_{t,n}$ accordingly.

$$s_{t,n} = \frac{S \cdot exp(\frac{\hat{x}_{t,n}}{2})}{\sum_{i=1}^{T/N} \sum_{j=1}^{N} exp(\frac{\hat{x}_{i,j}}{2})} \tag{3.13}$$

where S is the sample size of the intraday return series.

Besides checking the capability for two distinct filtration methods, our proposed framework will also be tested on different ARCH structures. Three structures, GARCH(1,1), FIGARCH(1,d,1) [13] and HYGARCH(1,d,1) [29], will be considered to handle the heteroskedasticity of the filtered series respectively. For the case that FIGARCH(1,d,1) is used, the variance equation of the filtered series (i.e. the second line in the equation set 3.4) will be changed to :

$$\begin{aligned} \tilde{\sigma}_{t,n}^2 &= \omega + [1 - \beta L - (1 - \alpha L) \cdot (1 - L)^d] \cdot (\tilde{\xi}_{t,n-1})^2 \\ &\quad + \beta \tilde{\sigma}_{t,n-1}^2 \end{aligned} \tag{3.14}$$

where the model parameters become ω, α, β and d. The notation L is the lag operator and $(1-L)^d$ is the fractional differencing operator. The parameters are constrained to obey $\omega > 0$, $0 \leq d \leq 1 - 2\alpha$ and $0 \leq \beta \leq \alpha + d$ to ensure the positivity of $\tilde{\sigma}_{t,n}^2$.

On the other hand, the HYGARCH(1,d,1) structure is defined as:

$$\begin{aligned} \tilde{\sigma}_{t,n}^2 &= \omega + \left\{1 - \beta L - (1 - \delta L) \cdot (1 + \alpha[(1 - L)^d - 1])\right\} \cdot (\tilde{\xi}_{t,n-1})^2 \\ &\quad + \beta \tilde{\sigma}_{t,n-1}^2 \end{aligned} \tag{3.15}$$

where $\omega, \alpha, \beta, \delta$ and d are the model parameters and $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$

and $d \geq 0$.

The HYGARCH model nests the GARCH, IGARCH and FIGARCH. The GARCH and the FIGARCH models correspond to $\alpha = 0$ and $\alpha = 1$ respectively when $0 < d < 1$. When $d = 1$, the HYGARCH model reduces to the GARCH for $\alpha = 0$ or the IGARCH for $\alpha = 1$.

3.4.2 Evaluation criteria

Several methods are selected for the comparison of model performance, including regression R^2, Mean Squared Error (MSE), Mean Absolute Error (MAE) and Diebold-Mariano test (DM test). According to a comprehensive review on the evaluation methods for volatility forecasting from Poon and Granger [65], these measurements reflect the performance in different perspectives and therefore, their results help to deliver a better overall picture. Furthermore, beside the investigation of one-step-ahead out-of-sample forecasts, several forecasting horizons will also be studied to reflect the trends among the multi-step ahead performances.

The regression R^2 from the Mincer and Zarnowitz type regression is employed in this study to measure the in-sample and out-of-sample model performances. The forecasted volatilities are tested against the 'observed volatility'. The regression takes the form as:

$$V_i^{1/2} = b_0 + b_1 \cdot \hat{\sigma}_{i,model} + u_i \tag{3.16}$$

where V_i is the 'observed variance' and $\hat{\sigma}_{i,model}$ denotes the forecasted volatility from the corresponding model. The subscript i is the index for the time series and u_i is a zero mean i.i.d. error term.

A commonly used representative for the 'observed variance' in the literature is the squared return. It is an unbiased estimator of the variance for a zero mean return process. The use of the squared return as the estimator of the 'observed variance' for a single intraday interval poses a natural extension to the multiple-step-ahead variance estimator, where multiple 'observed variance' are summed together to represent the variance for the involved period [7, 24, 67, 69]. This method matches the concept of realized variance [5, 6, 15] and it has been demonstrated that the summation of multiple squared returns can give a reasonably precise estimation of the variance for the involved period when the sampling-frequency is no more frequently than 5-minute [11, 9, 45, 14, 63].[3]

The application of the squared returns as the 'observed variance' suits well for the situation where the returns are zero-mean. To account for non-zero mean situations, the squared innovation may give a more accurate measurement instead [56, 57]. The results listed in this paper will be based on the use of the squared innovation as the 'observed variance'. [4]

The innovation is defined as:

$$\xi_i = r_i - E_i(r) \tag{3.17}$$

where ξ_i and r_i are the innovation and the return of the period i respectively. $E_i(r)$ is the expected value (unconditional mean) of the returns given the

[3]Properties of the realized variance estimator is provided in the supplementary note 3.8.4

[4]Same conclusions are drawn from the empirical analyses with the used of the squared return as the 'observed variance'. Empirical results with the use of squared return as the measurement can be obtained by contacting the author.

information up to the period i.

The Mean Squared Error (MSE) and the Mean Absolute Error (MAE) are selected as the second category of measurements. They indicate the differences between the forecasted volatility and the 'observed volatility':

$$\begin{aligned} MSE &= \frac{1}{F}\sum_{i=1}^{F}(V_i^{1/2} - \hat{\sigma}_{i,model})^2 && (3.18) \\ MAE &= \frac{1}{F}\sum_{i=1}^{F}|V_i^{1/2} - \hat{\sigma}_{i,model}| && (3.19) \end{aligned}$$

where F denotes the number of forecasted results.

Furthermore, the Diebold-Mariano test (DM test) [30] is used to compare the predictive accuracy between the sequential estimation method and our proposed integrated framework. The forecasted volatility from the sequential method ($\hat{\sigma}_{i,base}$) is considered as the benchmark for the comparison in this study. The test statistics (DM) is calculated as:

$$\begin{aligned} DM &= \frac{\overline{\delta_i}}{(\widehat{Var(\overline{\delta_i})})^{1/2}} \\ \delta_i &= L(V_i^{1/2}, \hat{\sigma}_{i,base}) - L(V_i^{1/2}, \hat{\sigma}_{i,model}) && (3.20) \end{aligned}$$

where $\overline{\delta_i}$ and $\widehat{Var(\overline{\delta_i})}$ denote the mean of δ_i and the Newey-West heteroskedasticity and autocorrelation consistent (HAC) variance of $\overline{\delta_i}$ respectively.

$L(V_i^{1/2}, \hat{\sigma}_{i,model})$ is the loss function for a specific measurement. The loss function for squared error and absolute error are defined by the equations 3.21 and 3.22.

$$L_{SE}(V_i^{1/2}, \hat{\sigma}_{i,model}) = (V_i^{1/2} - \hat{\sigma}_{i,model})^2 \quad (3.21)$$

$$L_{AE}(V_i^{1/2}, \hat{\sigma}_{i,model}) = |V_i^{1/2} - \hat{\sigma}_{i,model}| \tag{3.22}$$

The null hypothesis is $H_0 : E(\delta) \doteq 0$ while the alternatives are $H_1^A : E(\delta) > 0$ and $H_1^B : E(\delta) < 0$. The test statistics is asymptotically normal distributed. The hypothesis H_1^A indicates the selected model performs better than the base model and H_1^B gives a reverse indication.

In additional to the investigation of 1-step-ahead forecasts, the forecasting performance will also be checked for multiple horizon situations. The variances of involved periods are summed together to represent an accumulated variance that covers the corresponding multiple horizons. The forecasted volatility for m-horizon with the use of information up to the period i is defined as:

$$\left(\sum_{j=1}^{m}(\hat{\sigma}_{i+j,model})^2\right)^{1/2} \tag{3.23}$$

where $\hat{\sigma}_{i+j,model}$ denotes the forecasted volatility for the $(i+j)$-th period with the use of information up to the period i. For the period $i+1$ to $i+m$, the forecasted volatilities are calculated by the model with the same parameters. The $\hat{\sigma}_{i+1,model}$ is the first-step-ahead forecast, $\hat{\sigma}_{i+2,model}$ is the second-step-ahead forecast and so on.

Similarly, the square root of the 'observed variance' that covers m-horizon is defined as:

$$\left(\sum_{k=i+1}^{i+m}(V_k)^2\right)^{1/2} \tag{3.24}$$

3.5 Simulation study on the effectiveness of the integrated framework

The advantages of the proposed integrated framework over Andersen & Bollerslev's sequential approach has been discussed in section 3.3 and this section covers a set of simulation experiments which helps to reflect the magnitude of the potential improvement. The simulation data representing the intraday innovation series will be generated under the following procedure :

1. Generate a temporary conditional volatility series $\sigma'_{t,n}$ with a GARCH(1,1) structure.

2. Multiply $\sigma'_{t,n}$ with a pre-determined U-shape periodic factor series $U_{t,n}$ to form an intraday volatility series $\sigma_{t,n}$. The U-shape factor is determined by the empirical information (NASDAQ data series) used in this thesis.

3. Generate a temporary series $x'_{t,n}$ with $\sigma^2_{t,n}$ as its variance parameter. The $x'_{t,n}$ series is either normally distributed or t-distributed.

4. To reflect situations where the actual intraday series deviates from a specific ARCH structure, the final simulated data series $x_{t,n}$ is constructed as $x_{t,n} = x'_{t,n} + y_{t,n}$. The $y_{t,n}$ is a random term with user-specified occurrence probability and intensity (i.e. $y_{t,n}$ = intensity x occurrence probability) .

The second step in our data generating procedure takes care the repetitive U-shape pattern appeared in an intraday volatility process. The magnitude of

$U_{t,n}$ is calculated to mimic the conditions observed in the empirical NASDAQ data set.[5]

The above procedure is capable of producing a more 'realistic' intraday series where the generated series $x_{t,n}$ can be observed to deviate from a structure which is perfectly described by a simple ARCH structure. There are two possible methods to generate data with deviations from a perfect ARCH structure and they are : 1) altering the $\sigma_{t,n}$ series and use it as the underlying variance series; 2) adding random term to the data series.

The first method does not give a satisfactory result since the generated data does not produce the commonly observed U-shape pattern. However, the addition of a random term always shows the maintenance of the U-shape pattern in the ACF of squared observations. As a result, the second method is resorted to the fourth step in the data generating procedure. The random deviation $y_{t,n}$ is modeled as two components, intensity and occurrence probability. The intensity is specified as the k-th percentile of the absolute value of $x_{t,n}$ series whereas the occurrence probability denotes the likelihood of the presence of the deviation.

The simulation results with the assumption of normally distributed and t-distributed (with degree of freedom equals to 5) innovations are tabulated in table 3.1. The α and β of the GARCH(1,1) structure are specified as 0.07 and 0.925 respectively. The ω (long term volatility level) and the magnitude of the $U_{t,n}$ series are determined to minic the properties appeared in the empirical NASDAQ data set with the use of FFF deseasonalization method.

[5]The variation of $U_{t,n}$ series is determined by the FFF parameters obtained from the empirical NASDAQ data set and the overall magnitude of the series is maintained to follow the empirical ratio.

The intensity for the random term $y_{t,n}$ is set to the 98-th percentile of the absolute $x_{t,n}$ series. The length of the intraday series is 4,914 by setting the parameters with 126 days and 39 intraday intervals per day. Each specific setting is repeated 10,000 times to obtain the overall simulation result.

The simulation result in table 3.1 shows that the integrated framework approach achieves better results in terms of the averaged R^2, MSE and MAE for all the situations. The improvements obtained by using the integrated framework approach become more obvious when the level of deviation, represented by the occurrence probability of the random term, changes from 0 to 0.1. In other words, the degree of improvement increases when the simulated data becomes more deviated from the 'perfectly structured' intraday innovations. For the situation where the occurrence probability = 0.1, the improvement of R^2, MSE and MAE are 3.72%, 0.64% and 0.38% for normally distributed innovations while the improvements are 5.59%, 4.48% and 3.71% for t-distributed innovations with Dof = 5.

Apart from comparing the magnitude of MSE and MAE between the two estimation approaches, Diebold-Mariano hypothesis tests are conducted to reflect the model performance in another aspect. The averaged test statistics (t-statistics) are positive for the situations and this reflects the integrated framework tends to outperform the traditional sequential approach. Furthermore, similar as the trend observed in MSE and MAE, the benefits from the integrated framework become more obvious with the increase of level of deviation.

The framework significantly outperforms (5% of level of significance) the sequential approach in squared error and absolute error for 58.57% and

83.29% of the simulation samples (10,000 repetitions), for normally distributed innovations when the occurrence probability = 0.1. For t-distributed innovations, the proposed framework significantly gives better results in squared error and absolute error for 84.93% and 83.89% respectively.

Overall, the simulation results of R^2, MSE, MAE and Diebold-Mariano hypothesis tests indicates a positive preference on the integrated framework over the sequential estimation method.

Table 3.1: Simulation results (10,000 repetitions for each situation)

Normally distributed innovations

Random term $y_{t,n}$	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	DM-SE (Sig %)	DM-AE (Sig %)
Nil	Int. framework	0.000	0.781	**0.261**	**7.321E-04**	**1.876E-02**	0.426	-0.031
				(0.05%)	(0.12%)	(0.03%)	(21.93%)	(9.55%)
	Sequential	0.000	0.779	0.261	7.329E-04	1.876E-02	-	-
Occurrence Prob = 0.05	Int. framework	-0.007	0.848	**0.153**	**1.417E-03**	**2.750E-02**	1.780	2.860
				(2.30%)	(0.60%)	(0.39%)	(48.97%)	(76.90%)
	Sequential	-0.007	0.854	0.150	1.426E-03	2.760E-02	-	-
Occurrence Prob = 0.1	Int. framework	-0.009	0.863	**0.110**	**1.925E-03**	**3.356E-02**	2.087	3.015
				(3.72%)	(0.64%)	(0.38%)	(58.57%)	(83.29%)
	Sequential	-0.009	0.868	0.106	1.937E-03	3.369E-02	-	-

t-distributed innovations with degree of freedom = 5

Random term $y_{t,n}$	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	DM-SE (Sig %)	DM-AE (Sig %)
Nil	Int. framework	0.001	0.694	**0.184**	**1.635E-03**	**2.699E-02**	1.165	1.069
				(0.08%)	(0.06%)	(0.03%)	(37.82%)	(33.84%)
	Sequential	0.001	0.693	0.183	1.636E-03	2.700E-02	-	-
Occurrence Prob = 0.05	Int. framework	-0.003	0.658	**0.115**	**3.058E-03**	**4.179E-02**	8.909	11.550
				(2.83%)	(1.05%)	(1.09%)	(80.67%)	(79.53%)
	Sequential	-0.003	0.645	0.112	3.091E-03	4.226E-02	-	-
Occurrence Prob = 0.1	Int. framework	-0.008	0.686	**0.085**	**4.182E-03**	**5.188E-02**	19.640	25.752
				(5.59%)	(4.48%)	(3.71%)	(84.93%)	(83.89%)
	Sequential	-0.007	0.653	0.080	4.378E-03	5.388E-02	-	-

Note: The values of b_0, b_1, R^2, MSE, MAE, DM-SE and DM-AE are the averaged results from all the training samples. The DM-SE and DM-AE are the test statistics for squared error and absolute error respectively. The Sig % denotes the percentage of the acceptance of the alternative hypothesis, the integrated framework outperforms the sequential method, with p-value<0.05 for all the training samples. The improved % represents the averaged improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

3.6 Empirical investigation

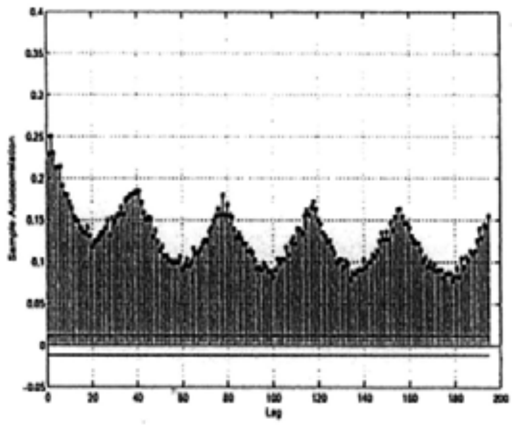
The data used in our experiment is composed of 10-minute returns for the NASDAQ composite stock index (NASDAQ) from 15 August 2005 to 12 Sept 2008, consisting of 28,392 observations (728 days), and the Standard and Poors composite stock index (S&P 500) from 10 March 2006 to 12 Sept 2008, consisting of 22,893 observations (587 days). The returns are recorded from 9:30 to 16:00 and therefore there are a total of 39 intraday returns per trading day. All the values of intraday returns are multiplied by 100 for the ease of presentation. The correlogram of the absolute return series of the two data sets, NASDAQ and S&P 500, are plotted in Figure 3.1 and their descriptive statistics are tabulated in Table 3.2.

The correlograms of the absolute returns in Figure 3.1a and Figure 3.1b illustrate the periodic autocorrelation pattern in both NASDAQ and S&P 500 data. The autocorrelations vary in a U-shape pattern and its magnitude declines slowly through time. The distinctive pattern suggests the existence of periodicity and heteroskedasticity in the series and validates the use of Andersen & Bollerslev's approach and periodic ARMA structures to model the underlying series.

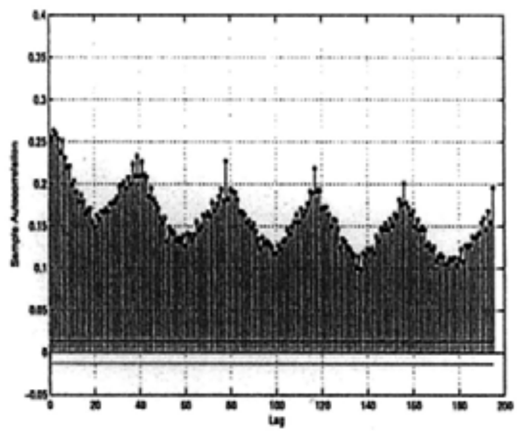
The results from Table 3.2 indicate that the intraday returns are significantly rejected by the Jarque-Bera normality test and exhibit excess Kurtosis, for both the NASDAQ and the S&P 500 data sets. It is advisable to consider other distribution types to model the intraday innovations beside the usage of the normal distribution. The empirical works from Bollerslev and Wilhelmsson demonstrated that the use of t-distributed errors helps

Figure 3.1: Correlogram of the absolute intraday returns, 5-day period - (a) NASDAQ, (b) S&P 500

(a) NASDAQ



(b) S&P 500



to handle the high kurtosis in the data when GARCH(1,1) model is employed [20, 71]. Therefore, apart from investigating the situation of normally distributed innovations, the assumption of t-distributed innovations will also be addressed. [6]

The model parameters are estimated by a rolling-sample method in this paper. This method updates the training samples step by step. For instance,

[6]The likelihood function for t-distribution is based on Bollerslev's work on 1987. [20]

Table 3.2: Descriptive statistics of the intraday returns (multiplied by 100)

	NASDAQ	S&P 500
Start date	16-Aug-05	13-Mar-06
End date	12-Sep-08	12-Sep-08
Observation	28,392	22,893
Mean	-0.0008	-0.0001
Max	1.1768	1.4236
Min	-1.4115	-1.4297
Standard deviation	0.1400	0.1362
Skewness	-0.0175	0.1770
Robust skewness	-0.0111	-0.0088
Kurtosis	7.7803	10.8744
Jarque-Bera (p-value)	<0.001	<0.001

Note: The robust skewness [49], which is robust against outliers, is calculated as $\frac{q3+q1-2q2}{q3-q1}$. The Jarque-Bera value is the p-value of the Jarque-Bera test [48] with the null hypothesis that the intraday returns are normally distributed.

if the sample size is S, the first sample for model training will be the 1 to S observations, the second sample will be the 2 to $S+1$ observations and so on. The sample size is set to be 200 days (7,800 intraday periods) in this study. There are 528 days (20,592 periods) and 387 days (15,093 periods) out-of-sample observations for the NASDAQ and the S&P 500 data sets respectively.

3.6.1 In-sample fitness

The performance of the integrated framework approach is investigated under several model settings. Combinations of two filtration processes, three ARCH structures and two distributional assumptions are engaged to illustrate the effects. For clarity, the filtration process with the use of estimated intraday periodicity determined by the equation 3.9 is named as the simple averaging method while the second way to estimate the periodicity (equations 3.11 to 3.13) is named as the FFF method. The results are tabulated in a multiple columns format, where the columns in the middle represents the values with the assumption of normally distributed innovations and the columns on the right side represent those from t-distributed innovations.

The in-sample results for various situations under the simple averaging filtration method are listed in Table 3.3 and the results with the use of the FFF method are listed in Table 3.4. Similarly, the values of the Diebold-Mariano tests are tabulated in Table 3.5 and Table 3.6 for the two filtration methods respectively. The values in the above tables are the averaged results from all the training samples.

According to the results from the pair-wise comparisons, the use of an integrated framework gives more favorable performance than the traditional sequential approach does for most situations. There are a total of 12 situations investigated for each data set (i.e. two filtration methods, three ARCH structures, two distributional assumptions) and all of them demonstrate positive improvements on the regression R^2. The improvement of R^2 ranges from 2.53% - 2.97% and 2.06% - 2.89% for the NASDAQ and the S&P 500 respectively.

The integrated framework approach also shows positive improvements of MSE and MAE. Among all situations, the MSE are shown to be increased by 0.66% - 1.67% and 0.23% - 2.37% for the two data sets respectively. For

MAE, all results indicate positive improvements, except the situation of using simple averaging filtration method with the FIGARCH(1,d,1) structure under the t-distributional assumption for the S&P 500 data set, which produces a worser performance (-0.13%).

Table 3.3: In-sample fit - Deseasonalized by simple averaging

NASDAQ

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	-0.0019	0.773	**0.162**	**8.500E-03**	**6.979E-02**	0.0041	0.727	**0.162**	**8.599E-03**	**6.984E-02**
				(2.77%)	(1.19%)	(0.57%)			(2.84%)	(1.43%)	(0.67%)
	Sequential	0.0002	0.756	0.158	8.603E-03	7.019E-02	0.0062	0.708	0.158	8.724E-03	7.032E-02
FIGARCH(1,d,1)	Int. framework	-0.0035	0.790	**0.167**	**8.401E-03**	**6.937E-02**	0.0017	0.750	**0.166**	**8.500E-03**	**6.954E-02**
				(2.58%)	(1.03%)	(0.53%)			(2.65%)	(0.66%)	(0.24%)
	Sequential	-0.0017	0.775	0.162	8.489E-03	6.974E-02	0.0035	0.736	0.162	8.557E-03	6.970E-02
HYGARCH(1,d,1)	Int. framework	-0.0033	0.787	**0.167**	**8.423E-03**	**6.949E-02**	0.0017	0.749	**0.166**	**8.493E-03**	**6.952E-02**
				(2.54%)	(1.16%)	(0.58%)			(2.53%)	(1.52%)	(0.72%)
	Sequential	-0.0014	0.771	0.162	8.522E-03	6.990E-02	0.0041	0.727	0.162	8.624E-03	7.003E-02

S&P 500

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	-0.0012	0.768	**0.221**	**8.433E-03**	**6.513E-02**	0.0019	0.745	**0.221**	**8.484E-03**	**6.508E-02**
				(2.36%)	(1.63%)	(0.73%)			(2.45%)	(2.00%)	(0.90%)
	Sequential	0.0012	0.747	0.216	8.573E-03	6.561E-02	0.0044	0.721	0.216	8.657E-03	6.567E-02
FIGARCH(1,d,1)	Int. framework	-0.0023	0.781	**0.224**	**8.352E-03**	**6.490E-02**	0.0011	0.751	**0.223**	**8.473E-03**	6.511E-02
				(2.26%)	(1.26%)	(0.46%)			(2.06%)	(0.23%)	(-0.13%)
	Sequential	0.0000	0.762	0.219	8.459E-03	6.520E-02	0.0028	0.741	0.219	8.492E-03	**6.503E-02**
HYGARCH(1,d,1)	Int. framework	-0.0022	0.782	**0.224**	**8.351E-03**	**6.486E-02**	0.0015	0.751	**0.223**	**8.454E-03**	**6.494E-02**
				(2.15%)	(1.84%)	(0.82%)			(2.13%)	(2.23%)	(1.01%)
	Sequential	0.0004	0.757	0.219	8.508E-03	6.539E-02	0.0041	0.725	0.218	8.647E-03	6.560E-02

Note: The values of b_0, b_1, R^2, MSE and MAE are the averaged results from all the training samples. The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.4: In-sample fit - Deseasonalized by FFF method

NASDAQ

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	0.0057	0.693	**0.153**	**8.824E-03**	**7.098E-02**	0.0118	0.650	**0.153**	**8.945E-03**	**7.101E-02**
				(2.82%)	(1.31%)	(0.61%)			(2.97%)	(1.66%)	(0.75%)
	Sequential	0.0081	0.675	0.149	8.941E-03	7.142E-02	0.0140	0.630	0.149	9.095E-03	7.154E-02
FIGARCH(1,d,1)	Int. framework	0.0036	0.716	**0.157**	**8.685E-03**	**7.047E-02**	0.0092	0.674	**0.157**	**8.817E-03**	**7.064E-02**
				(2.66%)	(1.17%)	(0.55%)			(2.76%)	(0.74%)	(0.26%)
	Sequential	0.0058	0.698	0.153	8.787E-03	7.087E-02	0.0111	0.660	0.153	8.882E-03	7.083E-02
HYGARCH(1,d,1)	Int. framework	0.0037	0.714	**0.157**	**8.704E-03**	**7.056E-02**	0.0092	0.675	**0.158**	**8.801E-03**	**7.058E-02**
				(2.59%)	(1.35%)	(0.66%)			(2.55%)	(1.67%)	(0.78%)
	Sequential	0.0061	0.694	0.154	8.824E-03	7.103E-02	0.0115	0.654	0.154	8.950E-03	7.114E-02

S&P 500

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	-0.0003	0.754	**0.206**	**8.660E-03**	**6.609E-02**	0.0032	0.728	**0.206**	**8.704E-03**	**6.597E-02**
				(2.81%)	(1.83%)	(0.77%)			(2.89%)	(2.23%)	(0.98%)
	Sequential	0.0027	0.727	0.200	8.821E-03	6.661E-02	0.0062	0.699	0.200	8.902E-03	6.662E-02
FIGARCH(1,d,1)	Int. framework	-0.0013	0.765	**0.209**	**8.574E-03**	**6.582E-02**	0.0035	0.724	**0.207**	**8.776E-03**	**6.625E-02**
				(2.70%)	(1.43%)	(0.51%)			(2.37%)	(0.78%)	(0.12%)
	Sequential	0.0015	0.742	0.203	8.698E-03	6.615E-02	0.0057	0.709	0.202	8.845E-03	6.633E-02
HYGARCH(1,d,1)	Int. framework	-0.0011	0.766	**0.208**	**8.576E-03**	**6.579E-02**	0.0030	0.733	**0.207**	**8.677E-03**	**6.584E-02**
				(2.47%)	(1.98%)	(0.86%)			(2.45%)	(2.37%)	(1.04%)
	Sequential	0.0018	0.738	0.203	8.750E-03	6.636E-02	0.0058	0.704	0.202	8.888E-03	6.653E-02

Note: The values of b_0, b_1, R^2_r, MSE and MAE are the averaged results from all the training samples. The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Although the magnitude of improvement percentages on MSE and MAE are not large, there are strong indications on the preference to our proposed framework. The results of Diebold-Mariano hypothesis test listed in Table 3.5 and Table 3.6 indicate that the integrated framework produces smaller squared errors and absolute errors in large portions of training samples at 5% significance level. For the NASDAQ data, our framework is verified to give smaller squared errors for 62% - 83% of training samples and smaller absolute error for 63.2% - 83% of samples. For the S&P 500, our framework significantly outperforms the sequential estimation method by giving smaller squared errors and absolute errors for at least 71.2% of samples in 10 out of 12 situations. The use of integrated framework is supported with fewer samples when the FIGARCH(1,d,1) structure is employed under the t-distributional assumption. Nevertheless, the overall results show a strong perference to our framework. [7]

---

[7]Perference to the integrated framework is also apparent when the squared return is used as the 'observed variance'. The two measurements, squared return vs. squared innovation, lead to almost the same results (i.e. less than 1% difference for in-sample and out-of-sample R^2, MSE and MAE). The corresponding results can be obtained by contacting the author.

Table 3.5: Diebold-Mariano test, In-sample fit for various ARCH structure - Deseasonalized by simple averaging

NASDAQ

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %
GARCH(1,1)	4.505	70.3%	4.793	80.1%	5.539	83.0%	5.542	83.0%
FIGARCH(1,d,1)	3.815	72.4%	4.436	83.0%	3.191	73.5%	2.518	63.2%
HYGARCH(1,d,1)	4.217	68.4%	4.783	78.9%	5.916	83.0%	6.093	83.0%

S&P 500

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %
GARCH(1,1)	5.719	78.0%	6.337	78.0%	6.594	78.0%	7.365	78.0%
FIGARCH(1,d,1)	4.321	71.2%	3.929	72.0%	1.612	53.3%	-0.658	30.6%
HYGARCH(1,d,1)	5.747	74.9%	6.649	78.0%	6.945	78.0%	8.030	78.0%

Note: The DM-SE and DM-AE are the averaged results of the test statistics for squared error and absolute error respectively. The alternative hypothesis indicates that the integrated framework outperforms the sequential method. The Sig % denotes the percentage of the acceptance of the alternative hypothesis with p-value<0.05 for all the training samples.

Table 3.6: Diebold-Mariano test, In-sample fit for various ARCH structure - Deseasonalized by the FFF method

NASDAQ

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %
GARCH(1,1)	4.080	59.8%	4.500	75.5%	5.614	83.0%	5.671	83.0%
FIGARCH(1,d,1)	3.553	63.7%	4.093	82.8%	3.167	71.2%	2.560	63.2%
HYGARCH(1,d,1)	4.121	62.0%	4.837	77.0%	5.867	82.1%	6.283	83.0%

S&P 500

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %	t-statistics	Sig. %
GARCH(1,1)	6.372	78.0%	6.428	78.0%	7.285	78.0%	7.677	78.0%
FIGARCH(1,d,1)	4.874	72.9%	4.199	75.4%	3.067	61.7%	1.021	39.6%
HYGARCH(1,d,1)	6.161	77.9%	6.776	78.0%	7.301	78.0%	8.045	78.0%

Note: The DM-SE and DM-AE are the averaged results of the test statistics for squared error and absolute error respectively. The alternative hypothesis indicates that the integrated framework outperforms the sequential method. The Sig % denotes the percentage of the acceptance of the alternative hypothesis with p-value<0.05 for all the training samples.

3.6.2 Out-of-sample intraday forecasting performance

The out-of-sample performance over different forecasting horizons for the GARCH(1,1), the FIGARCH(1,d,1) and the HYGARCH(1,d,1) structures are tabulated in Table 3.7, Table 3.8 and Table 3.9 respectively, for the simple averaging filtration method. In addition, the results for the FFF filtration method are tabulated in Table 3.10 to Table 3.12.

Similar to the findings from the in-sample fit analyses, the integrated framework is shown to produce superior performance in terms of R^2, squared errors and absolute errors empirically. The forecasts from the proposed framework produces higher R^2 than the corresponding results from the sequential approach under all forecasting horizons, varying from 1-step-ahead to 25-step-ahead, for both the NASDAQ and S&P 500 data.

Positive improvements in MSE and MAE are observed in all the 12 situations for the NASDAQ data. The improvement percentage of MSE ranges from 0.87% to 9.7% over various horizons while the improvement percentage of MAE ranges from 0.35% to 7.48%. For the S&P 500 data, all situations demonstrate positive improvements of MSE and MAE by the integrated framework, except the 1-step-ahead forecast under the FIGARCH(1,d,1) structure with the t-distributional assumption, when the simple averaging filtration method is employed.

The results of the Diebold-Mariano test are listed in Table 3.13 and Table 3.14. The integrated framework always produces significantly smaller (at 5% significance level) squared errors and absolute errors among various forecating horizons for both data sets. However, by comparing the results over several forecasting horizons, it is observed that the statistical significance of the Diebold-Mariano tests tend to decrease with the increase of forecasting horizon for both the NASDAQ and the S&P 500 data. Nevertheless, the test results still significantly indicate a positive preference to the integrated framework for 25-step-ahead forecasts under all situations.[8]

[8]For the ease of reading, graphical representations of the forecast performance in terms of R^2, MSE and MAE are also included in the supplementary note 3.8.3

Table 3.7: Out-of-sample fit, GARCH(1,1) - Deseasonalized by simple averaging

NASDAQ

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0049	0.749	**0.185**	**1.057E-02**	**7.692E-02**	0.0083	0.717	**0.186**	**1.073E-02**	**7.742E-02**
				(1.80%)	(1.50%)	(0.84%)			(1.67%)	(1.73%)	(0.93%)
	Sequential	0.0071	0.727	0.182	1.073E-02	7.757E-02	0.0104	0.695	0.183	1.092E-02	7.814E-02
3	Int. framework	0.0115	0.863	**0.360**	**1.514E-02**	**8.862E-02**	0.0191	0.823	**0.360**	**1.540E-02**	**8.936E-02**
				(2.25%)	(2.66%)	(1.62%)			(2.03%)	(3.01%)	(1.71%)
	Sequential	0.0160	0.835	0.352	1.556E-02	9.008E-02	0.0235	0.794	0.353	1.588E-02	9.092E-02
5	Int. framework	0.0188	0.885	**0.430**	**1.871E-02**	**9.687E-02**	0.0295	0.841	**0.431**	**1.907E-02**	**9.791E-02**
				(2.60%)	(3.58%)	(2.25%)			(2.30%)	(4.07%)	(2.30%)
	Sequential	0.0245	0.853	0.419	1.941E-02	9.910E-02	0.0354	0.808	0.421	1.987E-02	1.002E-01
10	Int. framework	0.0333	0.903	**0.494**	**2.699E-02**	**1.146E-01**	0.0518	0.850	**0.494**	**2.759E-02**	**1.163E-01**
				(3.01%)	(4.75%)	(3.34%)			(2.48%)	(5.54%)	(3.47%)
	Sequential	0.0398	0.867	0.479	2.834E-02	1.186E-01	0.0593	0.812	0.482	2.921E-02	1.205E-01
15	Int. framework	0.0429	0.918	**0.515**	**3.437E-02**	**1.295E-01**	0.0700	0.855	**0.515**	**3.518E-02**	**1.317E-01**
				(3.30%)	(5.13%)	(4.24%)			(2.52%)	(6.26%)	(4.31%)
	Sequential	0.0487	0.878	0.499	3.623E-02	1.353E-01	0.0776	0.812	0.503	3.753E-02	1.377E-01
25	Int. framework	0.0556	0.943	**0.522**	**4.834E-02**	**1.537E-01**	0.1023	0.861	**0.521**	**4.941E-02**	**1.565E-01**
				(3.36%)	(4.21%)	(4.91%)			(2.01%)	(6.14%)	(5.15%)
	Sequential	0.0571	0.897	0.505	5.046E-02	1.616E-01	0.1074	0.810	0.510	5.264E-02	1.650E-01

S&P 500

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0104	0.709	**0.224**	**1.128E-02**	**7.607E-02**	0.0115	0.695	**0.224**	**1.139E-02**	**7.643E-02**
				(1.51%)	(1.77%)	(0.88%)			(1.40%)	(2.14%)	(1.07%)
	Sequential	0.0123	0.688	0.220	1.148E-02	7.674E-02	0.0136	0.673	0.221	1.164E-02	7.726E-02
3	Int. framework	0.0224	0.817	**0.402**	**1.659E-02**	**8.783E-02**	0.0251	0.800	**0.403**	**1.677E-02**	**8.839E-02**
				(2.05%)	(3.12%)	(1.82%)			(1.91%)	(3.61%)	(2.08%)
	Sequential	0.0266	0.790	0.394	1.712E-02	8.947E-02	0.0294	0.770	0.396	1.739E-02	9.026E-02
5	Int. framework	0.0310	0.844	**0.476**	**2.060E-02**	**9.624E-02**	0.0349	0.825	**0.478**	**2.082E-02**	**9.695E-02**
				(2.25%)	(4.03%)	(2.51%)			(2.11%)	(4.66%)	(2.89%)
	Sequential	0.0363	0.814	0.466	2.146E-02	9.872E-02	0.0405	0.792	0.468	2.184E-02	9.984E-02
10	Int. framework	0.0489	0.866	**0.544**	**2.989E-02**	**1.153E-01**	0.0558	0.843	**0.546**	**3.020E-02**	**1.163E-01**
				(2.43%)	(5.09%)	(3.59%)			(2.20%)	(5.97%)	(4.02%)
	Sequential	0.0543	0.834	0.531	3.149E-02	1.196E-01	0.0617	0.807	0.535	3.211E-02	1.212E-01
15	Int. framework	0.0671	0.871	**0.564**	**3.877E-02**	**1.322E-01**	0.0770	0.845	**0.567**	**3.911E-02**	**1.334E-01**
				(2.59%)	(5.45%)	(4.25%)			(2.28%)	(6.61%)	(4.72%)
	Sequential	0.0709	0.838	0.550	4.100E-02	1.380E-01	0.0815	0.807	0.554	4.188E-02	1.400E-01
25	Int. framework	0.1041	0.871	**0.568**	**5.590E-02**	**1.613E-01**	0.1200	0.840	**0.573**	**5.612E-02**	**1.624E-01**
				(2.64%)	(4.82%)	(4.74%)			(2.22%)	(6.71%)	(5.35%)
	Sequential	0.1012	0.836	0.554	5.873E-02	1.693E-01	0.1185	0.797	0.561	6.016E-02	1.715E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.8: Out-of-sample fit, FIGARCH(1,d,1) - Deseasonalized by simple averaging

NASDAQ

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0023	0.765	0.189 (1.68%)	1.052E-02 (1.13%)	7.703E-02 (0.58%)	0.0060	0.732	0.188 (1.85%)	1.067E-02 (0.87%)	7.745E-02 (0.35%)
	Sequential	0.0043	0.746	0.186	1.064E-02	7.748E-02	0.0078	0.718	0.185	1.077E-02	7.772E-02
3	Int. framework	0.0046	0.889	0.369 (2.09%)	1.490E-02 (2.27%)	8.834E-02 (1.37%)	0.0134	0.845	0.367 (2.12%)	1.518E-02 (2.00%)	8.905E-02 (0.97%)
	Sequential	0.0085	0.864	0.361	1.525E-02	8.956E-02	0.0169	0.826	0.360	1.549E-02	8.992E-02
5	Int. framework	0.0074	0.919	0.443 (2.30%)	1.824E-02 (3.10%)	9.617E-02 (2.03%)	0.0202	0.869	0.440 (2.35%)	1.863E-02 (2.93%)	9.720E-02 (1.57%)
	Sequential	0.0123	0.890	0.433	1.882E-02	9.816E-02	0.0249	0.847	0.430	1.920E-02	9.875E-02
10	Int. framework	0.0094	0.955	0.512 (2.46%)	2.582E-02 (4.03%)	1.128E-01 (3.20%)	0.0312	0.894	0.508 (2.49%)	2.647E-02 (4.15%)	1.145E-01 (2.67%)
	Sequential	0.0148	0.921	0.500	2.690E-02	1.165E-01	0.0366	0.867	0.496	2.762E-02	1.176E-01
15	Int. framework	0.0069	0.982	0.540 (2.62%)	3.240E-02 (4.27%)	1.263E-01 (4.07%)	0.0367	0.912	0.535 (2.60%)	3.314E-02 (4.75%)	1.285E-01 (3.50%)
	Sequential	0.0118	0.943	0.526	3.385E-02	1.317E-01	0.0414	0.883	0.522	3.479E-02	1.332E-01
25	Int. framework	-0.0012	1.019	0.556 (2.46%)	4.457E-02 (3.28%)	1.484E-01 (4.82%)	0.0437	0.936	0.553 (2.33%)	4.502E-02 (4.57%)	1.507E-01 (4.36%)
	Sequential	0.0001	0.975	0.543	4.609E-02	1.559E-01	0.0441	0.905	0.540	4.717E-02	1.575E-01

S&P 500

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0082	0.721	0.225 (1.50%)	1.125E-02 (1.24%)	7.630E-02 (0.48%)	0.0104	0.699	0.224 (1.28%)	1.142E-02 (0.27%)	7.679E-02 (-0.06%)
	Sequential	0.0104	0.703	0.221	1.139E-02	7.666E-02	0.0116	0.693	0.222	1.145E-02	7.675E-02
3	Int. framework	0.0164	0.838	0.405 (2.04%)	1.643E-02 (2.55%)	8.802E-02 (1.33%)	0.0218	0.807	0.404 (1.73%)	1.677E-02 (1.32%)	8.886E-02 (0.56%)
	Sequential	0.0211	0.813	0.397	1.686E-02	8.921E-02	0.0244	0.798	0.397	1.700E-02	8.936E-02
5	Int. framework	0.0212	0.871	0.482 (2.17%)	2.025E-02 (3.39%)	9.621E-02 (1.96%)	0.0293	0.835	0.479 (1.85%)	2.076E-02 (2.01%)	9.745E-02 (0.94%)
	Sequential	0.0269	0.844	0.471	2.096E-02	9.813E-02	0.0323	0.825	0.471	2.119E-02	9.838E-02
10	Int. framework	0.0278	0.909	0.553 (2.17%)	2.892E-02 (4.21%)	1.145E-01 (3.06%)	0.0419	0.864	0.550 (1.81%)	2.981E-02 (2.72%)	1.163E-01 (1.78%)
	Sequential	0.0333	0.881	0.541	3.019E-02	1.181E-01	0.0435	0.856	0.540	3.064E-02	1.184E-01
15	Int. framework	0.0339	0.927	0.577 (2.26%)	3.692E-02 (4.49%)	1.308E-01 (3.86%)	0.0528	0.877	0.573 (1.80%)	3.805E-02 (2.99%)	1.332E-01 (2.31%)
	Sequential	0.0375	0.897	0.564	3.865E-02	1.361E-01	0.0514	0.870	0.563	3.923E-02	1.363E-01
25	Int. framework	0.0461	0.947	0.589 (2.10%)	5.180E-02 (3.78%)	1.588E-01 (4.14%)	0.0730	0.888	0.587 (1.60%)	5.293E-02 (2.77%)	1.609E-01 (2.84%)
	Sequential	0.0426	0.917	0.577	5.384E-02	1.656E-01	0.0629	0.885	0.577	5.444E-02	1.656E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.9: Out-of-sample fit, HYGARCH(1,d,1) - Deseasonalized by simple averaging

NASDAQ

Horizon	Estimation method	normally-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	t-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0036	0.752	**0.188**	**1.059E-02**	**7.730E-02**	0.0062	0.732	**0.188**	**1.066E-02**	**7.737E-02**
				(1.20%)	(1.34%)	(0.75%)			(1.66%)	(2.13%)	(1.11%)
	Sequential	0.0055	0.733	0.186	1.073E-02	7.789E-02	0.0091	0.702	0.185	1.089E-02	7.824E-02
3	Int. framework	0.0063	0.878	**0.369**	**1.497E-02**	**8.867E-02**	0.0134	0.847	**0.367**	**1.515E-02**	**8.892E-02**
				(1.98%)	(2.65%)	(1.65%)			(1.92%)	(3.41%)	(1.95%)
	Sequential	0.0107	0.849	0.362	1.538E-02	9.016E-02	0.0194	0.809	0.360	1.569E-02	9.069E-02
5	Int. framework	0.0096	0.907	**0.442**	**1.833E-02**	**9.660E-02**	0.0201	0.871	**0.440**	**1.861E-02**	**9.707E-02**
				(2.24%)	(3.62%)	(2.39%)			(2.18%)	(4.55%)	(2.72%)
	Sequential	0.0156	0.873	0.433	1.901E-02	9.897E-02	0.0284	0.827	0.431	1.949E-02	9.978E-02
10	Int. framework	0.0136	0.940	**0.512**	**2.594E-02**	**1.134E-01**	0.0309	0.897	**0.507**	**2.646E-02**	**1.143E-01**
				(2.45%)	(4.82%)	(3.69%)			(2.25%)	(6.12%)	(4.22%)
	Sequential	0.0217	0.897	0.499	2.725E-02	1.177E-01	0.0432	0.842	0.496	2.819E-02	1.194E-01
15	Int. framework	0.0142	0.962	**0.538**	**3.255E-02**	**1.271E-01**	0.0362	0.916	**0.534**	**3.318E-02**	**1.285E-01**
				(2.65%)	(5.33%)	(4.67%)			(2.35%)	(7.03%)	(5.24%)
	Sequential	0.0237	0.912	0.525	3.439E-02	1.333E-01	0.0525	0.851	0.522	3.569E-02	1.356E-01
25	Int. framework	0.0143	0.989	**0.554**	**4.469E-02**	**1.494E-01**	0.0427	0.942	**0.550**	**4.524E-02**	**1.509E-01**
				(2.65%)	(5.02%)	(5.63%)			(2.17%)	(7.44%)	(6.53%)
	Sequential	0.0253	0.929	0.540	4.705E-02	1.583E-01	0.0668	0.859	0.539	4.888E-02	1.615E-01

S&P 500

Horizon	Estimation method	normally-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	t-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0087	0.721	**0.225**	**1.122E-02**	**7.607E-02**	0.0106	0.700	**0.224**	**1.138E-02**	**7.654E-02**
				(1.32%)	(2.16%)	(1.13%)			(1.11%)	(2.66%)	(1.38%)
	Sequential	0.0111	0.695	0.222	1.147E-02	7.694E-02	0.0130	0.672	0.222	1.169E-02	7.761E-02
3	Int. framework	0.0176	0.837	**0.405**	**1.640E-02**	**8.772E-02**	0.0227	0.808	**0.403**	**1.672E-02**	**8.847E-02**
				(1.90%)	(3.51%)	(2.13%)			(1.61%)	(4.10%)	(2.43%)
	Sequential	0.0227	0.802	0.397	1.700E-02	8.963E-02	0.0277	0.771	0.397	1.744E-02	9.067E-02
5	Int. framework	0.0230	0.870	**0.481**	**2.023E-02**	**9.586E-02**	0.0309	0.836	**0.478**	**2.072E-02**	**9.695E-02**
				(2.08%)	(4.42%)	(2.97%)			(1.79%)	(5.17%)	(3.36%)
	Sequential	0.0295	0.831	0.471	2.117E-02	9.879E-02	0.0376	0.795	0.470	2.185E-02	1.003E-01
10	Int. framework	0.0313	0.906	**0.552**	**2.896E-02**	**1.140E-01**	0.0465	0.862	**0.547**	**2.991E-02**	**1.161E-01**
				(2.12%)	(5.33%)	(4.37%)			(1.77%)	(6.56%)	(4.83%)
	Sequential	0.0386	0.863	0.541	3.059E-02	1.193E-01	0.0546	0.815	0.537	3.200E-02	1.220E-01
15	Int. framework	0.0392	0.923	**0.575**	**3.709E-02**	**1.304E-01**	0.0612	0.871	**0.568**	**3.845E-02**	**1.331E-01**
				(2.28%)	(5.72%)	(5.29%)			(1.83%)	(7.47%)	(5.70%)
	Sequential	0.0464	0.874	0.562	3.933E-02	1.377E-01	0.0701	0.818	0.558	4.155E-02	1.412E-01
25	Int. framework	0.0554	0.940	**0.586**	**5.237E-02**	**1.587E-01**	0.0899	0.878	**0.577**	**5.441E-02**	**1.615E-01**
				(2.29%)	(5.19%)	(5.53%)			(1.72%)	(8.23%)	(6.68%)
	Sequential	0.0606	0.883	0.572	5.524E-02	1.680E-01	0.0991	0.813	0.567	5.929E-02	1.731E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.10: Out-of-sample fit, GARCH(1,1)-Deseasonalized by FFF method

NASDAQ

Horizon	Estimation method	normally-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	t-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0104	0.697	**0.180** (1.93%)	**1.090E-02** (2.03%)	**7.802E-02** (1.04%)	0.0139	0.667	**0.181** (1.74%)	**1.110E-02** (2.20%)	**7.853E-02** (1.11%)
	Sequential	0.0131	0.671	0.177	1.113E-02	7.885E-02	0.0162	0.643	0.178	1.135E-02	7.941E-02
3	Int. framework	0.0256	0.792	**0.344** (2.46%)	**1.599E-02** (3.65%)	**9.064E-02** (1.98%)	0.0329	0.755	**0.345** (2.10%)	**1.635E-02** (3.89%)	**9.146E-02** (2.00%)
	Sequential	0.0310	0.759	0.335	1.659E-02	9.247E-02	0.0378	0.724	0.338	1.702E-02	9.332E-02
5	Int. framework	0.0373	0.812	**0.410** (2.91%)	**2.000E-02** (4.94%)	**9.968E-02** (2.76%)	0.0476	0.771	**0.412** (2.39%)	**2.053E-02** (5.26%)	**1.008E-01** (2.70%)
	Sequential	0.0443	0.775	0.398	2.103E-02	1.025E-01	0.0543	0.735	0.402	2.167E-02	1.036E-01
10	Int. framework	0.0586	0.832	**0.468** (3.39%)	**2.918E-02** (6.65%)	**1.186E-01** (4.23%)	0.0761	0.782	**0.471** (2.51%)	**3.013E-02** (7.23%)	**1.205E-01** (4.00%)
	Sequential	0.0668	0.789	0.453	3.126E-02	1.238E-01	0.0847	0.741	0.459	3.248E-02	1.255E-01
15	Int. framework	0.0702	0.853	**0.488** (3.52%)	**3.708E-02** (7.15%)	**1.343E-01** (5.24%)	0.0954	0.795	**0.491** (2.37%)	**3.831E-02** (8.16%)	**1.368E-01** (4.94%)
	Sequential	0.0767	0.808	0.472	3.994E-02	1.418E-01	0.1039	0.749	0.480	4.171E-02	1.439E-01
25	Int. framework	0.0785	0.896	**0.492** (2.93%)	**5.160E-02** (5.96%)	**1.596E-01** (6.04%)	0.1219	0.819	**0.496** (1.37%)	**5.292E-02** (8.17%)	**1.620E-01** (6.19%)
	Sequential	0.0754	0.850	0.478	5.487E-02	1.698E-01	0.1262	0.768	0.489	5.763E-02	1.727E-01

S&P 500

Horizon	Estimation method	normally-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	t-distributed innovations b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0103	0.702	**0.218** (1.81%)	**1.145E-02** (2.16%)	**7.699E-02** (1.06%)	0.0114	0.690	**0.218** (1.63%)	**1.156E-02** (2.52%)	**7.733E-02** (1.22%)
	Sequential	0.0127	0.677	0.214	1.171E-02	7.781E-02	0.0140	0.663	0.215	1.186E-02	7.829E-02
3	Int. framework	0.0267	0.795	**0.392** (2.30%)	**1.706E-02** (3.69%)	**8.911E-02** (2.07%)	0.0289	0.781	**0.393** (2.11%)	**1.723E-02** (4.18%)	**8.962E-02** (2.30%)
	Sequential	0.0313	0.764	0.383	1.772E-02	9.100E-02	0.0340	0.747	0.384	1.798E-02	9.172E-02
5	Int. framework	0.0372	0.820	**0.463** (2.49%)	**2.135E-02** (4.67%)	**9.786E-02** (2.83%)	0.0403	0.805	**0.465** (2.31%)	**2.156E-02** (5.31%)	**9.853E-02** (3.10%)
	Sequential	0.0429	0.788	0.452	2.240E-02	1.007E-01	0.0467	0.768	0.454	2.277E-02	1.017E-01
10	Int. framework	0.0599	0.839	**0.529** (2.66%)	**3.129E-02** (5.74%)	**1.175E-01** (3.95%)	0.0646	0.822	**0.531** (2.41%)	**3.155E-02** (6.70%)	**1.184E-01** (4.33%)
	Sequential	0.0650	0.805	0.515	3.320E-02	1.224E-01	0.0715	0.782	0.519	3.381E-02	1.238E-01
15	Int. framework	0.0820	0.843	**0.548** (2.74%)	**4.076E-02** (5.85%)	**1.348E-01** (4.42%)	0.0879	0.825	**0.551** (2.44%)	**4.093E-02** (7.19%)	**1.357E-01** (4.80%)
	Sequential	0.0839	0.810	0.533	4.329E-02	1.410E-01	0.0930	0.784	0.538	4.410E-02	1.426E-01
25	Int. framework	0.1260	0.842	**0.553** (2.58%)	**5.884E-02** (4.54%)	**1.648E-01** (4.38%)	0.1337	0.823	**0.560** (2.24%)	**5.852E-02** (6.78%)	**1.652E-01** (4.94%)
	Sequential	0.1165	0.812	0.539	6.163E-02	1.724E-01	0.1310	0.780	0.548	6.278E-02	1.738E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.11: Out-of-sample fit, FIGARCH(1,d,1) - Deseasonalized by FFF method

NASDAQ

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0073	0.718	**0.184** (1.80%)	**1.081E-02** (1.63%)	**7.803E-02** (0.78%)	0.0115	0.682	**0.184** (2.05%)	**1.103E-02** (1.10%)	**7.858E-02** (0.42%)
	Sequential	0.0100	0.694	0.181	1.099E-02	7.864E-02	0.0135	0.666	0.181	1.115E-02	7.891E-02
3	Int. framework	0.0178	0.823	**0.354** (2.24%)	**1.561E-02** (3.18%)	**9.009E-02** (1.71%)	0.0273	0.775	**0.354** (2.36%)	**1.606E-02** (2.50%)	**9.111E-02** (1.14%)
	Sequential	0.0232	0.791	0.347	1.612E-02	9.166E-02	0.0313	0.755	0.346	1.647E-02	9.216E-02
5	Int. framework	0.0250	0.850	**0.424** (2.49%)	**1.930E-02** (4.35%)	**9.853E-02** (2.53%)	0.0387	0.796	**0.423** (2.64%)	**1.998E-02** (3.66%)	**9.997E-02** (1.80%)
	Sequential	0.0320	0.813	0.414	2.017E-02	1.011E-01	0.0441	0.772	0.412	2.073E-02	1.018E-01
10	Int. framework	0.0337	0.887	**0.489** (2.63%)	**2.754E-02** (5.77%)	**1.162E-01** (3.98%)	0.0571	0.819	**0.487** (2.86%)	**2.874E-02** (5.35%)	**1.186E-01** (3.09%)
	Sequential	0.0427	0.842	0.476	2.923E-02	1.210E-01	0.0638	0.791	0.474	3.036E-02	1.224E-01
15	Int. framework	0.0317	0.923	**0.516** (2.67%)	**3.442E-02** (6.18%)	**1.306E-01** (4.96%)	0.0645	0.844	**0.514** (2.96%)	**3.582E-02** (6.26%)	**1.337E-01** (4.02%)
	Sequential	0.0413	0.871	0.502	3.669E-02	1.374E-01	0.0708	0.813	0.500	3.821E-02	1.393E-01
25	Int. framework	0.0128	0.985	**0.533** (2.18%)	**4.674E-02** (5.16%)	**1.535E-01** (5.83%)	0.0652	0.887	**0.534** (2.63%)	**4.777E-02** (6.62%)	**1.563E-01** (5.28%)
	Sequential	0.0202	0.926	0.522	4.928E-02	1.630E-01	0.0671	0.854	0.520	5.116E-02	1.650E-01

S&P 500

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0079	0.715	**0.220** (1.75%)	**1.142E-02** (1.47%)	**7.723E-02** (0.54%)	0.0115	0.681	**0.219** (1.51%)	**1.170E-02** (0.71%)	**7.798E-02** (0.18%)
	Sequential	0.0105	0.694	0.216	1.159E-02	7.766E-02	0.0130	0.671	0.215	1.178E-02	7.812E-02
3	Int. framework	0.0205	0.816	**0.395** (2.16%)	**1.688E-02** (2.87%)	**8.927E-02** (1.42%)	0.0286	0.772	**0.393** (1.90%)	**1.746E-02** (1.86%)	**9.049E-02** (0.97%)
	Sequential	0.0255	0.789	0.387	1.738E-02	9.055E-02	0.0316	0.758	0.386	1.779E-02	9.138E-02
5	Int. framework	0.0267	0.849	**0.469** (2.26%)	**2.094E-02** (3.71%)	**9.788E-02** (2.07%)	0.0391	0.797	**0.466** (2.04%)	**2.184E-02** (2.65%)	**9.957E-02** (1.45%)
	Sequential	0.0330	0.820	0.459	2.175E-02	9.995E-02	0.0427	0.783	0.456	2.243E-02	1.010E-01
10	Int. framework	0.0368	0.885	**0.539** (2.25%)	**3.008E-02** (4.58%)	**1.167E-01** (3.13%)	0.0588	0.820	**0.533** (2.05%)	**3.178E-02** (3.59%)	**1.194E-01** (2.49%)
	Sequential	0.0431	0.854	0.527	3.152E-02	1.205E-01	0.0620	0.805	0.522	3.296E-02	1.224E-01
15	Int. framework	0.0446	0.904	**0.563** (2.31%)	**3.837E-02** (4.83%)	**1.332E-01** (3.84%)	0.0757	0.828	**0.555** (2.12%)	**4.082E-02** (4.08%)	**1.367E-01** (3.11%)
	Sequential	0.0492	0.872	0.550	4.032E-02	1.385E-01	0.0771	0.814	0.543	4.255E-02	1.411E-01
25	Int. framework	0.0591	0.926	**0.577** (2.07%)	**5.348E-02** (3.97%)	**1.611E-01** (4.06%)	0.1064	0.835	**0.568** (2.09%)	**5.694E-02** (4.33%)	**1.652E-01** (3.86%)
	Sequential	0.0562	0.895	0.565	5.569E-02	1.679E-01	0.1018	0.821	0.556	5.952E-02	1.718E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.12: Out-of-sample fit, HYGARCH(1,d,1) - Deseasonalized by FFF method

NASDAQ

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0087	0.707	0.184 (1.15%)	1.087E-02 (1.95%)	7.819E-02 (1.10%)	0.0116	0.683	0.184 (1.59%)	1.100E-02 (2.38%)	7.842E-02 (1.21%)
	Sequential	0.0108	0.684	0.182	1.109E-02	7.906E-02	0.0143	0.656	0.181	1.127E-02	7.938E-02
3	Int. framework	0.0197	0.814	0.355 (2.05%)	1.567E-02 (3.72%)	9.029E-02 (2.18%)	0.0273	0.778	0.353 (1.88%)	1.600E-02 (4.01%)	9.088E-02 (2.14%)
	Sequential	0.0246	0.779	0.348	1.628E-02	9.230E-02	0.0327	0.743	0.347	1.667E-02	9.287E-02
5	Int. framework	0.0273	0.841	0.425 (2.43%)	1.936E-02 (5.10%)	9.881E-02 (3.06%)	0.0387	0.799	0.422 (2.15%)	1.991E-02 (5.42%)	9.971E-02 (2.99%)
	Sequential	0.0342	0.800	0.415	2.041E-02	1.019E-01	0.0462	0.758	0.413	2.106E-02	1.028E-01
10	Int. framework	0.0380	0.875	0.490 (2.63%)	2.762E-02 (6.90%)	1.165E-01 (4.75%)	0.0571	0.823	0.486 (2.17%)	2.867E-02 (7.53%)	1.183E-01 (4.61%)
	Sequential	0.0476	0.824	0.478	2.967E-02	1.223E-01	0.0680	0.773	0.475	3.101E-02	1.241E-01
15	Int. framework	0.0393	0.905	0.516 (2.68%)	3.454E-02 (7.61%)	1.309E-01 (5.97%)	0.0645	0.848	0.512 (2.16%)	3.580E-02 (8.77%)	1.334E-01 (5.82%)
	Sequential	0.0505	0.846	0.503	3.738E-02	1.392E-01	0.0783	0.790	0.501	3.924E-02	1.416E-01
25	Int. framework	0.0305	0.956	0.533 (2.38%)	4.688E-02 (7.33%)	1.538E-01 (7.11%)	0.0648	0.893	0.529 (1.81%)	4.797E-02 (9.70%)	1.563E-01 (7.48%)
	Sequential	0.0427	0.885	0.520	5.058E-02	1.656E-01	0.0846	0.818	0.520	5.312E-02	1.689E-01

S&P 500

		normally-distributed innovations					t-distributed innovations				
Horizon	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
1	Int. framework	0.0085	0.715	0.219 (1.51%)	1.139E-02 (2.43%)	7.695E-02 (1.24%)	0.0109	0.691	0.219 (1.55%)	1.157E-02 (2.86%)	7.749E-02 (1.40%)
	Sequential	0.0113	0.685	0.216	1.167E-02	7.791E-02	0.0136	0.661	0.216	1.191E-02	7.859E-02
3	Int. framework	0.0218	0.815	0.395 (1.98%)	1.685E-02 (3.90%)	8.891E-02 (2.28%)	0.0273	0.784	0.394 (1.91%)	1.721E-02 (4.50%)	8.976E-02 (2.49%)
	Sequential	0.0273	0.778	0.387	1.753E-02	9.098E-02	0.0328	0.747	0.387	1.802E-02	9.205E-02
5	Int. framework	0.0289	0.847	0.468 (2.13%)	2.092E-02 (4.81%)	9.750E-02 (3.08%)	0.0373	0.811	0.467 (2.05%)	2.147E-02 (5.63%)	9.861E-02 (3.32%)
	Sequential	0.0357	0.807	0.459	2.198E-02	1.006E-01	0.0444	0.770	0.457	2.275E-02	1.020E-01
10	Int. framework	0.0412	0.881	0.537 (2.07%)	3.017E-02 (5.68%)	1.163E-01 (4.39%)	0.0566	0.835	0.535 (2.08%)	3.115E-02 (7.21%)	1.182E-01 (4.70%)
	Sequential	0.0486	0.837	0.526	3.198E-02	1.217E-01	0.0655	0.789	0.524	3.357E-02	1.240E-01
15	Int. framework	0.0519	0.898	0.560 (2.09%)	3.869E-02 (5.84%)	1.330E-01 (5.13%)	0.0733	0.846	0.557 (2.10%)	3.999E-02 (8.08%)	1.354E-01 (5.52%)
	Sequential	0.0584	0.850	0.548	4.109E-02	1.401E-01	0.0827	0.794	0.545	4.350E-02	1.433E-01
25	Int. framework	0.0728	0.915	0.571 (1.81%)	5.455E-02 (4.81%)	1.616E-01 (5.17%)	0.1043	0.857	0.568 (1.76%)	5.612E-02 (8.44%)	1.640E-01 (6.05%)
	Sequential	0.0750	0.862	0.561	5.730E-02	1.705E-01	0.1126	0.795	0.558	6.129E-02	1.745E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.13: Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by simple averaging

NASDAQ

		normally-distributed innovations				t-distributed innovations			
		DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	Horizon	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	1	10.644	0.000	12.395	0.000	11.809	0.000	13.490	0.000
	3	8.767	0.000	11.905	0.000	9.293	0.000	12.124	0.000
	5	8.407	0.000	11.702	0.000	8,749	0.000	11.427	0.000
	10	7.445	0.000	11.057	0.000	7.719	0.000	10.821	0.000
	15	6.561	0.000	10.661	0.000	6.978	0.000	10.105	0.000
	25	3.876	0.000	7.967	0.000	4.789	0.000	7.761	0.000
FIGARCH(1,d,1)	1	7.691	0.000	8.453	0.000	6.464	0.000	5.412	0.000
	3	8.025	0.000	10.605	0.000	7.094	0.000	7.796	0.000
	5	8.255	0.000	11.790	0.000	7.619	0.000	9.334	0.000
	10	7.743	0.000	12.571	0.000	7.747	0.000	11.116	0.000
	15	6.765	0.000	11.911	0.000	7.627	0.000	11.496	0.000
	25	3.542	0.000	8.784	0.000	5.671	0.000	9.699	0.000
HYGARCH(1,d,1)	1	8.703	0.000	10.462	0.000	14.381	0.000	15.954	0.000
	3	9.185	0.000	12.759	0.000	10.620	0.000	14.160	0.000
	5	9.108	0.000	13.654	0.000	9.985	0.000	14.176	0.000
	10	8.136	0.000	13.486	0.000	8.619	0.000	13.598	0.000
	15	6.978	0.000	12.344	0.000	7.583	0.000	12.174	0.000
	25	4.224	0.000	9.048	0.000	5.136	0.000	9.215	0.000

S&P 500

		normally-distributed innovations				t-distributed innovations			
		DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	Horizon	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	1	9.690	0.000	11.764	0.000	10.691	0.000	13.271	0.000
	3	8.324	0.000	11.905	0.000	8.679	0.000	12.486	0.000
	5	7.710	0.000	11.616	0.000	7.965	0.000	12.263	0.000
	10	6.540	0.000	10.673	0.000	6.726	0.000	10.686	0.000
	15	5.654	0.000	9.486	0.000	5.894	0.000	9.233	0.000
	25	3.590	0.000	6.893	0.000	4.152	0.000	6.675	0.000
FIGARCH(1,d,1)	1	6.908	0.000	6.482	0.000	1.866	0.031	-0.851	0.803
	3	7.258	0.000	9.187	0.000	4.698	0.000	4.279	0.000
	5	7.281	0.000	10.039	0.000	5.357	0.000	5.316	0.000
	10	6.973	0.000	11.090	0.000	5.385	0.000	7.218	0.000
	15	6.478	0.000	11.093	0.000	5.395	0.000	7.990	0.000
	25	4.087	0.000	7.536	0.000	4.178	0.000	7.116	0.000
HYGARCH(1,d,1)	1	10.835	0.000	14.484	0.000	12.009	0.000	16.329	0.000
	3	8.775	0.000	13.594	0.000	8.925	0.000	13.920	0.000
	5	8.139	0.000	13.625	0.000	8.100	0.000	13.562	0.000
	10	6.791	0.000	12.868	0.000	6.855	0.000	11.911	0.000
	15	5.663	0.000	11.224	0.000	5.970	0.000	10.062	0.000
	25	3.369	0.000	7.169	0.000	4.304	0.000	7.378	0.000

Note: The DM-SE and DM-AE are the test statistics for squared error and absolute error respectively. The alternative hypothesis indicates that the integrated framework outperforms the sequential method.

Table 3.14: Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by FFF method

NASDAQ

		normally-distributed innovations				t-distributed innovations			
		DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	Horizon	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	1	11.167	0.000	13.290	0.000	12.537	0.000	14.683	0.000
	3	8.763	0.000	12.553	0.000	9.621	0.000	12.908	0.000
	5	8.173	0.000	12.392	0.000	8.890	0.000	12.290	0.000
	10	7.279	0.000	12.206	0.000	7.884	0.000	11.575	0.000
	15	6.585	0.000	11.512	0.000	7.281	0.000	10.812	0.000
	25	4.309	0.000	8.528	0.000	5.468	0.000	8.734	0.000
FIGARCH(1,d,1)	1	8.879	0.000	9.955	0.000	7.292	0.000	6.242	0.000
	3	8.603	0.000	11.711	0.000	7.748	0.000	8.734	0.000
	5	8.531	0.000	12.699	0.000	8.158	0.000	10.167	0.000
	10	7.850	0.000	13.312	0.000	8.286	0.000	12.277	0.000
	15	7.023	0.000	12.444	0.000	8.336	0.000	12.779	0.000
	25	4.379	0.000	9.113	0.000	7.298	0.000	11.548	0.000
HYGARCH(1,d,1)	1	10.472	0.000	13.445	0.000	14.584	0.000	16.719	0.000
	3	10.236	0.000	14.842	0.000	11.125	0.000	14.965	0.000
	5	9.839	0.000	15.096	0.000	10.434	0.000	14.899	0.000
	10	8.769	0.000	14.741	0.000	9.355	0.000	14.594	0.000
	15	7.690	0.000	13.416	0.000	8.660	0.000	13.509	0.000
	25	5.098	0.000	9.722	0.000	6.638	0.000	10.741	0.000

S&P 500

		normally-distributed innovations				t-distributed innovations			
		DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	Horizon	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	1	11.146	0.000	13.178	0.000	11.702	0.000	14.340	0.000
	3	9.052	0.000	12.654	0.000	9.073	0.000	13.185	0.000
	5	8.090	0.000	12.305	0.000	8.125	0.000	12.525	0.000
	10	6.637	0.000	11.000	0.000	6.723	0.000	11.046	0.000
	15	5.575	0.000	9.249	0.000	5.800	0.000	9.071	0.000
	25	3.225	0.001	5.928	0.000	3.940	0.000	5.944	0.000
FIGARCH(1,d,1)	1	7.855	0.000	7.128	0.000	4.532	0.000	2.471	0.007
	3	7.646	0.000	9.680	0.000	5.865	0.000	6.826	0.000
	5	7.453	0.000	10.420	0.000	6.126	0.000	7.476	0.000
	10	6.961	0.000	10.936	0.000	5.843	0.000	9.013	0.000
	15	6.326	0.000	10.666	0.000	5.609	0.000	9.146	0.000
	25	3.992	0.000	7.158	0.000	4.781	0.000	7.891	0.000
HYGARCH(1,d,1)	1	11.641	0.000	15.148	0.000	12.688	0.000	16.182	0.000
	3	9.088	0.000	14.186	0.000	9.509	0.000	14.141	0.000
	5	8.258	0.000	13.875	0.000	8.608	0.000	13.359	0.000
	10	6.819	0.000	12.693	0.000	7.467	0.000	11.781	0.000
	15	5.565	0.000	10.869	0.000	6.516	0.000	10.061	0.000
	25	3.112	0.001	6.710	0.000	4.577	0.000	6.893	0.000

Note: The DM-SE and DM-AE are the test statistics for squared error and absolute error respectively. The alternative hypothesis indicates that the integrated framework outperforms the sequential method.

3.6.3 Out-of-sample 1-day-ahead forecasting performance

Apart from investigating the results on various intraday horizons, the performance of 1-day-ahead forecasts is studied to reflect the capability of the framework. To eliminate the influence from forecasts which cover an interday period, the findings will be based on consecutive forecasts which cover the horizons starting from the first intraday period to the last period on the same day. The 1-day-ahead forecasts are calculated with the use of intraday information up to the last period on the most recent day. (i.e. the $t+1$ 1-day-ahead forecast is calculated with the information up to the last intraday period on day t. The results are tabulated in Table 3.15 to Table 3.18.

Our framework produces better results under all situations for both data sets. The averaged improvements of R^2 is 2.41%, MSE is 8.27% and MAE is 6.23% for the NASDAQ data while the averaged improvements of R^2 is 2.80%, MSE is 8.96% and MAE is 6.74% for the S&P 500 data. Almost all the results of Diebold-Mariano test indicate that the proposed framework reduces squared errors and absolute errors significantly at a 5% significane level.

Overall, the empirical analysis demonstrated that the proposed framework produces better in-sample and out-of-sample R^2, MSE and MAE in most situations. Besides, the Diebold-Mariano hypothesis tests always significantly accept (with p-value<0.05) the alternative hypothesis that the integrated framework outperforms the sequential estimation method by reducing squared errors and absolute errors. The superiority of the proposed framework is demonstrated.

Table 3.15: Out-of-sample fit, 1-step-ahead daily forecast - Deseasonalized by simple averaging

NASDAQ

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	0.1540	0.854	**0.469**	**7.683E-02**	**1.955E-01**	0.2205	0.766	**0.462**	**8.134E-02**	**2.051E-01**
				(2.93%)	(3.90%)	(4.29%)			(1.10%)	(7.02%)	(4.07%)
	Sequential	0.1534	0.805	0.456	7.996E-02	2.043E-01	0.2238	0.712	0.457	8.748E-02	2.138E-01
FIGARCH(1,d,1)	Int. framework	0.0596	0.956	**0.533**	**6.573E-02**	**1.811E-01**	0.1165	0.874	**0.529**	**6.719E-02**	**1.868E-01**
				(2.54%)	(3.36%)	(5.23%)			(3.17%)	(5.92%)	(4.54%)
	Sequential	0.0598	0.910	0.520	6.801E-02	1.911E-01	0.1196	0.836	0.513	7.141E-02	1.957E-01
HYGARCH(1,d,1)	Int. framework	0.0767	0.928	**0.534**	**6.552E-02**	**1.823E-01**	0.1157	0.880	**0.525**	**6.771E-02**	**1.875E-01**
				(2.97%)	(6.10%)	(6.27%)			(2.49%)	(9.82%)	(6.99%)
	Sequential	0.0911	0.864	0.519	6.978E-02	1.945E-01	0.1497	0.788	0.512	7.508E-02	2.016E-01

S&P 500

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	0.2130	0.777	**0.520**	**9.231E-02**	**2.083E-01**	0.2395	0.742	**0.521**	**9.547E-02**	**2.144E-01**
				(3.59%)	(6.98%)	(6.19%)			(2.80%)	(9.55%)	(6.39%)
	Sequential	0.2090	0.737	0.502	9.924E-02	2.220E-01	0.2366	0.694	0.507	1.056E-01	2.290E-01
FIGARCH(1,d,1)	Int. framework	0.1275	0.867	**0.557**	**8.016E-02**	**1.960E-01**	0.1641	0.806	**0.551**	**[illegible]440E-02**	**2.036E-01**
				(2.88%)	(5.52%)	(5.37%)			(2.75%)	(4.58%)	(3.54%)
	Sequential	0.1237	0.833	0.542	8.485E-02	2.071E-01	0.1564	0.794	0.536	8.846E-02	2.111E-01
HYGARCH(1,d,1)	Int. framework	0.1359	0.863	**0.557**	**8.060E-02**	**1.950E-01**	0.1928	0.790	**0.531**	**8.903E-02**	**2.086E-01**
				(2.56%)	(7.26%)	(7.67%)			(1.64%)	(11.35%)	(8.42%)
	Sequential	0.1415	0.806	0.543	8.691E-02	2.112E-01	0.2032	0.721	0.522	1.004E-01	2.278E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.16: Out-of-sample fit, 1-step-ahead daily forecast - Deseasonalized by FFF method

NASDAQ

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1)	Int. framework	0.1774	0.801	**0.442**	**8.152E-02**	**2.033E-01**	0.2419	0.716	**0.436**	**8.812E-02**	**2.145E-01**
				(2.50%)	(8.82%)	(7.02%)			(0.02%)	(11.85%)	(6.38%)
	Sequential	0.1790	0.746	0.431	8.941E-02	2.186E-01	0.2462	0.661	0.436	9.996E-02	2.292E-01
FIGARCH(1,d,1)	Int. framework	0.0755	0.914	**0.509**	**6.900E-02**	**1.892E-01**	0.1429	0.815	**0.508**	**7.263E-02**	**1.967E-01**
				(2.39%)	(7.71%)	(6.96%)			(3.56%)	(9.64%)	(6.14%)
	Sequential	0.0880	0.849	0.497	7.477E-02	2.033E-01	0.1505	0.774	0.490	8.038E-02	2.095E-01
HYGARCH(1,d,1)	Int. framework	0.0911	0.892	**0.514**	**6.870E-02**	**1.893E-01**	0.1429	0.821	**0.502**	**7.289E-02**	**1.970E-01**
				(3.13%)	(11.10%)	(8.46%)			(2.12%)	(14.02%)	(8.43%)
	Sequential	0.1134	0.812	0.499	7.728E-02	2.068E-01	0.1723	0.739	0.492	8.478E-02	2.152E-01

S&P 500

		normally-distributed innovations					t-distributed innovations				
ARCH structure	Estimation method	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)	b0	b1	R2 (improved %)	MSE (improved %)	MAE (improved %)
GARCH(1,1,)	Int. framework	0.2450	0.727	**0.510**	**9.858E-02**	**2.158E-01**	0.2623	0.703	**0.511**	**1.017E-01**	**2.213E-01**
				(3.80%)	(9.72%)	(6.95%)			(2.97%)	(12.76%)	(7.28%)
	Sequential	0.2380	0.688	0.492	1.092E-01	2.320E-01	0.2608	0.654	0.497	1.165E-01	2.386E-01
FIGARCH(1,d,1)	Int. framework	0.1584	0.816	**0.545**	**8.465E-02**	**2.022E-01**	0.2197	0.724	**0.530**	**9.673E-02**	**2.184E-01**
				(2.81%)	(7.54%)	(6.24%)			(2.82%)	(7.01%)	(4.88%)
	Sequential	0.1568	0.781	0.530	9.155E-02	2.156E-01	0.2150	0.706	0.515	1.040E-01	2.296E-01
HYGARCH(1,d,1)	Int. framework	0.1663	0.814	**0.545**	**8.476E-02**	**2.010E-01**	0.2182	0.745	**0.527**	**9.379E-02**	**2.154E-01**
				(2.54%)	(10.12%)	(8.68%)			(2.43%)	(15.13%)	(9.31%)
	Sequential	0.1728	0.757	0.532	9.430E-02	2.200E-01	0.2305	0.679	0.515	1.105E-01	2.375E-01

Note: The improved % represents the improvement obtained from the integrated framework when its result is compared to the sequential esimation method.

Table 3.17: Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by simple averaging

NASDAQ

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	1.640	0.051	3.186	0.001	2.152	0.016	2.465	0.007
FIGARCH(1,d,1)	1.451	0.074	4.601	0.000	2.229	0.013	3.808	0.000
HYGARCH(1,d,1)	2.166	0.015	4.453	0.000	2.471	0.007	3.932	0.000

S&P 500

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	2.463	0.007	5.772	0.000	2.597	0.005	3.966	0.000
FIGARCH(1,d,1)	2.082	0.019	5.210	0.000	1.779	0.038	3.870	0.000
HYGARCH(1,d,1)	2.221	0.013	4.476	0.000	2.456	0.007	3.940	0.000

Note: The DM-SE and DM-AE are the test statistics for squared error and absolute error respectively. The alternative hypothesis indicates that the integrated framework outperforms the sequential method.

Table 3.18: Diebold-Mariano test, Out-of-sample fit for various ARCH structure, 1-step-ahead daily forecast - Deseasonalized by the FFF method

NASDAQ

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	2.314	0.011	4.480	0.000	2.654	0.004	3.514	0.000
FIGARCH(1,d,1)	2.046	0.021	5.104	0.000	2.864	0.002	5.149	0.000
HYGARCH(1,d,1)	2.540	0.006	5.128	0.000	2.863	0.002	4.812	0.000

S&P 500

	normally-distributed innovations				t-distributed innovations			
	DM-SE		DM-AE		DM-SE		DM-AE	
ARCH structure	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value	t-statistics	p-value
GARCH(1,1)	2.698	0.004	5.398	0.000	2.754	0.003	4.041	0.000
FIGARCH(1,d,1)	2.344	0.010	5.305	0.000	2.684	0.004	5.496	0.000
HYGARCH(1,d,1)	2.476	0.007	4.564	0.000	2.559	0.005	3.655	0.000

Note: The DM-SE and DM-AE are the test statistics for squared error and absolute error respectively. The alternative hypothesis indicates that the integrated framework outperforms the sequential method.

3.7 Concluding remarks

The limitations of Andersen & Bollerslev's sequential estimation approach for modeling an intraday volatility process is addressed in this paper. To search for better ARCH parameters, an integrated framework approach that utilizes the interaction effect between the periodicity and the heteroskedasticity is proposed. The proposed framework improves the subsequent ARCH structure in the sequential method by integrating the filtration process and the ARCH process in a united setting and optimizing the model parameters for the raw series instead of the filtered series. The simultaneous presence of the estimated periodicity and the ARCH volatility enables the consideration of their interaction while the use of the raw series assures the optimal fit to the target series. The integrated framework can be written as a modified P-ARCH structure where the periodically varying autoregressive coefficients are represented as the product of the estimated periodicity times the ARCH parameters. On the other hand, the effectiveness of applying the integrated framework to different ARCH structures is also discussed in this paper.

The performance of the proposed approach is tested empirically under two filtration approaches with three ARCH structure, including GARCH(1,1), FIGARCH(1,d,1) and HYGARCH(1,d,1) structures by using 10-minute returns of the NASDAQ index (3-year period) and the S&P 500 index (2.5-year period). Moreover, both normally distributed and t-distributed innovations are considered in our investigation. The performance measures include regression R^2, mean squared error, mean absolute error and Diebold-Mariano hypothesis tests on squared error and absolute error. Apart from studying the 1-step-ahead out-of-sample performance, several multiple-step-ahead forecasting results (up to a one-day-ahead forecast) are also addressed. Under the same level of model flexibility (parameterized portions), the integrated framework approach is demonstrated to achieve better performances in both in-sample fitness and out-of-sample forecasts for most cases. The improvements become more substantial when the volatility models are applied on 1-day-ahead forecasting. The proposed method is demonstrated to give 3.36% - 15.13% reduction on squared errors and 3.54% - 9.31% reduction on absolute errors under different situations.

Overall, apart from considering the integrated framework as an alternative method for intraday volatility modeling, its superior modeling accuracy can help practitioners to better estimate the intrinsic value of an important financial instrument, the variance swap, in the hedge fund industry.

3.8 Supplementary note

3.8.1 Limitation of the integration framework approach for EGARCH(1,1) structure

This supplementary note illustrates the limitation of applying the integrated framework approach to a specific type of ARCH model, where the log of the conditional variances are modeled through ARMA structure (e.g. EGARCH [61] and FIEGARCH [22]). The following content is based on using the EGARCH(1,1) structure specified below for modeling the filtered series:

$$\begin{aligned} log(\tilde{\sigma}^2_{t,n}) &= \omega + \alpha \cdot g(\tilde{\epsilon}_{t,n-1}) + \beta \cdot log(\tilde{\sigma}^2_{t,n-1}) \\ g(\tilde{\epsilon}_{t,n}) &= \theta \cdot \tilde{\epsilon}_{t,n} + \gamma(|\tilde{\epsilon}_{t,n}| - E(|\tilde{\epsilon}_{t,n}|)) \\ \tilde{\epsilon}_{t,n} &\sim D(0, \tilde{\sigma}^2_{t,n}) \end{aligned} \tag{3.25}$$

where $\tilde{\epsilon}_{t,n}$ is a i.i.d. zero-mean symmetrically distributed innovation for the time index (t, n). It represents the innovation of the filtered series for an intraday volatility modeling process. The variance of $\tilde{\epsilon}_{t,n}$ is $\tilde{\sigma}^2_{t,n}$. ω, α, β, θ and γ are the model parameters. Furthermore, $g(\tilde{\epsilon}_{t,n})$ is a zero-mean i.i.d. variable under the above setting. By letting $v_{t,n} \equiv g(\tilde{\epsilon}_{t,n})$, the EGARCH(1,1) can be simplified as:

$$log(\tilde{\sigma}^2_{t,n}) = \omega + \alpha \cdot v_{t,n-1} + \beta \cdot log(\tilde{\sigma}^2_{t,n-1}) \tag{3.26}$$

On the other hand, a typical Periodic EGARCH(1,1) structure [21], which is capable of capturing a repetitive periodical autocorrelation pattern, exists with the following structure when it is applied to model an intraday innovation series.

$$
\begin{aligned}
log(\sigma_{t,n}^2) &= \omega_{s(t,n)} + \alpha_{s(t,n)} \cdot g(\epsilon_{t,n-1}) + \beta_{s(t,n)} \cdot log(\sigma_{t,n-1}^2) \\
&\equiv \omega_{s(t,n)} + \alpha_{s(t,n)} \cdot v_{t,n-1} + \beta_{s(t,n)} \cdot log(\sigma_{t,n-1}^2) \\
\epsilon_{t,n} &\sim D(0, \sigma_{t,n}^2)
\end{aligned}
\tag{3.27}
$$

where $\omega_{s(t,n)}$, $\alpha_{s(t,n)}$ and $\beta_{s(t,n)}$ are periodically changing coefficients. $\epsilon_{t,n}$ denotes an i.i.d. zero-mean innovation of the raw intraday return series. $g(\tilde{\epsilon}_{t,n})$ is the same function as specified in the equation set 3.25.

For the case that the proposed integrated framework approach is applied to the EGARCH(1,1) structure, the formulation of the volatility process of the intraday innovation series will be:

$$
\begin{aligned}
\tilde{\epsilon}_{t,n} &= \frac{\epsilon_{t,n}}{s_{t,n}} - \kappa \\
log(\tilde{\sigma}_{t,n}^2) &= \omega + \alpha \cdot g(\tilde{\epsilon}_{t,n-1}) + \beta \cdot log(\tilde{\sigma}_{t,n-1}^2) \\
\epsilon_{t,n} &\sim D(0, s_{t,n}^2 \tilde{\sigma}_{t,n}^2)
\end{aligned}
\tag{3.28}
$$

Letting $\sigma_{t,n}^2 \equiv s_{t,n}^2 \tilde{\sigma}_{t,n}^2$, the above variance equation can be re-written as:

$$
\begin{aligned}
log(\tilde{\sigma}_{t,n}^2) &= \omega + \alpha \cdot g(\tilde{\epsilon}_{t,n-1}) + \beta \cdot log(\tilde{\sigma}_{t,n-1}^2) \\
log(\tilde{\sigma}_{t,n}^2) + log(s_{t,n}^2) &= \omega + log(s_{t,n}^2) + \alpha \cdot g(\tilde{\epsilon}_{t,n-1}) + \beta \cdot log(\tilde{\sigma}_{t,n-1}^2) \\
&\quad + \beta \cdot log(s_{t,n-1}^2) - \beta \cdot log(s_{t,n-1}^2) \\
log(s_{t,n}^2 \tilde{\sigma}_{t,n}^2) &= \omega + log\left(\frac{s_{t,n}^2}{[s_{t,n-1}]^{2\beta}}\right) + \alpha \cdot g(\tilde{\epsilon}_{t,n-1})
\end{aligned}
$$

$$
\begin{aligned}
& + \beta \cdot log(s^2_{t,n-1}\tilde{\sigma}^2_{t,n-1}) \\
log(\sigma^2_{t,n}) &= \omega + log\left(\frac{s^2_{t,n}}{[s_{t,n-1}]^{2\beta}}\right) + \beta \cdot log(\sigma^2_{t,n-1}) \\
&\quad + \alpha \cdot \frac{s_{t,n-1}}{s_{t,n-1}} \cdot g(\tilde{\epsilon}_{t,n-1}) \\
&= \omega + log\left(\frac{s^2_{t,n}}{[s_{t,n-1}]^{2\beta}}\right) + \beta \cdot log(\sigma^2_{t,n-1}) \\
&\quad + \frac{\alpha}{s_{t,n-1}} \cdot g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}) \\
&= \omega + log\left(\frac{s^2_{t,n}}{[s_{t,n-1}]^{2\beta}}\right) + \beta \cdot log(\sigma^2_{t,n-1}) \\
&\quad + \frac{\alpha}{s_{t,n-1}} \cdot [g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}) - E(g(s_{t,n-1}\tilde{\epsilon}_{t,n-1})) \\
&\quad + E(g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}))] \\
&= \omega + log\left(\frac{s^2_{t,n}}{[s_{t,n-1}]^{2\beta}}\right) + \beta \cdot log(\sigma^2_{t,n-1}) \\
&\quad + \frac{\alpha}{s_{t,n-1}} \cdot [g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}) - E(g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}))] \\
&\quad + \frac{\alpha}{s_{t,n-1}} \cdot E(g(s_{t,n-1}\tilde{\epsilon}_{t,n-1})) \\
&= \left\{\omega + log\left(\frac{s^2_{t,n}}{[s_{t,n-1}]^{2\beta}}\right) + \frac{\alpha}{s_{t,n-1}} \cdot E(g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}))\right\} \\
&\quad + \frac{\alpha}{s_{t,n-1}} \cdot v'_{t,n-1} + \beta \cdot log(\sigma^2_{t,n-1})
\end{aligned}
\tag{3.29}
$$

where $s_{t,n} \cdot g(\tilde{\epsilon}_{t,n}) = g(s_{t,n}\tilde{\epsilon}_{t,n})$ because of $s_{t,n} > 0$. Since $\epsilon_{t,n}$ is an i.i.d. variable, $v'_{t,n}$ will be a zero-mean i.i.d. variable when $v'_{t,n} \equiv g(s_{t,n}\tilde{\epsilon}_{t,n}) - E(g(s_{t,n}\tilde{\epsilon}_{t,n}))$.

The above equation can be re-written as:

$$log(\sigma_{t,n}^2) \;=\; \omega'_{s(t,n)} + \alpha'_{s(t,n)} \cdot v'_{t,n-1} + \beta \cdot log(\sigma_{t,n-1}^2) \qquad (3.30)$$

where $\omega'_{s(t,n)}$ and $\alpha'_{s(t,n)}$ are periodically changing coefficients. The notation $\omega'_{s(t,n)}$ is equivalent to $\left\{\omega + log\left(\frac{s_{t,n}^2}{[s_{t,n-1}]^{2\beta}}\right) + \frac{\alpha}{s_{t,n-1}} \cdot E(g(s_{t,n-1}\tilde{\epsilon}_{t,n-1}))\right\}$ and $\alpha'_{s(t,n)} \equiv \alpha / s_{t,n-1}$.

The coefficient of the term $log(\sigma_{t,n-1}^2)$ in the above equation is constrained to be the same along different time indexes. Comparing with the periodic EGARCH structure in the equation set 3.27, the capability of modeling the repetitive periodical autocorrelation pattern is much restricted and therefore, the performance gain by using our proposed method may be limited. Our preliminary empirical investigation does not indicate a significant performance gain from our proposed approach over Andersen & Bollerslev's sequentially estimation method.

3.8.2 Proposed integrated framework approach with the use of GARCH(1,1) structure - Maximum likelihood estimation method

The sequential approach proposed by Andersen & Bollerslev does not consider the interaction effect between the intraday periodic component and the ARCH process. As the ARCH model parameters are estimated with the use of the filtered series, the estimated model can only assure an optimal fit for the filtered series only. There is no indication that the recovered intraday volatility, product of the intraday periodic component times the ARCH conditional volatility, fits the raw intraday series optimally.

An integrated framework that utilizes the interaction effect between the intraday periodic component and the ARCH conditional volatility is proposed to model the volatility of the raw intraday return series. The framework incorporates the filtration process and the ARCH structure in a united setting such that the intraday periodic component and the ARCH conditional volatility are present in a time. The recovered intraday volatility is revealed and its fitness can be assessed during the model estimation process. Furthermore, to ensure an optimal fit, the integrated framework makes use of the likelihood of the raw intraday return series instead of the use of the filtered series as the sequential approach does. As a result, under the same level of model flexibility (parameterized portions), the optimal parameter set from the integrated framework should achieve the largest likelihood of the raw return series.

The key advantage of the integrated framework relies on its ability to adjust the model parameters such that an intraday periodic component and it corresponding ARCH conditional volatility can work in a complementary way. Consider a situation that an intraday innovation series $\{\xi_{t,n}\}$ is equal to the product of a specific intraday periodic component series $\{s^*_{t,n}\}$ times a series (filtered series) which is based on using $\{\tilde{\sigma}^*_{t,n}\}$ as the underlying volatilities.

To model the corresponding intraday volatility process, the intraday periodic component is needed to be estimated in the very beginning. However, it is clear that the estimation process is not error free in practice. Denote the estimated intraday periodic component as $s_{t,n}$, which is different from

$s^*_{t,n}$, and the estimated ARCH volatility as $\tilde{\sigma}_{t,n}$. For the sequential approach, the ARCH parameters are estimated with the filtered series only. Since the erroneous $s_{t,n}$ produces an erroneous filtered series, the estimated $\tilde{\sigma}_{t,n}$ will most probably different from the $\tilde{\sigma}^*_{t,n}$. As a result, the overall error of the recovered intraday volatility $s_{t,n} \cdot \tilde{\sigma}_{t,n}$ is equal to the product of the errors from the $s_{t,n}$ and $\tilde{\sigma}_{t,n}$. The optimality of the recovered intraday volatility series for the original $\{\xi_{t,n}\}$ series is not guaranteed.

On the other hand, under our proposed integrated framework, the model parameters are adjusted to ensure the $\tilde{\sigma}_{t,n}$ works complementarily to the $s_{t,n}$. The composite term $s_{t,n} \cdot \tilde{\sigma}_{t,n}$, which is equal to the recovered intraday volatility, is optimized to fit the $\{xi_{t,n}\}$ series in the best possible way. The adverse impact from the erroneous $s_{t,n}$ is alleviated.

The following formulations detail our proposed framework. Define the likelihood function of the raw intraday innovations as:

$$LF_{raw} = \prod_{\forall (t,n) \in \mathbf{S}} \phi(\sigma_{t,n} | \xi_{t,n}) \tag{3.31}$$

where $\phi(\sigma|x)$ denotes a probability density function of a variable x with its mean equals to zero and its standard deviation equals to σ. $\mathbf{S}$ is a set containing all the day and interval index pairs (t,n) for the sample data. $\xi_{t,n}$ and $\sigma_{t,n}$ are the innovations of the raw series and its volatility respectively.

As the intraday volatility is modeled as the product of the intraday periodic component and the ARCH conditional volatility, the following equation holds:

$$\sigma_{t,n} = s_{t,n} \cdot \tilde{\sigma}_{t,n} \tag{3.32}$$

Our proposed approach makes use of the identical values of $s_{t,n}$ as Andersen & Bollerslev's sequential approach does. The estimate of the composite term $s_{t,n} \cdot \tilde{\sigma}_{t,n}$ is regarded as the recovered intraday volatility.

Let $g(\theta|\{\tilde{r}_{t,n}\}, t, n)$ be a function to calculate the conditional volatility series for a particular n-th intraday interval on day t by a specified ARCH structure with the use of a particular parameter set θ given the filtered series $\{\tilde{r}_{t,n}\}$. The optimal parameter set $\hat{\theta}$ for the likelihood LF_{raw} can be estimated as:

$$\hat{\theta} = \arg_{\theta} max \prod_{\forall(t,n)\in \mathbf{S}} \phi(s_{t,n} \cdot g(\theta|\{\tilde{r}_{t,n}\}, t, n)|\xi_{t,n}) \tag{3.33}$$

With the use of the above framework, the optimal parameter $\hat{\theta}$ will be driven in a way such that the estimated conditional volatility $g(\theta|\tilde{r}_{t,n}, t, n)$ works with the $s_{t,n}$ cooperatively to achieve the best fit for the intraday innovations series.

For the situation that GARCH(1,1) is selected as the ARCH structure and the innovations are assumed to be i.i.d. normally distributed, our proposed framework can be formulated as below:

$$\begin{aligned} \hat{\theta} &= \arg_{\theta} max \sum_{\forall(t,n)\in \mathbf{S}} l(\theta) \\ l(\theta) &= -\frac{1}{2}\left(log(2\pi) + 2 \cdot log(s_{t,n}\tilde{\sigma}_{t,n}) + (\frac{\xi_{t,n}}{s_{t,n}\tilde{\sigma}_{t,n}})^2\right) \end{aligned}$$

$$
\begin{aligned}
\tilde{r}_{t,n} &= \frac{\xi_{t,n}}{s_{t,n}} \\
\tilde{\xi}_{t,n} &= \tilde{r}_{t,n} - \kappa \\
\tilde{\sigma}^2_{t,n} &= \omega + \alpha(\tilde{\xi}_{t,n-1})^2 + \beta\tilde{\sigma}^2_{t,n-1} \\
\xi_{t,n} &\sim N(0, s^2_{t,n}\tilde{\sigma}^2_{t,n}) \\
\omega, \alpha, \beta &> 0 \\
\theta &= \{\kappa, \omega, \alpha, \beta\}
\end{aligned}
\tag{3.34}
$$

The first two lines indicate that the optimal parameter set θ is estimated with the log likelihood function which is based on the intraday innovation $\xi_{t,n}$ in this framework. $\xi_{t,n}$ is assumed to follow an i.i.d. zero-mean standard normal distribution with its variance equals to $s^2_{t,n}\tilde{\sigma}^2_{t,n}$. The intraday periodic component $s_{t,n}$ is calculated by the same procedures as in Andersen & Bollerslev's sequential estimation method. The ARCH structure $g(\theta|\tilde{r}_{t,n}, t, n)$ is defined by the fourth and fifth lines with the parameter set θ to model the conditional volatility series $\tilde{\sigma}_{t,n}$.

Block diagrams of the two approaches are included to illustrate the differences. In Figure 3.2, the Andersen & Bollerslev's sequential approach handles the intraday periodicity and the heteroskedasticity independently. Their approach estimates the ARCH parameters purely based on the deseasonalized series $\tilde{r}_{t,n}$, therefore, it may be possible that the estimated parameters may not provide the optimal fit for the raw intraday innovation series $\xi_{t,n}$. On the other hand, the proposed integrated approach optimizes its parameters for the $\xi_{t,n}$ series instead and it also considers the interaction between the intraday periodicity and the heteroskedasticity. The proposed framework integrates the filtration and the ARCH structure in a unified framework.

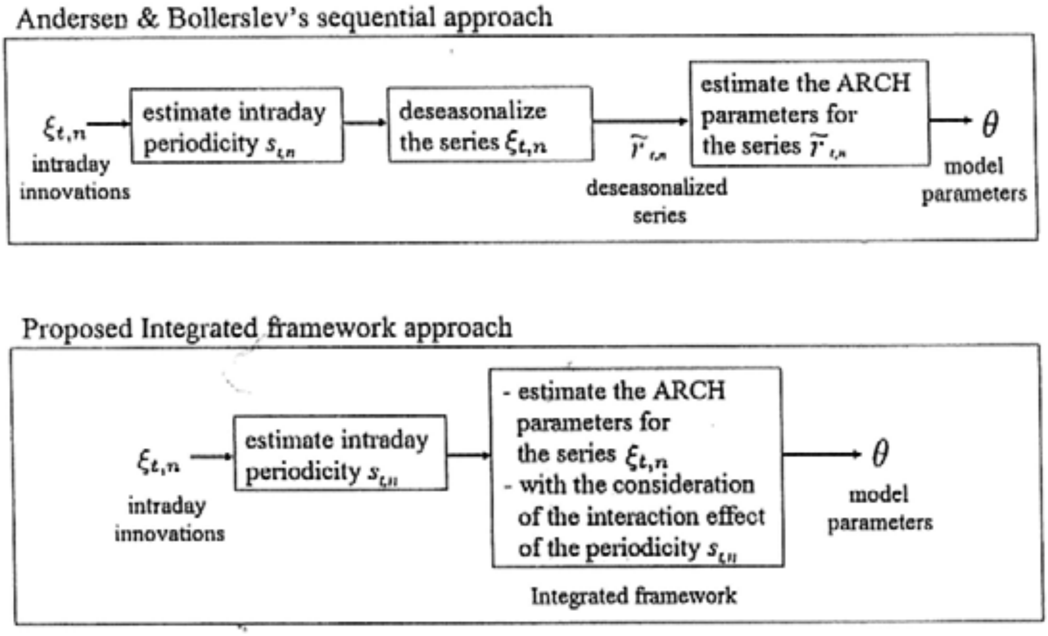


Figure 3.2: Block diagram of the two approaches

3.8.3 Graphical representations on the out-of-sample forecast performance

The following charts try to provide a quick overview on the out-of-sample forecast performance in terms of R^2, MSE and MAE. There are totally 24 tested situations in our study and all of them give very similar findings. The situation - using the GARCH(1,1) structure, deseasonalized by the simple averaging method, with normally distributed innovations and tested by NASDAQ data - is selected as an illustration.

The Figure 3.3, 3.4 and 3.5 indicate the corresponding improvements from using the proposed integrated framework approach over the original

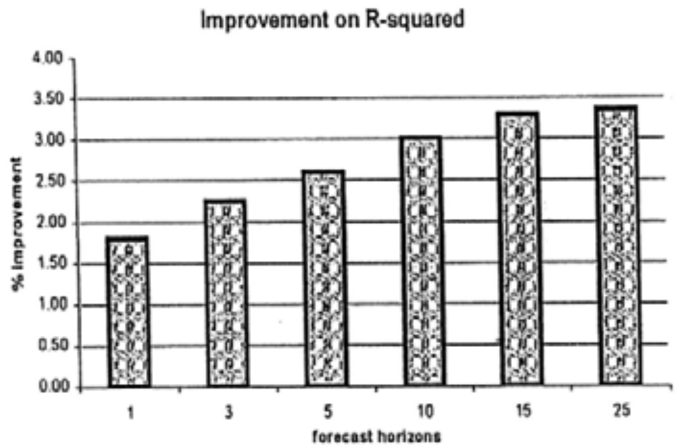


Figure 3.3: Improvement on R-square by using the integrated framework approach

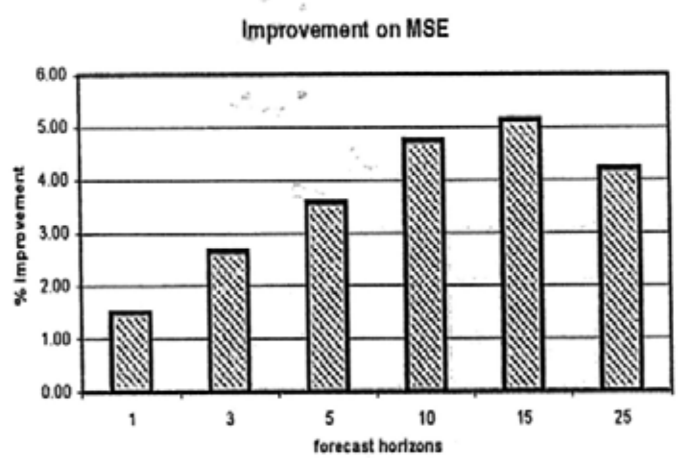


Figure 3.4: Improvement on MSE by using the integrated framework approach

sequential approach. It can be observed that our approach outperforms in the aspects of R^2, MSE and MAE over various forecasting horizons, ranging from 1-step-ahead to 25-step-ahead. The three bar charts represent the results for a particular case out of the total 24 situations tested in the study. Nevertheless, the same findings (i.e. the proposed approach produces higher R^2, smaller MSE and MAE) can also be observed in the other 23 situations.

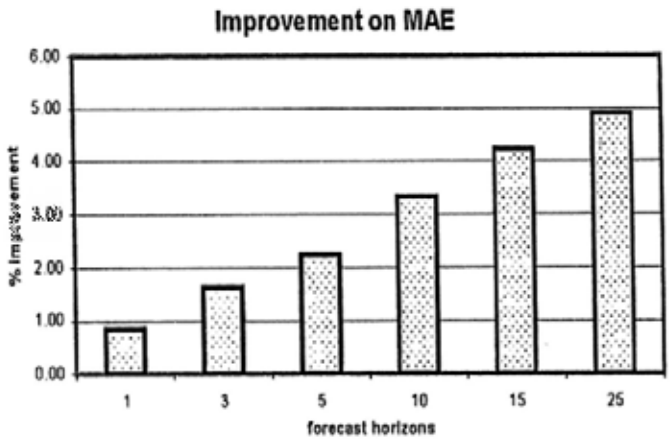


Figure 3.5: Improvement on MAE by using the integrated framework approach

3.8.4 Properties of the realized variance estimator

The following paragraphs provide the mathematical foundations for using the realized variance, sum of squared innovations, as an unbiased estimator of the latent variance parameter of a daily innovation process. It can be shown that the realized variance is more efficient (estimator with smaller estimation error) than a traditional estimator based on a daily innovation.

Consider the situation that a daily innovation is composed of N number of i.i.d. intraday innovations :

$$\begin{aligned} \epsilon_t &= \sum_{i=1}^{N} \epsilon_{t,i} \\ \epsilon_{t,i} &\sim N(0, \frac{\sigma_t^2}{N}) \end{aligned} \tag{3.35}$$

The ϵ_t and $\epsilon_{t,i}$ denote a daily innovation for day t and an intraday innovation for the i-the period on day t respectively. $\epsilon_{t,i}$ is assumed to follow a standard normal distribution with variance equals to σ_t^2/N.

As $\epsilon_{t,i}$ is independent of each other, the variance of ϵ_t equals to :

$$\begin{aligned} Var(\epsilon_t) &= Var(\sum_{i=1}^{N} \epsilon_{t,i}) \\ &= \sum_{i=1}^{N} Var(\epsilon_{t,i}) \\ &= \sum_{i=1}^{N} \frac{\sigma_t^2}{N} \\ &= \sigma_t^2 \end{aligned} \tag{3.36}$$

However, it is important to know that the parameter σ_t^2 is not directly observable in practice. Its value can only be estimated through the information obtained from the observed innovations ϵ_t and $\epsilon_{t,i}$. The properties of two estimators, the simple unbiased estimator and the realized variance estimator, are compared as follows :

Unbiasedness property

The traditional estimator of the latent variable σ_t^2 is a squared daily innovation (i.e. ϵ_t^2). It is an unbiased when the intraday innovations are independent of each other.

$$\begin{aligned} E(\epsilon_t^2) &= E[(\epsilon_t - 0)^2] \\ &= Var(\epsilon_t) \\ &= \sigma_t^2 \end{aligned} \tag{3.37}$$

On the other hand, the another candidate, realized variance estimator, is defined as the sum of squared intraday innovations (i.e. $\sum_{i=1}^{N} \epsilon_{t,i}^2$). It is also an unbiased estimator since :

$$\begin{aligned} E(\sum_{i=1}^{N} \epsilon_{t,i}^2) &= \sum_{i=1}^{N} E(\epsilon_{t,i}^2) \\ &= \sum_{i=1}^{N} \frac{\sigma_t^2}{N} \\ &= N * \frac{\sigma_t^2}{N} \\ &= \sigma_t^2 \end{aligned} \tag{3.38}$$

Variance (accuracy) of the estimators

The traditional estimator (ϵ_t^2) and the realized variance estimator ($\sum_{i=1}^{N} \epsilon_{t,i}^2$) are proved to be unbiased in the above section. However, it can be shown that the realized variance estimator does a better job as it is more efficient (estimator with smaller estimation error) than the traditional estimator. Consider the variability (variance of the estimator) of the realized variance estimator.

$$\begin{aligned} Var(\sum_{i=1}^{N} \epsilon_{t,i}^2) &= \sum_{i=1}^{N} Var(\epsilon_{t,i}^2) \\ &= \sum_{i=1}^{N} \left\{ E(\epsilon_{t,i}^4) - [E(\epsilon_{t,i}^2)]^2 \right\} \\ &= N * \left\{ E([\frac{\sigma_t}{\sqrt{N}} * z]^4) \right\} - N * [\frac{\sigma_t^2}{N}]^2 \\ &= N * \frac{(\sigma_t)^4}{N^2} E(z^4) - \frac{\sigma_t^4}{N} \\ &= \frac{(\sigma_t)^4}{N}(E(z^4) - 1) \end{aligned} \tag{3.39}$$

where z is a standard normal variate with $E(z^4) = 3$.

The above value is shown to be smaller than the variance of the traditional estimator :

$$\begin{aligned} Var(\epsilon_t^2) &= Var([\sigma_t z]^2) \\ &= Var(\sigma_t^2 z^2) \\ &= (\sigma_t)^4 Var(z^2) \\ &= (\sigma_t)^4 \left\{E(z^4) - [E(z^2)]^2\right\} \\ &= (\sigma_t)^4 (E(z^4) - 1) \\ &> Var(\sum_{i=1}^{N} \epsilon_{t,i}^2) \end{aligned} \quad (3.40)$$

As a result, the realized variance is a more efficient estimator when the number of intraday intervals $N > 1$. Furthermore, the realized variance is also a consistent estimator as its variance decrease with the increase of N (equation 3.39).

Applicability under a general setting

The argument of the superiority of using the sum of intraday innovations as the estimator is also valid when the daily innovation is defined as the aggregation of intraday innovations with different level of variability.

Consider a daily innovation is re-defined as:

$$\begin{aligned} \epsilon_t &= \sum_{i=1}^{N} \epsilon_{t,i} \\ \epsilon_{t,i} &\sim N(0, w_i \sigma_t^2) \end{aligned} \quad (3.41)$$

where the w_i reflect the level of variability of the i-th intraday innovation. Without loss of generality, the weight factor is formulated to follow $\sum_{i=1}^{N} w_i = 1$ and $w_i > 0$. Therefore, it can be shown that $Var(\epsilon_t) = \sigma_t^2$.

The unbiasedness property still holds:

$$
\begin{aligned}
E(\sum_{i=1}^{N}\epsilon_{t,i}^2) &= \sum_{i=1}^{N}E(\epsilon_{t,i}^2) \\
&= \sum_{i=1}^{N}(w_i\sigma_t^2) \\
&= \sigma_t^2\sum_{i=1}^{N}w_i \\
&= \sigma_t^2
\end{aligned}
\tag{3.42}
$$

$$
\begin{aligned}
E(\epsilon_t^2) &= E[(\sum_{i=1}^{N}\epsilon_{t,i})^2] \\
&= E[\sum_{i=1}^{N}\epsilon_{t,i}^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\epsilon_{t,i}\epsilon_{t,j}] \\
&= \sum_{i=1}^{N}E(\epsilon_{t,i}^2) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}E(\epsilon_{t,i}\epsilon_{t,j}) \\
&= \sum_{i=1}^{N}E(\epsilon_{t,i}^2) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}0 \\
&= \sigma_t^2
\end{aligned}
\tag{3.43}
$$

The variance of the estimators are :

$$
\begin{aligned}
Var(\sum_{i=1}^{N}\epsilon_{t,i}^2) &= \sum_{i=1}^{N}Var(\epsilon_{t,i}^2) \\
&= \sum_{i=1}^{N}\left\{E(\epsilon_{t,i}^4) - [w_i\sigma_t^2]^2\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} E(\epsilon_{t,i}^4) - \sum_{i=1}^{N} w_i^2 \sigma_t^4 \\
&= \sum_{i=1}^{N} \left\{ E([\sqrt{w_i}\sigma_t z]^4) \right\} - \sum_{i=1}^{N} w_i^2 \sigma_t^4 \\
&= \sum_{i=1}^{N} w_i^2 \sigma_t^4 E(z^4) - \sum_{i=1}^{N} w_i^2 \sigma_t^4 \\
&= \sum_{i=1}^{N} w_i^2 * \sigma_t^4 (E(z^4) - 1) \qquad (3.44)
\end{aligned}
$$

$$
\begin{aligned}
Var(\epsilon_t^2) &= Var([\sigma_t z]^2) \\
&= (\sigma_t)^4 (E(z^4) - 1) \\
&> \sum_{i=1}^{N} w_i^2 * \sigma_t^4 (E(z^4) - 1) \qquad (3.45)
\end{aligned}
$$

Under the initial setting where $\sum_{i=1}^{N} w_i = 1$ and $w_i > 0$, it implies $w_i < 1$. The inequality in the equation set 3.45 holds because as $\sum_{i=1}^{N} w_i^2 < 1$. Furthermore, the variability of the estimator, sum of squared intraday innovations, tends to decrease with the increase of the number of intraday interval N.

□ **End of chapter.**

Chapter 4

Day-varying structure for modeling intraday periodicity

Summary

The periodicity appearing in an intraday volatility process is always modeled to follow an identically repeating structure. It varies in a U-shape pattern within a trading day while its value for every particular intraday interval is identical across days. This rigid day-invariant structure may hinder the potential usage of the variability of periodicity across days. Andersen & Bollerslev provide a method that is capable of estimating the periodicity with the allowance of day-variability. However, the performance of the day-variant periodicity is always demonstrated to be inferior to the corresponding day-invariant version empirically.

We improve their normalization method by adjusting the estimated values to fulfill the implicit constraint for the construction of daily variances from their corresponding intraday variances. For the situation that the periodicity is modeled to be day-variant, the proposed method is shown to be less susceptible to heteroskedastic errors through numerical simulations. For day-invariant periodicity, our method is proven to give the same performance as Andersen and Bollerslev's method does mathematically. Furthermore, our method is tested by using 10-minute returns of NASDAQ index (3-year period) and S&P 500 index (2.5-year period). Preference for using the proposed method is supported empirically.

4.1 Introduction and the research objective

The periodicity appearing in an intraday volatility process is always modeled to follow an identically repeating structure [4, 8, 17, 24, 53, 54, 55]. It varies in a U-shape pattern within a trading day while its value for every particular intraday interval is identical across days. It is speculated that this rigid day-invariant structure may hinder the potential usage of the variability of periodicity across days.

Andersen & Bollerslev provide a method to estimate the intraday periodicity with the allowance of day-variability [3]. The periodicity is modeled in two steps. The dynamics of an intraday return series is firstly approximated by a smoothing function (Flexible Fourier Form) and, secondly, the periodicity is recovered by a normalization procedure with the use of the approximation results. Their method is capable to define the estimated periodicity to be either day-invariant or day-variant with proper adjustments. However, when the method is applied for day-variant situation, it can be shown that the resultant periodicity violates the implicit constraint, which is derived from the initial modeling assumption, in some situations. As a result, the correctness of the periodicity cannot be assured at all times.

Furthermore, the normalization procedure proposed by Andersen & Bollerslev is shown to be susceptible to heteroskedastic errors. It is demonstrated that the time series of the expected periodicity deviates from its true value when the approximated series is contaminated with heteroskedastic errors. The conjecture of the presence of heteroskedastic errors is supported empirically. The approximation errors of the smoothing function are demonstrated to have non-constant variances. The Bartlett's Test [16] and the Brown-Forsythe's Test [23] significantly rejected the null hypothesis of constant error variance at 5% significance.

A modified normalization procedure that ensures the fulfillment of the implicit constraint is proposed in this study. Our procedure regulates the magnitude of the periodicities with reference to the size of their corresponding daily variances, guaranteeing that the coherence of the intraday variances and their aggregated results is met. The procedure is capable for handling both day-variant and day-invariant periodicities. When the proposed procedure is applied to the day-variant situations, the results turn out to be more robust to heteroskedastic errors than the original method under numeric simulations. Besides, the modified method is demonstrated to give superior performance in intraday volatility modeling empirically.

For the day-invariant situation, the series of the estimated periodicity resulting from the modified method can be shown to be a scaled version of the results from the original method. Besides, it can be mathematically proven that the modified method gives identical performance in intraday volatility modeling as the original normalization does.

The remainder of this work is organized as follows: Section 4.2 contains a brief review of Andersen & Bollerslev's procedure for estimating the intraday periodicity. The limitations of direct application to model day-variant periodicity is also addressed in this section. The rationale behind our modification together with its formulation are discussed in Section 4.3. Section 4.4 and 4.5 illustrate the influence from heteroskedastic errors on the estimated periodicity. The existence of heteroskedastic errors is verified empirically in Section 4.4 and the robustness of the normalizations are checked by a simulation study in Section 4.5.

The effect of applying the modified day-variant periodicity to model an intraday volatility process will be investigated in Section 4.6. The modified day-variant periodicity will be tested against the original day-variant version and the day-invariant version, with the use of GARCH(1,1) and FIGARCH(1,d,1) structures under the assumption of either normally distributed innovations or t-distributed innovations. The measurements criteria includes R^2, mean squared error(MSE) and mean absolute error (MAE). Beside the investigation of one-step-ahead out-of-sample forecasts, several forecasting horizons will also be studied to reflect the trends among the multi-step ahead performances. Two sets of 10-minute intraday returns, NASDAQ and S&P 500 indexes, are used in the empirical investigation. Section 4.7 contains the concluding remarks.

4.2 Andersen & Bollerslev's intraday periodicity estimation and its limitation

The seminal work by Andersen & Bollerslev establishes a good example for modeling the intraday periodicity [3]. They notice a prominent U-shape periodic pattern in an absolute intraday return series and discuss the inappropriateness to employ a ARCH type model on an intraday return series directly. To solve the problem, they propose the periodicity should be filtered out before the ARCH model is applied. The estimation of periodicity is carried out in two steps. The dynamics of an intraday return series is firstly approximated by a smoothing function (Flexible Fourier Form) and, secondly, the periodicity is recovered by a normalization procedure with the use of the smoothed results.

The estimation of intraday periodicity is based on an assumption on an intraday return process:

$$r_{t,n} = E(r_{t,n}) + s_{t,n} \cdot \frac{\sigma_t}{\sqrt{N}} \cdot z_{t,n} \tag{4.1}$$

where $r_{t,n}$ is an intraday return of day t in n-th intraday interval and $E(r_{t,n})$ is the unconditional expectation of $r_{t,n}$. The $r_{t,n}$ is calculated as $log(P_{t,n}/P_{t,n-1})$. $P_{t,n}$ is the index value and $P_{t,0}$ is the value at the market open.

The intraday periodicity is denoted as $s_{t,n}$ and it reflects the U-shape regularity across days. σ_t is a daily volatility (standard deviation) of the return in day t and N is the number of intraday interval per day. $z_{t,n}$ is a i.i.d. standard normal random term. The GARCH(1,1) conditional volatility is chosen as the substitute of σ_t in this paper.[1] The above formula provides a simplified expression of an intraday return process and makes the estimation of the intraday periodicity feasible.

By squaring and taking logs on both sides on the equation 4.1, the $s_{t,n}$ can be expressed in terms of an intermediate proxy variable for the intraday dynamics $x_{t,n}$ under the following settings:

$$\begin{aligned} x_{t,n} &\equiv 2log[|r_{t,n} - E(r_{t,n})|] - log(\sigma_t^2) + log(N) \\ x_{t,n} &= f(\Theta|\sigma_t, n) + u_{t,n} \\ \hat{x}_{t,n} &\equiv f(\hat{\Theta}|\hat{\sigma}_t, n) \end{aligned} \tag{4.2}$$

The function $f(\Theta|\sigma_t, n)$ is a modified version of the flexible Fourier form[39] which is defined as :

$$f(\Theta|\sigma_t, n) = \sum_{j=0}^{J} \sigma_t^j [\sum_{k=0}^{K} \nu_{kj} \cdot n^k + \sum_{p=1}^{P} (\gamma_{pj} cos\frac{pn2\pi}{N} + \delta_{pj} sin\frac{pn2\pi}{N})] \tag{4.3}$$

where Θ represents the parameter set $\{\nu_{kj}, \gamma_{pj}, \delta_{pj}\}$. Variables J, K and P control the order of expansion. The parameter set Θ is estimated by setting $f(\Theta|\sigma_t, n)$ as a regressor (independent variable) for a dummy variable $x_{t,n}$ of a linear regression with $u_{t,n}$ as a zero mean i.i.d. error term. $\hat{x}_{t,n}$ is an approximation of the underlying periodic dynamic. The $\hat{x}_{t,n}$ is allowed to vary across days by setting $J > 0$.

After obtaining the approximated value of the intraday dynamics $\hat{x}_{t,n}$, a

[1] There are a number of estimators available to evaluate the value of σ_t, including ARCH type conditional volatility [33], realized variance [6] and range-based estimators [41, 64, 66, 68, 73].

normalization step is used to retrieve the intraday periodicity $s_{t,n}$ accordingly.

$$s_{t,n} = \frac{S \cdot exp(\frac{\hat{x}_{t,n}}{2})}{\sum_{i=1}^{S/N} \sum_{j=1}^{N} exp(\frac{\hat{x}_{i,j}}{2})} \tag{4.4}$$

where S is the sample size of the intraday return series.

The inclusion of a day-varying factor in the flexible Fourier form function (by setting $J > 0$ in the equation 4.3) always helps to improve the approximation accuracy. However, the day-varying property of the approximation results can cause the recovered intraday periodicities to violate the the implicit constraint, which is derived from the initial modeling assumption of an intraday return process, in some situations. The following paragraphs elaborate the concept.

With the initial modeling assumption of an intraday return process listed in the equation 4.1, the calculation of the variance of daily return can be formulated as:

$$\begin{aligned}
\text{Since} \qquad r_t &= \sum_{n=1}^{N} r_{t,n} \\
&= \sum_{n=1}^{N} \left(E(r_{t,n}) + s_{t,n} \cdot \frac{\sigma_t}{\sqrt{N}} \cdot z_{t,n} \right) \\
\Rightarrow \qquad Var(r_t) &= Var\left(\sum_{n=1}^{N} \left(s_{t,n} \cdot \frac{\sigma_t}{\sqrt{N}} \cdot z_{t,n} \right) \right)
\end{aligned} \tag{4.5}$$

where r_t denotes an accumulated return in the active trading period for day t. The variance equation can be simplified as below when the intraday returns are assumed to be independent among themselves.

$$Var(r_t) = \sigma_t^2 \cdot \frac{\sum_{n=1}^{N} s_{t,n}^2}{N} \tag{4.6}$$

Since σ_t represents the daily volatility (square root of variance), the above equation implies the following constraint which should hold for all the possible day t in the sample data.

$$\sum_{n=1}^{N} s_{t,n}^2 = N \tag{4.7}$$

As the main purpose of finding the intraday periodicity is for modeling an intraday volatility process, the exact magnitude of the $s_{t,n}$ is not an important concern. In fact, it is proved that the use of a scaled $s_{t,n}$ series gives equivalent performance on modeling an intraday volatility process as the original series does.[2] Therefore, any intraday periodicity series which fulfills the below condition will perform as good as one which obeys the equation 4.7.

$$\sum_{n=1}^{N} s_{t,n}^2 = k \tag{4.8}$$

where k is any real number. The above condition can be regarded as an implicit compliance to the constraint listed in the equation 4.7 and it is called as the 'implicit constraint' in this paper.

For the situations that the intraday periodicity is restricted to be day-invariant (i.e. the periodicity varies in a U-shape pattern within a trading day while its values for every particular intraday interval are identical across days), the results from normalization in the equation 4.4 are shown to fulfill the implicit constraint automatically.

[2]Mathematical elaboration is provided in 4.8.1

Consider the series of intraday periodicity for a particular day t_1, let:

$$k_1 = \sum_{n=1}^{N} s_{t_1,n}^2 \tag{4.9}$$

When the periodicity is day-invariant (i.e. $s_{d_1,n} = s_{d_2,n}$ for all possible d_1, d_2 within the sample data), we know for sure that the below condition holds for all possible day d. As a result, the implicit constraint is fulfilled automatically.

$$\sum_{n=1}^{N} s_{d,n}^2 = k_1 \tag{4.10}$$

On the other hand, for day-variant situations, the Andersen & Bollerslev's normalization can be shown to violate the implicit constraint in most situations. when the periodicity is day-variant (i.e. $s_{d_1,n} \neq s_{d_2,n}$ for distinct day d_1 and d_2), the condition in the equation 4.8 will not hold for most circumstances. It is because the sum of the $s_{t,n}^2$ per day will vary along days as there is not any constraint to ensure the fulfillment of the above condition for distinct days.

4.3 Proposed normalization procedure for estimating intraday periodicity

As the estimation of intraday periodicity is based on a pre-specified assumption (the equation 4.1), we speculate that the fulfillment of the implicit constraint, which is derived from the assumption, may improve the performance of the estimated results. A modified normalization procedure is proposed as:

$$s_{t,n} = \frac{\sqrt{N} \cdot exp(\frac{\hat{x}_{t,n}}{2})}{\left(\sum_{j=1}^{N} exp(\hat{x}_{t,j})\right)^{1/2}} \tag{4.11}$$

The above formulation ensures $\sum_{n=1}^{N} s_{t,n}^2 = N$ for all possible day t. As a result, the variance equation listed in the equation 4.6 holds and the implicit constraint is fulfilled.

For the situation that the intraday periodicity $s_{t,n}$ is restricted to be non day-varying, the performance of the estimated periodicity for intraday volatility modeling becomes the same as those from the use of the original procedure.[3] The modified normalization gives different results when $s_{t,n}$ is day-varying (by specifying the order of expansion of the flexible Fourier form to $J \geq 1$).

[3]Full elaboration is provided in 4.8.2

4.4 Characteristics of the residuals of the approximation function

Apart from the violation of the implicit constraint, the original normalization procedure can be demonstrated to be susceptible to heteroskedastic errors. The conjecture of the presence of heteroskedasticity is supported empirically. The residuals from the approximation function is shown to be heteroskedastic through the Bartlett's test and the Brown-Forsythe's test. As the approximated results are used as the substitute of the true underlying dynamics, it is reasonable to assume the approximated series deviates from its true value with heteroskedastic errors. The following subsections state our findings.

4.4.1 Data description

The data used in our experiment is composed of 10-minute returns for the NASDAQ composite stock index (NASDAQ) from 15 August 2005 to 12 Sept 2008, consisting of 28,392 observations (728 days), and the Standard and Poors composite stock index (S&P 500) from 10 March 2006 to 12 Sept 2008, consisting of 22,893 observations (587 days). The returns are recorded from 9:30 to 16:00 and therefore there are a total of 39 intraday returns per trading day. All the values of intraday return are multiplied by 100 for the ease of presentation. The the descriptive statistics of the two data sets are tabulated in Table 4.1.

Table 4.1: Descriptive statistics of the intraday returns (multiplied by 100)

	NASDAQ	S&P 500
Start date	16-Aug-05	13-Mar-06
End date	12-Sep-08	12-Sep-08
Observation	28,392	22,893
Mean	-0.0008	-0.0001
Max	1.1768	1.4236
Min	-1.4115	-1.4297
Standard deviation	0.1400	0.1362
Skewness	-0.0175	0.1770
Robust skewness	-0.0111	-0.0088
Kurtosis	7.7803	10.8744
Jarque-Bera (p-value)	<0.001	<0.001

Note: The robust skewness [49], which is robust against outliers, is calculated as $\frac{q3+q1-2q2}{q3-q1}$. The Jarque-Bera value is the p-value of the Jarque-Bera test [48] with the null hypothesis that the intraday returns are normally distributed.

4.4.2 Specification of the approximation function

The order of expansion of the approximation function $f(\Theta|\sigma_t, n)$ is set to $J = 1$, $K = 2$ and $P = 4$. Expansion beyond this order gives insignificant coefficients for the additional parameters.

On the other hand, the GARCH(1,1) conditional volatilities used in the equations 4.2 and 4.3 are based on a daily return series with the latest daily return defined as a convex combination of a m-day moving average of historical daily returns and an accumulated sum of intraday returns:

$$\hat{r}_t = (1 - \frac{n}{N})\frac{\sum_{i=t-1}^{t-m} r_i}{m} + \frac{n}{N}\sum_{j=1}^{n} r_{t,j} \qquad (4.12)$$

where r_i means the daily return for day i and $\hat{r}_t$ is the forecast of the latest daily return given the information up to the n-th interval. The variable m is set to 5 for all the empirical investigations in this paper.

4.4.3 Heteroskedasticity of the residuals

The residuals of the approximation function (i.e. $\hat{x}_{t,n} - x_{t,n}$ in the equation set 4.2) are shown to have non-constant variances based on the empirical findings. Two hypothesis tests, the Bartlett's test [16] and the Brown-Forsythe's test [23], are used to check the homoskedasticity of the errors. The Bartlett's test verifies the equality of variances from different samples for normally distributed variables while the Brown-Forsythe's test works for non-normality situations. The null hypothesis of the tests state the variances from the different samples are homoskedastic.

Our investigation organizes the residuals into 10 bins by two different ways. The first way groups the errors according to their calendrical order, where the first group contains the farthest data and the last group contains the most current data. The second way arranges the residuals by the magnitude of their daily volatility, where the errors from the least volatile days are grouped together and the errors from the most volatile days are grouped as another group.

The rolling-sample method is employed for the empirical analysis in this paper. This method updates the training samples step by step. For instance, if the sample size is S, the first sample will be the 1 to S observations, the second sample will be the 2 to $S+1$ observations and so on. The sample size is set to be 200 days (7,800 intraday periods) in this study. Therefore, there are totally 20,592 sample sets for NASDAQ data and 15,093 sets for S&P

500 data. The results of the hypothesis tests are tabulated in Table 4.2.

Table 4.2: Results of Bartlett's test and Brown-Forsythe's test

		Bartlett's test		**Brown-Forsythe's test**	
		by calendrical order	by daily volatility size	by calendrical order	by daily volatility size
NASDAQ	p-value	0.0030	0.0038	0.0539	0.0670
	(Sig. %)	(98.73%)	(98.45%)	(80.22%)	(68.12%)
S&P 500	p-value	0.0017	0.0002	0.0389	0.0159
	(Sig. %)	(99.50%)	(99.98%)	(77.02%)	(91.06%)

Note: The p-values, with the null hypothesis of homoskedasticity, are the averaged results from all the training samples (i.e. 20,592 sets for NASDAQ and 15,093 sets for S&P 500). The Sig. % denotes the percentage of rejecting the null hypothesis with 5% significant level.

In Table 4.2, the p-values are the averaged results from all the training samples. The Sig. % denotes the percentage of the acceptance of the alternative hypothesis, the residuals are heteroskedastic with 5% significant level. The results from the Bartlett's test strongly indicate that the residuals are heteroskedastic. For both NASDAQ and S&P 500 data sets, over 98% of the samples reject the null hyopthesis with 5% significant level, no matter the errors are grouped by calendrical order or by their corresponding daily volatlity size. The Brown-Forsythe's test gives similar findings, but with fewer supporting samples. Nevertheless, large portion of samples (ranges from 68% to 91% for different combinations) indicate the existence of heteroskedasticity with 5% significant level empirically. Therefore, it is reasonable to assume the residuals are heteroskedastic instead of homoskedastic.

4.5 Simulation study on the robustness of the normalization procedures

The robustness of the two normalization procedures with the presence of two types of errors, homoskedastic and heteroskedastic, are investigated in this subsection. The robustness is measured by the correlation between the series of the true intraday periodicity and the series of the expected periodicity. A larger correlation means the method is the more robust. Theoretically, the expected intraday periodicity recovered from the two normalization of Andersen & Bollerslev's method is defined as:

$$E(s_{t,n}) \;=\; \int_{i=1}^{I}\int_{j=1}^{J}\left(\frac{I\cdot J\cdot exp(\frac{x'_{t,n}+\delta_{i,j}}{2})}{\sum_{i=1}^{I}\sum_{j=1}^{J}exp(\frac{x'_{t,n}+\delta_{i,j}}{2})}\cdot\phi(\delta_{i,j})\right)dj\cdot di \quad (4.13)$$

where T and N denote the number of sample day and the number of intraday interval per day respectively. $x'_{t,n}$ is the true series and $\delta_{i,j}$ is the error, which can be either homoskedastic or heteroskedastic. The $\phi(\delta_{i,j})$ is the pdf of the random error, which is modelled as normally distributed in this study.

For the modified normalization, the expected periodicity is calculated as:

$$E(s_{t,n}) \;=\; \int_{j=1}^{J}\left(\frac{\sqrt{J}\cdot exp(\frac{x'_{t,n}+\delta_{i,j}}{2})}{\left(\sum_{j=1}^{J}exp(x'_{t,n}+\delta_{i,j})\right)^{1/2}}\cdot\phi(\delta_{i,j})\right)dj \qquad (4.14)$$

However, direct evaluation of the above two integrals are very difficult as they involve the calculation of the sum of log normal variables ($e^{x'_{t,n}+\delta_{i,j}}$ can be written as $e^{x'_{t,n}}\cdot e^{\delta_{i,j}}$, which is log normal). Studies have shown that the integration of sum of log normal variables cannot be expressed in any closed-form expression [58]. As a result, the robustness will be investigated through a simulation study.

The first step to conduct the experiment is to define a time series representing the true underlying intraday dynamics $x'_{t,n}$. Since the normalizations involve the use of a nonlinear function (log function), the magnitude of the series should be concerned. Three series with different mean values are used as the benchmark for checking the robustness of the expected values. The three benchmark series are obtained through a linear scaling of the following series:

$$x_{t,n}^{shape} \quad = \quad d_t + cos\left(\frac{\pi}{2} + \frac{n\pi}{N}\right) \tag{4.15}$$

The $x_{t,n}^{shape}$ is a day-varying periodic series which is composed of two components, d_t and $cos\left(\frac{\pi}{2} + \frac{n\pi}{N}\right)$. The former represents the day-varying portion and the latter represents the periodic portion. The variable N is set to 39 to match the actual number of intraday interval per day of our experiment data.

The day varying portion d_t is formulated to follow a simple oscillating sequence:

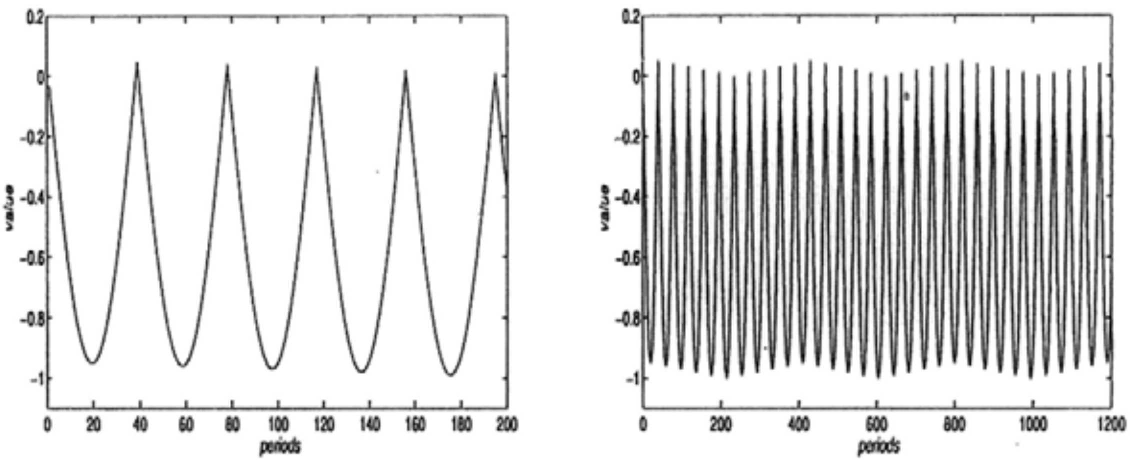
$$d_t \quad = \quad d_{t-1} + \Delta_t \tag{4.16}$$

where $\Delta_t = -\Delta_{t-1}$ when $d_t = \overline{d}$ or $d_t = \underline{d}$, otherwise $\Delta_t = \Delta_{t-1}$. The symbols, $\overline{d}$ and $\underline{d}$, denote the upper bound and lower bound respectively. The series d_t is defined by setting $d_1 = \overline{d}$, $\Delta_1 = -0.01$, $\overline{d} = 0.05$, $\underline{d} = 0$ in this simulation experiment.

On the other hand, the periodic portion $cos\left(\frac{\pi}{2} + \frac{n\pi}{N}\right)$ is used to introduce the periodic pattern in the series $x_{t,n}^{shape}$. The shape of the time series $x_{t,n}^{shape}$ for our simulation study is depicted in figure 4.1.

Figure 4.1: Shape of the time series $x_{t,n}^{shape}$

(a) 200 intervals (b) 1200 intervals



Overall, the true underlying intraday dynamics $x'_{t,n}$ is formulated as:

$$x'_{t,n} \quad = \quad scaling \cdot x_{t,n}^{shape} \tag{4.17}$$

where the parameter *scaling* is set in a way to transform the series $x'_{t,n}$ to have a specified mean value.

The second assumption made in our simulation is about the characteristics of the heteroskedastic error. The heteroskedasticity of the error is mimicked by the daily change of its standard deviation. The standard deviation of the error $\sigma(\delta_i^{shape})$ is assumed to follow an oscillating sequence:

$$\sigma(\delta_i^{shape}) \quad = \quad \sigma(\delta_{i-1}^{shape}) + \Delta'_t \tag{4.18}$$

where δ_i^{shape} is the time index in the above function. $\Delta_t' = -\Delta_{t-1}'$ when $\sigma(\delta_i^{shape}) = \overline{\sigma(\delta)}$ or $\sigma(\delta_i^{shape}) = \underline{\sigma(\delta)}$, otherwise $\Delta_t' = \Delta_{t-1}'$. $\overline{\sigma(\delta)}$ and $\underline{\sigma(\delta)}$ are the pre-specified upper bound and lower bound respectively. The series $\sigma(\delta_i^{shape})$ is defined by setting $\sigma(\delta_i^{shape}) = \overline{\rho(\delta)}$, $\Delta_1' = -0.5$, $\overline{\sigma(\delta)} = 1.1$, $\underline{\sigma(\delta)} = 0.1$.

Like the procedures to handle the $x_{t,n}'$ series, an underlying error series $\delta_{i,j}^{shape}$ is generated at the beginning and the actual error series δ_t is then obtained through a linear scaling. The $\delta_{i,j}^{shape}$ is defined as:

$$\delta_{i,j}^{shape} \sim N(0, [\sigma(\delta_i^{shape})]^2) \tag{4.19}$$

The magnitudes of the series $x_{t,n}'$ and the error series $\delta_{i,j}$ is controlled to represent the scenarios with different error intensities. The intensity is categorized by the ratio of the averaged value of absolute $\delta_{i,j}$ to the averaged value of absolute $x_{t,n}'$ (i.e. $\mathrm{E}[|\delta_{i,j}|]/\mathrm{E}[|x_{t,n}'|]$). In addition, as the sensitivity of the normalizations vary with the change of the magnitude of $x_{t,n}'$ (log function is involved in the normalizations), several mean values of the series (i.e. $\mathrm{E}[x_{t,n}']$) are employed to reflect the properties of the normalizations under different operating ranges.

Based on the properties of the empirical data, the mean value of the approximated intraday dynamics $\hat{x}_{t,n}$ ranges from -2.017 to -2.394 and -1.840 to -2.578 for NASDAQ and S&P 500 data sets respectively. Furthermore, the maximum and minimum values are 1.963 and -7.011 for for NASDAQ, 4.230 and -6.540 for S&P 500. In our simulation study, the series of the true value $x_{t,n}'$ is assumed to lie within a similar range of $\hat{x}_{t,n}$. The length of the series is constructed to have 7,800 periods (i.e. 200 days times 39 intraday intervals) and the errors series are simulated for 100,000 times.

Lastly, the expected periodicities for the two normalizations are defined as:

$$E(s_{t,n}) = E\left(\frac{I \cdot J \cdot exp(\frac{x'_{t,n}+\delta_{i,j}}{2})}{\sum_{i=1}^{I}\sum_{j=1}^{J} exp(\frac{x'_{t,n}+\delta_{i,j}}{2})}\right) \quad (4.20)$$

$$E(s_{t,n}) = E\left(\frac{\sqrt{J} \cdot exp(\frac{x'_{t,n}+\delta_{i,j}}{2})}{\left(\sum_{j=1}^{J} exp(x'_{t,n}+\delta_{i,j})\right)^{1/2}}\right) \quad (4.21)$$

The equation 4.20 denotes the expectation for the original normalization and the equation 4.21 is for the modified procedure.

The correlations between the two series, the series of the intraday periodicity recovered from $x'_{t,n}$ and the series of the expected periodicity, under various situations are tabulated in Table 4.3.

Table 4.3: Heteroskedastic error - Correlation between the true series and the expected series

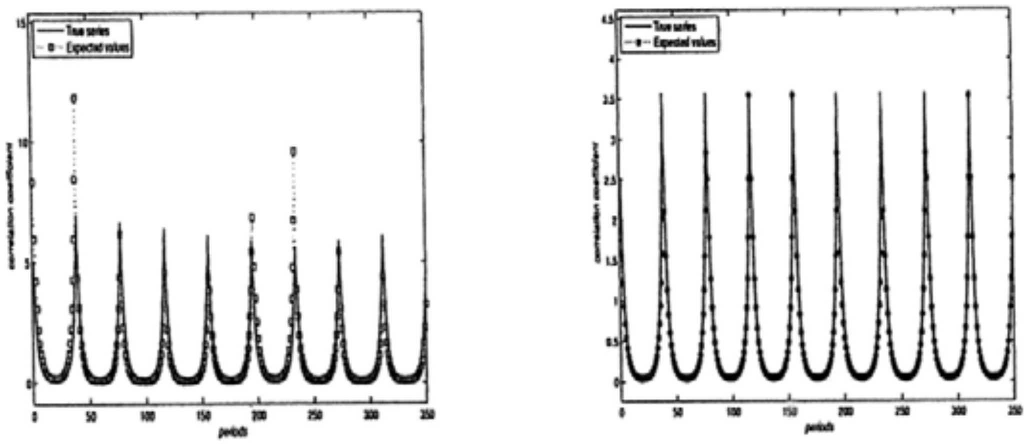
	Error-to-signal ratio					
	0.1		0.2		0.3	
	Original normalization	Modified normalization	Original normalization	Modified normalization	Original normalization	Modified normalization
Mean = -2.6	0.999	>0.999	0.983	0.994	0.912	0.982
Mean = -2.2	>0.999	>0.999	0.987	0.995	0.932	0.985
Mean = -1.8	>0.999	>0.999	0.990	0.996	0.948	0.987

Note: The numbers represent the correlation between the series of the intraday periodicity recovered from the series of $x'_{t,n}$ and the series of the corresponding expected value. The Error-to-signal denotes the ratio between the average of the absolute error series and the average of the true absolute $x'_{t,n}$ series (i.e. $E[|\delta_{i,j}|]/E[|x'_{t,n}|]$). The Mean value indicates the the average of the true $x'_{t,n}$ series (i.e. $E[x'_{t,n}]$). The boundaries of $x'_{t,n}$ are 2.379 to -6.621, 2.779 to -6.221 and 3.179 to -5.821 for the mean values of -2.6, -2.2 and -1.8 respectively.

With the presence of heteroskedastic error, the correlations from the mod-

Figure 4.2: Expected values: Error-to-signal ratio=0.3, Mean=-2.6

(a) Original Normalization (b) Modified Normalization



ified normalization procedure are shown to be greater than those from the original normalization. The correlations from the modified normalization range from >0.999 to 0.982, and the results from the original normalization are from >0.999 to 0.912, for the nine simulation conditions. The gain from the modified procedure becomes larger when either the mean value or the error intensity is increased or both. The correlation from the original normalization decreases to 0.912 for the situation of $E[x'_{t,n}]$=-2.6 and error intensity ($\mathrm{E}[|\delta_{i,j}|]/\mathrm{E}[|x'_{t,n}|]$)) is 0.3, whereas the correlation of the modified procedure is 0.982 under the same simulation condition.

On the other hand, for the presence of homoskedastic errors, the correlations from the original normalization and the modified procedure are all greater than 0.999 for the above nine combinations. The results are not tabulated here for clarity.

Overall, as the residuals of the approximation function of the intraday dynamics are demonstrated to be heteroskedastic in the subsection 4.4.3, it is reasonable to assume the approximated series deviates from its true value with heteroskedastic errors instead of homoskedastic errors. With the presence of heteroskedastic errors, the expected intraday periodicity from the modified normalization gives higher correlations than those from the original normalization under different simulation conditions. The next section illustrates the benefits of applying the modified normalization for intraday volatility modeling empirically.

4.6 Application on intraday volatility modeling

It is speculated that the violation of the implicit constraint may give adverse effects on the estimated intraday periodicity. However, as the magnitude of the periodicity is not directly measurable, there is no way to obtain its true underlying value from the market data and therefore, the performance of the estimated periodicity cannot be check straight forwardly. To show the benefits for real applications, the two normalizations are compared on the overall performance of modeling intraday volatility series.

4.6.1 Data description and sampling method

The descriptive statistics of the two data sets, the NASDAQ and the S&P 500 index returns, are summarized in Table 4.1 under the sub-section 4.4.1. The results indicate that the two data sets exhibit excess Kurtosis and their returns are significantly rejected by the Jarque-Bera normality test. It is advisable to consider other distribution type to model the intraday innovations besides the usage of normal distribution. The empirical works from Bollerslev and Wilhelmsson demonstrated that the use of t-distributed errors helps to handle the high kurtosis when GARCH(1,1) model is employed [20, 71]. Therefore, both normally distributed and t-distributed innovations [4] will be employed in this paper. On the other hand, the sample size is set to be 200 days (7,800 intraday periods) in this study. Under the rolling-sample method described in the sub-section 4.4.3, there will be 528 days (20,592 periods) and

[4]The likelihood function for t-distribution is based on Bollerslev's work on 1987. [20]

387 days (15,093 periods) out-of-sample observations for the NASDAQ and the S&P 500 data sets respectively.

4.6.2 Specification of Engle's Multiplicative Component ARCH model

Based on the formulation of the high frequency intraday volatility in Engle et al's paper [37], an intraday volatility can be modeled by the following multiplicative components:

$$\sigma_{t,n} = s_{t,n} \cdot \sigma_t \cdot \sqrt{q}_{t,n} \tag{4.22}$$

where $\sigma_{t,n}$ is an intraday volatility, $s_{t,n}$ and σ_t are the intraday periodicity and the daily conditional volatility respectively. The $q_{t,n}$ is a ARCH conditional volatility.

Two ARCH structures, GARCH(1,1) and FIGARCH(1,d,1) are used to model the $q_{t,n}$ in our empirical investigation. Besides, both normal and t-distributed innovations will be considered. The performance from three distinct sets of estimated intraday periodicity $s_{t,n}$, including the day-varying periodicity from modified normalization, the day-varying results form the original normalization and the day-invariant periodicity, will be studied.

For instance, the overall framework for modeling the intraday volatility with GARCH(1,1) [19] structures with normally distributed innovation can be described as:

$$
\begin{aligned}
\tilde{\xi}_{t,n} &= \frac{r_{t,n} - E(r_{t,n})}{s_{t,n} \cdot \sigma_t} - \kappa \\
q_{t,n} &= \omega + \alpha \cdot \xi_{t,n-1}^2 + \beta \cdot q_{t,n-1} \\
r_{t,n} &\sim N(E(r_{t,n}), [s_{t,n}\sigma_t\sqrt{q}_{t,n}]^2)
\end{aligned}
\tag{4.23}
$$

where $\xi_{t,n}$ denotes a series with heteroskedasticity and κ is its conditional mean. ω, α and β are the GARCH(1,1) parameters. The parameters κ, ω, α and β are estimated with the maximum likelihood method.

Three sets of $s_{t,n}$ are provided for the above framework. Two day-varying periodicity series are obtained by setting $J = 1$ in the modified Fourier form function (the equation 4.3) while the day-invariant periodicity is obtained by setting $J = 0$. The other parameters for the order of expansion are set to $K = 2$ and $P = 4$. Expansion beyond this order gives insignificant coefficients for the additional parameters.

On the other hand, apart from studying the situation of normally distributed innovations, the situation of t-distributed innovations will also be considered in our experiment. Furthermore, the performance of the normalizations will be checked with FIGARCH(1,d,1) in addition to the GARCH(1,1) structure. The variance equation of FIGARCH(1,d,1) [13] is formulated as:

$$
q_{t,n}^2 = \omega + \beta \cdot q_{t,n-1}^2 + [1 - \beta L - (1 - \alpha' L) \cdot (1 - L)^d] \cdot (\tilde{\xi}_{t,n-1})^2 \tag{4.24}
$$

where the model parameters become ω, β, α' and d. The notation L is the lag operator and $(1-L)^d$ is the fractional differencing operator. The parameters are constrained to obey $0 \leq \omega$, $0 \leq d \leq 1 - 2\alpha'$ and $0 \leq \beta \leq \alpha' + d$ to ensure the positivity of $\bar{\sigma}^2_{t,n}$.

The above equation will replace the second line (variance equation) in the equation set 4.23 for representing the FIGARCH(1,d,1) structure.

4.6.3 Evaluation criteria

The performance of the modeling an intraday volatility process is measured by several criteria, including regression R^2, Mean Squared Error (MSE) and Mean Absolute Error (MAE). Furthermore, besides the investigation of one-step-ahead out-of-sample forecast, several forecasting horizons will also be studied to reflect the trends among the trends among the multi-step ahead performance.

The regression R^2 from the Mincer and Zarnowitz type regression is employed in this study to measure the in-sample and out-of-sample model performances. The forecasted volatilities are tested against the 'observed volatility'. The regression takes the form as:

$$V_i^{1/2} \;=\; b_0 + b_1 \cdot \hat{\sigma}_{i,model} + u_i \tag{4.25}$$

where V_i is the 'observed variance' and $\hat{\sigma}_{i,model}$ denotes the forecasted volatility from the corresponding model. The subscript i is the index for the time series and u_i is a zero mean i.i.d. error term. The results listed in this paper will be based on the use of the squared innovation as the 'observed variance'. [5]

[5] Same conclusions are drawn from the empirical analyses with the used of the squared return as the 'observed variance'.

The innovation is defined as:

$$\xi_i = r_i - E(r_i) \tag{4.26}$$

where ξ_i and r_i are the innovation and the return of the period i respectively. $E(r_i)$ is the expected value (unconditional mean) of r_i given the information up to the period i.

The Mean Squared Error (MSE) and the Mean Absolute Error (MAE) are selected as the other category of measurements. They indicate the differences between the forecasted volatility and the 'observed volatility':

$$MSE = \frac{1}{F}\sum_{i=1}^{F}(V_i^{1/2} - \hat{\sigma}_{i,model})^2 \tag{4.27}$$

$$MAE = \frac{1}{F}\sum_{i=1}^{F}|V_i^{1/2} - \hat{\sigma}_{i,model}| \tag{4.28}$$

where F denotes the number of forecasted results.

In additional to the investigation of 1-step-ahead forecast, the forecasting performance will also be checked for multiple horizon situations. The variances of involved periods are summed together to represent an accumulated variance that covers the corresponding multiple horizons. The forecasted volatility for m-horizon with the use of information up to the period i is defined as:

$$\left(\sum_{j=1}^{m}(\hat{\sigma}_{i+j,model})^2\right)^{1/2} \tag{4.29}$$

where $\hat{\sigma}_{i+j,model}$ denotes the forecasted volatility for the $(i+j)$-th period with the use of information up to the period i. For the period $i+1$ to $i+m$, the forecasted volatilities are calculated by the model with the same parameters. The $\hat{\sigma}_{i+1,model}$ is the first-step-ahead forecast, $\hat{\sigma}_{i+2,model}$ is the second-step-ahead forecast and so on.

Similarly, the square root of the 'observed variance' that covers m-horizon is defined as:

$$\left(\sum_{k=i+1}^{i+m}(V_k)^2\right)^{1/2} \tag{4.30}$$

4.6.4 In-sample fitness

Three sets of estimated intraday periodicities are compared under the conditions of normally distributed innovations and t-distributed innovations. The in-sample results for the GARCH(1,1) and the FIGARCH(1,d,1) structure are tabulated in Table 4.4 and 4.5 respectively. The results are the average of the 20,592 and 15,093 training samples for the NASDAQ and the S&P 500 data set correspondingly.

For the GARCH(1,1) structure, the modified day-variant periodicity gives the best R^2 and MSE among the three normalization methods for both NASDAQ and S&P 500 data sets. In contrast, direct application of Andersen & Bollerslev's normalization procedure for evaluating day-variant periodicity (original day-variant) gives the worst performance for all the measurement criteria. The performance of typical day-invariant periodicity ranks in the middle for the criteria of R^2 and MSE. There is not a clear consensus on which normalization achieves the smallest MAE, as the modified day-variant method gives the best result for the S&P 500 data while the day-invariant method gives the best result for the NASDAQ data.

Table 4.4: In-sample fit for GARCH(1,1)

NASDAQ

	normally-distributed innovations					t-distributed innovations				
Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
Modified (Day-variant)	0.004	0.707	**0.1563**	**0.00875**	0.07077	0.010	0.662	**0.1563**	**0.00887**	0.07084
Original (Day-variant)	0.007	0.675	0.1544	0.00889	0.07116	0.013	0.637	0.1546	0.00901	0.07120
Day-invariant	0.005	0.701	0.1554	0.00877	**0.07076**	0.011	0.657	0.1554	0.00890	**0.07082**

S&P 500

	normally-distributed innovations					t-distributed innovations				
Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
Modified (Day-variant)	0.001	0.740	**0.2060**	**0.00867**	**0.06599**	0.005	0.709	**0.2061**	**0.00876**	**0.06598**
Original (Day-variant)	0.007	0.667	0.1940	0.00909	0.06663	0.011	0.639	0.1937	0.00921	0.06664
Day-invariant	0.001	0.734	0.2029	0.00870	0.06608	0.005	0.704	0.2029	0.00879	0.06608

Note: The values of b_0, b_1, R^2, MSE and MAE are the averaged results from all the training samples.

Table 4.5: In-sample fit for FIGARCH(1,d,1)

NASDAQ

	normally-distributed innovations					t-distributed innovations				
Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
Modified (Day-variant)	0.003	0.714	**0.1596**	**0.00870**	0.07068	0.009	0.673	**0.1599**	**0.00883**	0.07084
Original (Day-variant)	0.006	0.683	0.1577	0.00886	0.07115	0.011	0.649	0.1583	0.00897	0.07124
Day-invariant	0.004	0.708	0.1585	0.00873	**0.07066**	0.010	0.668	0.1587	0.00886	**0.07082**

S&P 500

	normally-distributed innovations					t-distributed innovations				
Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
Modified (Day-variant)	0.001	0.740	**0.2104**	**0.00866**	**0.06612**	0.007	0.688	**0.2081**	**0.00901**	**0.06697**
Original (Day-variant)	0.007	0.663	0.1991	0.00921	0.06721	0.013	0.616	0.1958	0.00960	0.06795
Day-invariant	0.001	0.736	0.2077	0.00869	0.06620	0.007	0.684	0.2053	0.00903	0.06703

Note: The values of b_0, b_1, R^2, MSE and MAE are the averaged results from all the training samples.

The same findings are observed for the FIGARCH(1,d,1) structure. The modified day-variant periodicity gives the best R^2 and MSE, the typical day-invariant periodicity gives the second best performance and the original day-variant gives the worst. The results of MAE indicates the modified day-variant method and the day-invariant method achieve the smallest values for the S&P 500 and the NASDAQ data set respectively. The original day-invariant gives the worst MAE for both data sets.

Overall, the in-sample results of R^2 and MSE show a positive preference on the modified day-variant periodicity under the GARCH(1,1) and

FIGARCH(1,d,1) structure for both normally and t-distributed innovations.

4.6.5 Out-of-sample forecasting performance

The out-of-sample performance over different forecasting horizons for using GARCH(1,1) structure are tabulated in Table 4.6 and Table 4.7.

Regardless of the distributional assumption (either normally-distributed innovations or t-distributed innovations), the modified day-variant periodicity gives the best R^2 and MSE for the forecasting horizons ranging from 1 to 25 and the best MAE for horizons 3 to 25, for the NASDAQ data set. For the S&P 500 data set, it achieves the best R^2, MSE and MAE for horizons ranging from 1 to 25, except for one specific case. The R^2 of 1-step-ahead forecast with normally-distributed innovations from the use of the modified day-variant periodicity is slight lower than the day-invariant version (i.e. 0.2178 vs. 0.2180).

On the other hand, for the FIGARCH(1,d,1) structure, the modified day-variant periodicity also shows a strong preference over the other methods for most situations. According to the results in Table 4.8, the modified day-variant periodicity achieves the best R^2 and MSE for both normally-distributed and t-distributed innovations for all the forecasting horizons when NASDAQ data is employed. However, the MAE for the short horizons ranging from 1 to 5 by using the modified day-variant periodicity are slight worser than the results from using day-invariant periodicity. Nevertheless, as people always place more emphasis on MSE instead of MAE for performance evaluation purpose, the modified day-variant periodicity can be considered a better candidate as a whole.

Table 4.6: Out-of-sample fit for GARCH(1,1) - NASDAQ

NASDAQ		normally-distributed innovations					t-distributed innovations				
Horizon	Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
1	Modified (Day-variant)	0.006	0.726	**0.1830**	**0.01081**	0.07815	0.010	0.690	**0.1835**	**0.01100**	0.07861
	Original (Day-variant)	0.008	0.688	0.1791	0.01127	0.08026	0.011	0.667	0.1820	0.01134	0.08016
	Day-invariant	0.009	0.707	0.1816	0.01089	**0.07814**	0.013	0.673	0.1822	0.01108	**0.07856**
3	Modified (Day-variant)	0.016	0.830	**0.3503**	**0.01565**	**0.09045**	0.025	0.785	**0.3517**	**0.01599**	**0.09120**
	Original (Day-variant)	0.021	0.785	0.3413	0.01654	0.09378	0.028	0.758	0.3477	0.01664	0.09360
	Day-invariant	0.021	0.808	0.3468	0.01585	0.09054	0.030	0.765	0.3485	0.01622	0.09123
5	Modified (Day-variant)	0.024	0.857	**0.4187**	**0.01938**	**0.09920**	0.037	0.806	**0.4205**	**0.01987**	**0.10017**
	Original (Day-variant)	0.030	0.808	0.4067	0.02065	0.10354	0.040	0.777	0.4149	0.02080	0.10329
	Day-invariant	0.031	0.833	0.4144	0.01970	0.09937	0.043	0.785	0.4166	0.02023	0.10028
10	Modified (Day-variant)	0.036	0.888	**0.4781**	**0.02801**	**0.11748**	0.059	0.824	**0.4810**	**0.02879**	**0.11897**
	Original (Day-variant)	0.047	0.834	0.4621	0.03002	0.12333	0.064	0.792	0.4735	0.03031	0.12331
	Day-invariant	0.046	0.864	0.4743	0.02845	0.11763	0.068	0.804	0.4777	0.02930	0.11897
15	Modified (Day-variant)	0.040	0.918	**0.4976**	**0.03562**	**0.13237**	0.074	0.841	**0.5017**	**0.03644**	**0.13393**
	Original (Day-variant)	0.054	0.859	0.4790	0.03807	0.13953	0.079	0.808	0.4936	0.03829	0.13906
	Day-invariant	0.050	0.897	0.4949	0.03601	0.13261	0.082	0.825	0.4995	0.03693	0.13408
25	Modified (Day-variant)	0.028	0.985	**0.5011**	**0.05028**	**0.15675**	0.087	0.880	**0.5078**	**0.05024**	**0.15733**
	Original (Day-variant)	0.054	0.911	0.4774	0.05285	0.16516	0.094	0.844	0.4984	0.05218	0.16342
	Day-invariant	0.042	0.962	0.4983	0.05055	0.15708	0.097	0.865	0.5058	0.05073	0.15755

Table 4.7: Out-of-sample fit for GARCH(1,1) - S&P 500

S&P 500

		normally-distributed innovations					t-distributed innovations				
Horizon	Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
1	Modified (Day-variant)	0.008	0.717	0.2178	**0.01137**	**0.07690**	0.011	0.693	**0.2182**	**0.01155**	**0.07739**
	Original (Day-variant)	0.014	0.651	0.2050	0.01220	0.07976	0.015	0.639	0.2077	0.01227	0.07974
	Day-invariant	0.009	0.714	**0.2180**	0.01141	0.07708	0.011	0.691	0.2181	0.01159	0.07758
3	Modified (Day-variant)	0.022	0.818	**0.3937**	**0.01681**	**0.08863**	0.027	0.788	**0.3949**	**0.01711**	**0.08939**
	Original (Day-variant)	0.031	0.751	0.3743	0.01835	0.09299	0.034	0.736	0.3801	0.01843	0.09286
	Day-invariant	0.023	0.812	0.3903	0.01696	0.08908	0.028	0.784	0.3915	0.01726	0.08981
5	Modified (Day-variant)	0.030	0.849	**0.4668**	**0.02089**	**0.09712**	0.037	0.817	**0.4686**	**0.02127**	**0.09813**
	Original (Day-variant)	0.041	0.783	0.4443	0.02299	0.10247	0.046	0.765	0.4519	0.02305	0.10232
	Day-invariant	0.031	0.842	0.4605	0.02120	0.09762	0.039	0.810	0.4623	0.02159	0.09859
10	Modified (Day-variant)	0.046	0.880	**0.5321**	**0.03043**	**0.11592**	0.058	0.841	**0.5359**	**0.03086**	**0.11702**
	Original (Day-variant)	0.061	0.814	0.5055	0.03362	0.12281	0.069	0.792	0.5167	0.03354	0.12217
	Day-invariant	0.049	0.873	0.5238	0.03103	0.11667	0.061	0.835	0.5278	0.03147	0.11775
15	Modified (Day-variant)	0.061	0.896	**0.5512**	**0.03950**	**0.13259**	0.078	0.851	**0.5567**	**0.03980**	**0.13355**
	Original (Day-variant)	0.077	0.833	0.5244	0.04311	0.13965	0.090	0.807	0.5387	0.04273	0.13848
	Day-invariant	0.064	0.891	0.5422	0.04034	0.13364	0.081	0.846	0.5481	0.04063	0.13450
25	Modified (Day-variant)	0.093	0.915	**0.5538**	**0.05780**	**0.16267**	0.119	0.860	**0.5643**	**0.05711**	**0.16213**
	Original (Day-variant)	0.109	0.855	0.5271	0.06120	0.16854	0.129	0.825	0.5492	0.05962	0.16570
	Day-invariant	0.094	0.914	0.5445	0.05904	0.16409	0.122	0.859	0.5560	0.05823	0.16343

The forecasting performance on S&P 500 data tabulated in Table 4.9 also reflects the superiority of the modified day-variant periodicity. Among the 12 distinct situations with different forecasting horizons and distributional assumptions, the modified day-variant periodicity achieves the highest R^2 for 10 times and lowest MSE and MAE for all 12 times.

Overall, the modified day-variant periodicity tends to produce better out-of-sample R^2 and MSE for various forecasting horizons (i.e. 1 to 25 periods). The above conclusion is valid for both GARCH(1,1) and FIGARCH(1,d,1) structure with either normally or t-distributed innovations. The superiority of the modified day-variant periodicity is demonstrated.

Table 4.8: Out-of-sample fit for FIGARCH(1,d,1) - NASDAQ

NASDAQ

		normally-distributed innovations					t-distributed innovations				
Horizon	Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
1	Modified (Day-variant)	0.005	0.729	**0.1858**	**0.01084**	0.07853	0.009	0.694	**0.1867**	**0.01102**	0.07901
	Original (Day-variant)	0.008	0.690	0.1821	0.01121	0.08002	0.011	0.669	0.1851	0.01130	0.08007
	Day-invariant	0.007	0.714	0.1847	0.01090	**0.07848**	0.011	0.679	0.1854	0.01109	**0.07893**
3	Modified (Day-variant)	0.012	0.841	**0.3582**	**0.01552**	0.09063	0.022	0.793	**0.3594**	**0.01590**	0.09159
	Original (Day-variant)	0.020	0.792	0.3487	0.01629	0.09314	0.027	0.761	0.3547	0.01648	0.09339
	Day-invariant	0.017	0.821	0.3550	0.01569	**0.09061**	0.026	0.774	0.3561	0.01610	**0.09148**
5	Modified (Day-variant)	0.016	0.874	**0.4297**	**0.01905**	0.09906	0.031	0.817	**0.4309**	**0.01963**	0.10041
	Original (Day-variant)	0.028	0.820	0.4165	0.02020	0.10247	0.038	0.782	0.4239	0.02052	0.10293
	Day-invariant	0.023	0.852	0.4254	0.01932	**0.09905**	0.037	0.797	0.4266	0.01995	**0.10037**
10	Modified (Day-variant)	0.020	0.917	**0.4962**	**0.02699**	**0.11628**	0.045	0.843	**0.4966**	**0.02800**	0.11869
	Original (Day-variant)	0.040	0.851	0.4770	0.02899	0.12145	0.057	0.802	0.4861	0.02966	0.12273
	Day-invariant	0.029	0.895	0.4917	0.02738	0.11639	0.053	0.825	0.4926	0.02847	**0.11862**
15	Modified (Day-variant)	0.016	0.952	**0.5224**	**0.03378**	**0.13005**	0.051	0.866	**0.5233**	**0.03490**	**0.13310**
	Original (Day-variant)	0.046	0.876	0.4986	0.03646	0.13649	0.069	0.819	0.5102	0.03722	0.13815
	Day-invariant	0.025	0.933	0.5188	0.03414	0.13033	0.059	0.851	0.5203	0.03533	0.13321
25	Modified (Day-variant)	-0.007	1.019	**0.5406**	**0.04604**	**0.15174**	0.050	0.909	**0.5428**	**0.04651**	**0.15463**
	Original (Day-variant)	0.045	0.920	0.5077	0.04978	0.16041	0.080	0.850	0.5246	0.04994	0.16170
	Day-invariant	0.004	1.000	0.5363	0.04644	0.15247	0.059	0.896	0.5395	0.04703	0.15519

Table 4.9: Out-of-sample fit for FIGARCH(1,d,1) - S&P 500

S&P 500

		normally-distributed innovations					t-distributed innovations				
Horizon	Normalization	b_0	b_1	R^2	MSE	MAE	b_0	b_1	R^2	MSE	MAE
1	Modified (Day-variant)	0.008	0.708	0.2185	**0.01155**	**0.07791**	0.012	0.668	0.2193	**0.01193**	**0.07900**
	Original (Day-variant)	0.016	0.627	0.2021	0.01252	0.08076	0.017	0.614	0.2106	0.01272	0.08139
	Day-invariant	0.008	0.705	**0.2193**	0.01158	0.07807	0.012	0.662	**0.2198**	0.01201	0.07934
3	Modified (Day-variant)	0.020	0.811	**0.3965**	**0.01700**	**0.09009**	0.029	0.758	**0.3977**	**0.01774**	**0.09182**
	Original (Day-variant)	0.035	0.721	0.3693	0.01909	0.09497	0.037	0.703	0.3853	0.01931	0.09579
	Day-invariant	0.021	0.806	0.3941	0.01712	0.09058	0.030	0.750	0.3955	0.01796	0.09252
5	Modified (Day-variant)	0.025	0.846	**0.4719**	**0.02099**	**0.09906**	0.039	0.784	**0.4728**	**0.02208**	**0.10153**
	Original (Day-variant)	0.047	0.749	0.4383	0.02414	0.10534	0.049	0.727	0.4585	0.02437	0.10657
	Day-invariant	0.027	0.839	0.4667	0.02125	0.09951	0.041	0.774	0.4680	0.02248	0.10215
10	Modified (Day-variant)	0.034	0.884	**0.5424**	**0.02999**	**0.11805**	0.058	0.806	**0.5427**	**0.03194**	**0.12190**
	Original (Day-variant)	0.070	0.773	0.4981	0.03578	0.12713	0.075	0.745	0.5251	0.03603	0.12906
	Day-invariant	0.037	0.878	0.5360	0.03045	0.11850	0.061	0.796	0.5372	0.03260	0.12268
15	Modified (Day-variant)	0.041	0.905	**0.5670**	**0.03806**	**0.13393**	0.075	0.814	**0.5664**	**0.04078**	**0.13901**
	Original (Day-variant)	0.089	0.784	0.5176	0.04605	0.14444	0.096	0.751	0.5489	0.04627	0.14728
	Day-invariant	0.044	0.901	0.5607	0.03865	0.13445	0.079	0.805	0.5612	0.04157	0.13999
25	Modified (Day-variant)	0.057	0.927	**0.5813**	**0.05289**	**0.16084**	0.110	0.815	**0.5804**	**0.05668**	**0.16651**
	Original (Day-variant)	0.127	0.791	0.5253	0.06448	0.17325	0.138	0.750	0.5638	0.06448	0.17658
	Day-invariant	0.059	0.926	0.5759	0.05360	0.16154	0.113	0.811	0.5769	0.05738	0.16729

4.7 Concluding remarks

The limitations of direct application of Andersen & Bollerslev's method for modeling day-variant intraday periodicity are addressed in this paper. Firstly, the original method does not consider the potential conflict appearing in the normalization procedure. It is shown that the resultant periodicity violates the implicit constraint, which is derived from the initial modeling assumption, in some situations. As a result, the overall intraday variances cannot be aggregated to form reasonable substitutes for their corresponding daily variances.

Secondly, the original normalization procedure is shown to be susceptible to heteroskedastic errors. It is demonstrated that the time series of the expected periodicity is deviated from its true value when the approximated series is contaminated with heteroskedastic errors. The conjecture of the presence of heteroskedastic errors is supported empirically. The approximation errors of the smoothing function are demonstrated to have non-constant variances. The Bartlett's Test and the Brown-Forsythe's Test significantly rejected the null hypothesis of constant error variance at 5% significance. As the approximated results are used as the substitute of the true underlying dynamics, it is reasonable to assume the approximated series to be deviated from its true value with heteroskedastic errors instead of homoskedastic errors.

To formulate a better day-variant intraday periodicity, a modified normalization procedure that ensures the fulfillment of the implicit constraint is proposed. The modified procedure regulates the magnitude of the periodicities with reference to the size of their corresponding daily variances, guaranteeing that the coherence of the intraday variances and their aggregated results is met. For the robustness on heteroskedastic errors, the modified procedure is demonstrated to outperform the original normalization through numerical simulations. The time series of the expected periodicity from the modified method match the underlying true series closer with higher R^2 than the original normalization does under various heteroskedastic situations.

On the other hand, the application of the modified periodicity in volatility modeling is also investigated in this study. Our method is tested with GARCH(1,1) and FIGARCH(1,d,1) structures by using 10-minute returns of the NASDAQ index (3-year period) and the S&P 500 index (2.5-year period). Both normally distributed and t-distributed innovations are considered in our investigation. The performance measures include regression R^2, mean squared error and mean absolute error. Apart from studying the 1-step-ahead forecast performance, several multiple-step-ahead forecasting results (up to a 25-step-ahead forecast) are also involved.

By comparing with the original day-variant intraday periodicity and the day-invariant version, our proposed day-variant periodicity is demonstrated to give superior performance in both in-sample fitness and out-of-sample forecast. The proposed method is always shown to produce the highest R^2 and the lowest MSE. Our findings indicates that the allowance of the periodicity to be day-variant can help to improve the modeling accuracy of an intraday volatility process as a whole.

4.8 Supplementary note

4.8.1 Equivalent modeling performance by scaled intraday periodicity series

Based on the application of intraday periodicity for modeling an intraday volatility process, it can be demonstrated that the use of a scaled intraday periodicity series produces the same overall performance as the original series does.

Let $s_{t,n}$ be the original intraday periodicity series and $s'_{t,n}$ be the scaled version:

$$s'_{t,n} = C \cdot s_{t,n} \tag{4.31}$$

where C is a real number constant.

Consider the deseasonalized series $\tilde{r}'_{t,n}$ obtained with the use of the scaled series $s'_{t,n}$.

$$\begin{aligned}\tilde{r}'_{t,n} &= \frac{r_{t,n} - E(r_{t,n})}{s'_{t,n}} \\ &= \frac{r_{t,n} - E(r_{t,n})}{C \cdot s_{t,n}} \\ &= \frac{1}{C} \cdot \frac{r_{t,n} - E(r_{t,n})}{s_{t,n}} \\ &= \frac{1}{C} \cdot \tilde{r}_{t,n}\end{aligned} \tag{4.32}$$

The $\tilde{r}_{t,n}$ equals to the series deseasonalized by the original series $s_{t,n}$. It can be observed that the deseasonalized series $\tilde{r}'_{t,n}$ can be expressed in terms of

$\tilde{r}_{t,n}$.

Let $\tilde{\sigma}_{t,n}$ be the ARCH volatility of the series $\tilde{r}_{t,n}$. Based on the above equation, the volatility of the deseasonalized series $\tilde{r}'_{t,n}$ can be written as:

$$\tilde{\sigma}'_{t,n} = \frac{1}{C} \cdot \tilde{\sigma}_{t,n} \tag{4.33}$$

The $\tilde{\sigma}'_{t,n}$ denotes the conditional volatility of the series $\tilde{r}'_{t,n}$.

Since the volatility of an intraday return $\sigma_{t,n}$ is defined as a multiple of the intraday periodicity and the volatility of the corresponding deseasonalized series, the $\sigma_{t,n}$ can be written as:

$$\begin{aligned} \sigma_{t,n} &= s'_{t,n} \cdot \tilde{\sigma}'_{t,n} \\ &= (C \cdot s_{t,n}) \cdot (\frac{1}{C} \cdot \tilde{\sigma}_{t,n}) \\ &= s_{t,n} \cdot \tilde{\sigma}_{t,n} \end{aligned} \tag{4.34}$$

The second line in the above equation set is supported by the equations 4.31 and 4.33. This demonstrates the two normalizations give identical results for the situation that the intraday periodicity is non day-varying. When the periodicity varies along days, the ratio C will no longer be constant for all days and its influence cannot be offset.

4.8.2 Characteristics of the two normalization procedures for evaluating day-invariant periodicity

The Andersen & Bollerslev's normalization procedure and the modified normalization procedure give different influence on modeling the intraday volatility process when the intraday periodicity $s_{t,n}$ is allowed to be changed along days. However, they give identical performance for the case that $s_{t,n}$ does not vary along days (i.e. $s_{t_1,n} = s_{t_2,n}$ for all possible t_1, t_2 within the sample data). The following paragraphs elaborate the concept.

Bearing in mind that the Andersen & Bollerslev's normalization procedure is defined as:

$$s_{t,n} = \frac{S \cdot exp(\frac{\hat{x}_{t,n}}{2})}{\sum_{i=1}^{S/N} \sum_{j=1}^{N} exp(\frac{\hat{x}_{i,j}}{2})} \tag{4.35}$$

where S is the sample size of the intraday return series. $\hat{x}_{i,j}$ is the approximated intraday dynamics of the return series. It can be re-written as:

$$s_{t,n} = k \cdot exp(\frac{\hat{x}_{t,n}}{2}) \tag{4.36}$$

where $k = S / \sum_{i=1}^{S/N} \sum_{j=1}^{N} exp(\frac{\hat{x}_{i,j}}{2})$

On the other hand, the modified normalization procedure is:

$$s_{t,n} = \frac{\sqrt{N} \cdot exp(\frac{\hat{x_{t,n}}}{2})}{\left(\sum_{j=1}^{N} exp(\hat{x}_{t,j})\right)^{1/2}} \tag{4.37}$$

where N is the number of intraday interval per day.

For the situation that the intraday periodicity $s_{t,n}$ is restricted to be non day-varying, the $\hat{x}_{t,n}$ must be invariant across days.[6] Therefore, the denominator of the modified normalization becomes constant across days

[6]The order J of the approximating function $f(\Theta|\sigma_t, n)$ (Flexible Fourier form) in the equation 4.3 decides whether $s_{t,n}$ is allowed to be changed along the day index t. When $J \geq 1$, the $s_{t,n}$ depends on its corresponding daily volatility σ_t and the values of $s_{t,n}$ vary along the day index t. The condition of non day-varying $s_{t,n}$ implies $J = 0$.

and the modified normalization can be regarded as:

$$s_{t,n} \quad = \quad k' \cdot exp(\frac{\hat{x}_{t,n}}{2}) \tag{4.38}$$

where $k' = \sqrt{N}/\left(\sum_{j=1}^{N} exp(\hat{x}_{t_1,j})\right)^{1/2} = \sqrt{N}/\left(\sum_{j=1}^{N} exp(\hat{x}_{t_2,j})\right)^{1/2}$ for all t_1, t_2 in the data sample.

As a result, it can be observed that the values of intraday periodicity recovered from the two normalization methods only differ from each other in a constant ratio when the the $\hat{x}_{i,j}$ is day-invariant.

Using the result from the supplementary note 4.8.1, we know for sure that the two normalizations give equivalent performance on modeling an intraday volatility process when the $s_{t,n}$ is non day-varying.

□ **End of chapter.**

Chapter 5

Impact of the overnight information on daily volatility modeling

Summary

The overnight return is one kind of information that can reflect the volatility of the corresponding financial instrument. However, all volatility estimators, either based on range-based or high-frequency data, do not include this information in their formulations due to the incompatibility of the involved sampling interval.

In this study, we explore the possibility of incorporating the overnight variance indirectly through the use of linearly combined daily volatility estimators. Our empirical results demonstrate that the inclusion of overnight variance can produce substantial influence when the minimum-variance constraints are relaxed. . Besides, the influence is revealed to be not monotonic as an increase of the overnight proportion does not necessarily produce a larger influence. Furthermore, it is demonstrated that the inclusion of overnight variance can improve the prediction accuracy of the CBOE volatility indexes (VIX and VXD) under specific weight combinations. Our findings contradict the common perception that overnight return does not contain useful information for volatility modeling.

5.1 Introduction and the research objective

The overnight return, the return that results from the price difference between last market close and current market open, is one kind of information that can reflect the volatility of the corresponding financial instrument. However, not all volatility estimators make use of this information in their formulation. For example, the estimators proposed by Parkinson [64], Garman & Klass (σ_4^2) [41], Rogers & Satchell [66] and even the standard realized variance [10] do not incorporate overnight returns whereas estimators proposed by Garman & Klass (σ_6^2) [41], Yang & Zhang [73] and Hansen & Lunde [46] do. There is no consensus on which kind of estimators, with and without overnight return, can better capture the underlying volatility in stock market. Furthermore, some estimators are based on daily range quotes while some are base on high frequency quotes and this further complicates the problem about the appropriateness of incorporating overnight return in the formulation of volatility estimators.

One practical usage of volatility estimators is to set them as the information sources for volatility prediction. Popular volatility prediction models such as variants of ARCH/GARCH [19, 36, 35, 47] and RiskMetrics [60] are shown to satisfactorily capture the underlying dynamics. However, most of their variants are not suitable for handling non-negative time series as an information source. To properly handle the non-negativity, Engle proposed to model the error in the series as the multiple of the conditional mean estimates, adopting it as the mean equation in the GARCH framework [36].

This model is known as the Multiplicative Error Model (MEM) and it can provide consistent results for various distributions of error terms under its quasi maximum likelihood estimation method, making it robust to ambiguous error assumptions [52, 38].

Besides using econometric models to measure future volatility, there is another way to indicate the level of fluctuation in the future. The Chicago Board of Options Exchange (CBOE), the world's largest options exchange, has compiled volatility indexes by averaging the weighted prices of put and call options to measure the market expectation on future volatility[75]. The CBOE market based volatility index has been related to the model based conditional volatility in recent literatures. Blair et al. investigated the information content of VIX for the prediction of GARCH volatility and found out the current VIX value contains the richest information content for 1-step-ahead predicted realized volatility [18]. Besides, Engle & Gallo studied the possibility of using MEM volatilities to improve the prediction of VIX in 2006. They demonstrated multi-step average volatilities can be incorporated as statistically significant regressors in the auto-regression of VIX [38].

In this study, the impact of overnight information on volatility prediction is explored. We investigate the characteristics of MEM volatilities resulting from various degree of overnight component and assess their incremental information content accordingly. Our study aims to address the following issues: 1) To what extent overnight information affects MEM outcomes; 2) Whether range-based and high-frequency estimators behaves differently; 3) Whether the inclusion of overnight information provides additional information for predicting market-based volatilities. This study is divided in two phases to tackle these issues systematically. In the first phase, the relationships among the predicted conditional volatilities from range-based and high frequency estimators with and without overnight information are studied. In additional to the defined minimum-variance situations [41, 46], a generalized framework is proposed to broaden the investigation. Afterward, the incremental information content of the predicted volatilities is assessed by the improvements on the auto-regression of market based volatility indexes in the second phase.

Our work is organized as follows. Specifications of various volatility estimators are in Section 5.2. A brief introduction of Multiplicative Error Model (MEM) and the model settings for our study are in Section 5.3. Empirical investigations of impacts of overnight information are discussed in Section 5.4 and 5.5. Section 5.6 summarizes our findings.

5.2 Specifications of daily volatility estimators

Four estimators are selected as input sources for MEM models in our study. Garman and Klass's estimators are selected as range-based estimators to measure the volatility in active trading period ($\hat{\sigma}_4^2$) and whole day period ($\hat{\sigma}_6^2$) respectively. For high frequency estimators, Newey-West realized variance ($\hat{\sigma}_{NW}^2$) and Hansen and Lunde's whole-day-based minimum-variance realized variance ($\hat{\sigma}_{wholeRV}^2$) are chosen. Figure one illustrates the time spans of these estimators. Overnight information is embedded in both whole-day-based estimators, $\hat{\sigma}_6^2$ and $\hat{\sigma}_{wholeRV}^2$.

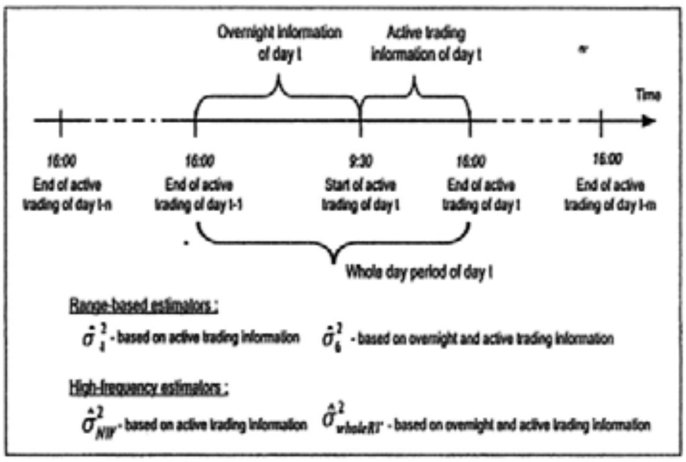


Figure 5.1: Illustration of the scope of four estimators

5.2.1 Garman & Klass's volatility estimators

Garman and Klass proposed a number of estimators in their seminal paper in 1980 [41]. Among the few estimators, there is one for estimating the volatility in active trading period and one for whole day period. The estimator for active trading period is formulated as follows.

$$\hat{\sigma}_4^2 = 0.511(u-d)^2 - 0.019[c(u+d) - 2ud] - 0.383c^2 \tag{5.1}$$

where C_t and O_t are the log values of the closing price and the opening price, H_t and L_t are the log values of the highest price and the lowest price for day t respectively. The normalized high, low and close are expressed as u, d and c individually, where $u = H_t - O_t$, $d = L_t - O_t$ and $c = C_t - O_t$.

The whole day version is formulated as :

$$\hat{\sigma}_6^2 = \frac{a}{f}(O_t - C_{t-1})^2 + \frac{1-a}{1-f}\hat{\sigma}_4^2 \tag{5.2}$$

where f is the fraction of that day that trading is closed and it is set to 1050/1440 in our study. a is a weight parameter and it is set to 0.12 to achieve minimum-variance property regardless of the value of f.

5.2.2 Realized variance and Hansen & Lunde's whole-day-based variance estimator

Standard Realized variance is the most well-known high frequency estimator to measure the volatility in an active trading period. Its formulation for day t is :

$$\hat{\sigma}_{RV}^2 = \sum_{i=1}^{m}\{(p(x_i) - p(x_i - \Delta)\}^2 \tag{5.3}$$

where x_i denotes the time, $p(x_i)$ is the price at time x_i, Δ is the sampling interval, m is the total number of intraday price recorded (excluding the opening price) in a day and $p(x_0)$ is the opening price of day t.

An important concern about the estimation of realized variance is whether the return series is autocorrelated or not [63]. The standard realized variance becomes biased when the returns are autocorrelated and the result should then be adjusted accordingly. The autocorrelation phenomenon always happen in high frequency data and there are various ways to offset the bias. Hansen & Lunde suggested an estimation method to handle the bias and regarded it as the Newey-West modified realized variance [45]. This method is based on Bartlett kernel and is guaranteed to be nonnegative. The Newey-West modified variance for day t is defined as:

$$\begin{aligned} \hat{\sigma}^2_{NW} &= \sum_{i=1}^{m} y_i^2 + 2\sum_{h=1}^{q}(1 - \frac{h}{q+1}) \sum_{i=1}^{m-h} y_i y_{i+h} \\ y_i &= p(x_i) - p(x_i - \Delta) \end{aligned} \tag{5.4}$$

the variable q represents the lag-length and it is set to $q = ceil(\frac{mw}{d})$ where w is the desired length of lag window and d is the total length of sampling period(trading period) in minutes[45]

For the whole day period, Hansen and Lunde defined an optimally combined whole day variance estimator in 2005 [46]. It is a minimum-variance estimate constructed by weighted squared overnight return and Newey-West modified realized variance. The following equations show the settings for

Hansen and Lunde's minimum-variance whole-day-based estimator ($\hat{\sigma}^2_{wholeRV}$).

$$\begin{aligned}
\hat{\sigma}^2_{wholeRV} &= \omega_1^* r_{1,t}^2 + \omega_2^* \hat{\sigma}^2_{NW,t} \\
\omega_1^* &= (1-\varphi)\frac{\mu_0}{\mu_1} \\
\omega_2^* &= \varphi\frac{\mu_0}{\mu_2} \\
\varphi &= \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}}
\end{aligned} \tag{5.5}$$

where ω_1^* and ω_2^* are the optimal weight for the min-variance estimator. $r_{1,t}^2$ is the squared overnight return. μ_0, μ_1 and μ_2 are the expected value of integrated variance, overnight variance and Newey-West realized variance respectively. η_1, η_2 and η_{12} are the variance of overnight variance and Newey-West realized variance and their covariance respectively.

5.3 Specification of Engle's Multiplicative Error Model

The way to employ volatility estimators as input series for GARCH type models is different from those for treating returns as the input series. Due to the non-negative nature of the volatility estimators, it is difficult to use traditional GARCH type models, which is based on linear formulation on return process, to estimate the model parameters as the variance and higher moments of the error distribution are unlikely to be constant [36]. Engle proposed an efficient way to model non-negative series in GARCH framework by treating the series as a composition of its conditional mean multiplied by a unit-mean error term. This multiplicative error structure is able to provide consistent results for error terms belong to a family of gamma distribution as the corresponding first order optimality conditions on the log-likelihood functions is the same.

The MEM (1,1) model is defined by the following two equations.

$$x_t = \mu_t \epsilon_t \qquad \text{Mean eqt.} \tag{5.6}$$

$$\mu_t = \omega + \alpha x_{t-1} + \beta \mu_{t-1} + c' z_{t-1} \qquad \text{Variance eqt.} \tag{5.7}$$

In the mean equation, x_t is the non-negative time series, μ_t is the conditional mean estimates and ϵ_t represents a unit-mean gamma-distributed i.i.d. error process. The variance equation is similar as that in the GARCH framework by replacing the error squared term with x_t in the ARCH term. Furthermore, exogenous variables are treated by including z_t in the variance equation.

The first order optimality condition for maximum likelihood estimation is:

$$\sum_{t=1}^{T}[-log(\mu_t(\theta)) - \frac{x_t}{\mu_t(\theta)}] \tag{5.8}$$

where θ is the parameter set $\{\alpha, \beta$ and $c'\}$ to be estimated and T is the size of training sample.

The MEM(1,1) framework is employed to predict the future volatility based on the most up-to-date volatility estimates. This framework has been successfully modeled the dynamics of non-negative volatility series, including range-based and high-frequency estimates, in some applications [38, 51, 50] and has been extended to multivariate cases recently [25]. We treat x_t in equations 5.6, 5.7 as a proxy to represent the input series(volatility estimates) for the model.

5.4 Correlation analysis among models with and without overnight information

Correlation analysis on the predicted conditional volatilities from MEM with and without overnight information are used to study the overnight impact. As stated in Section 5.2, $\hat{\sigma}_4^2$ and $\hat{\sigma}_{NW}^2$ are estimators without the consideration of overnight return in their formulation while $\hat{\sigma}_6^2$ and $\hat{\sigma}_{wholeRV}^2$ are estimators with overnight return.

To model the MEM 1-step-ahead conditional variances, we leave the exogenous term in equation 5.7 empty and replace the x_t in equation 5.6 by the estimators defined in equations 5.1- 5.5 correspondingly. Since the estimators represent variance for different time spans as stated in figure 5.1, their magnitudes are naturally different. Therefore, we employ the correlation coefficient as a scale-independent measure to evaluate the impact of overnight information. For instance, if all overnight returns equal to zero, the correlation will be 1. A small correlation demonstrates a large impact from overnight returns. The predicted conditional volatilities (i.e. square root of the conditional mean estimates) are named as *MEM*-σ_4, *MEM*-σ_6, *MEM*-σ_{NW}, and *MEM*-$\sigma_{wholeRV}$ respectively for the rest of this paper for clarity.

5.4.1 Generalized whole-day-based variance formulation

Observing that the minimum-variance estimators in equations 5.2 and 5.5 can be formulated as:

$$Var_t = w_1 On_t^2 + w_2 Act_t \tag{5.9}$$

where Var_t and Act_t represent variance estimates of whole day period and active trading period on day t respectively. w_1 and w_2 are weight factors and On_t is the overnight return. Var_t can then be considered as a linear combination of squared overnight return and variance estimate of active trading period.

The minimum-variance estimators, $\hat{\sigma}_6^2$ and $\hat{\sigma}_{wholeRV}^2$, can be considered as a particular weight combination. Scholars may use weights other than the minimum-variance weights to construct whole-day-based variance estimators. For example, Blair [18] and Gallo [40] use weights with w_1:w_2 = 1:1 to construct their estimators for whole day period. Sometimes, people may simply set $w_1 = 0$ as it is supposed that the magnitude of overnight variance is comparatively small and cannot give significant influence on the outcomes. To broaden our investigation of impact from overnight information, we try to consider all the possible combinations of positive w_1 and w_2 and formulate the estimator for whole day period as:

$$\begin{aligned} Var_t &= (w_1 + w_2)[\frac{w_1}{w_1 + w_2} On_t^2 + \frac{w_2}{w_1 + w_2} Act_t] \\ &= (w_1 + w_2)(\lambda On_t^2 + (1 - \lambda) Act_t) \end{aligned} \tag{5.10}$$

where λ is some positive scalar and $0 \leq \lambda \leq 1$. λ can be considered as a variable that governs the composition and its value indicates the influence of overnight information on a whole-day-based estimator. The portion $\lambda On_t^2 + (1-\lambda)Act_t$ represents a scaled version of Var_t.

As our MEM models do not have any exogenous variable, the scaling factor, $w_1 + w_2$, only adds a constant term, $Tlog(w_1 + w_2)$, to the likelihood function in equation 5.8. The optimal conditional variances for the original series therefore only differs the scaled version in a factor of $w_1 + w_2$. In other words, the value of conditional volatilities for the original series equals to those from scaled series multiply $\sqrt{w_1 + w_2}$. Since we use correlation to compare the overnight impact and linear regression to reflect the incremental information content of the predicted conditional volatilities, the properties of scaled form $\lambda On_t^2 + (1-\lambda)Act_t$ can therefore demonstrate the corresponding properties of the original series. Our generalized whole-day-based formulations are as follows.

$$general\ \hat{\sigma}_{4,t}^2 = \lambda On_t^2 + (1-\lambda)\hat{\sigma}_{4,t}^2 \tag{5.11}$$

$$general\ \hat{\sigma}_{RV,t}^2 = \lambda On_t^2 + (1-\lambda)\hat{\sigma}_{NW,t}^2 \tag{5.12}$$

where $general\ \hat{\sigma}_{4,t}^2$ and $general\ \hat{\sigma}_{RV,t}^2$ represent generalized variance estimators based on range information and high-frequency information respectively. Our refined study will use these two estimators to substitute the x_t in equations 5.6 and 5.7 for evaluating the impact of overnight return. Their corresponding predicted conditional volatilities are labeled as *MEM-general* σ_4 and *MEM-general* σ_{RV}.

5.4.2 Preliminary study on the empirical data

Two index data sets, S&P 500 and Dow Jones Industrial Average (DJIA), in the period of 13 March 2006 to 11 January 2008 (458 trading days) are used in our study. The index values are sampled every 10 minutes starting from 9:30 to 16:00 inclusively. In addition, the daily close of their corresponding volatility indexes, VIX and VXD, are used for the regression analysis to investigate the information content of the predicted volatilities.

Our data selection scheme for model training is based on a rolling sample approach. For instance, if the sample size is m, the first sample for model training will be the 1 to m observations, the second sample will be the 2 to m+1 observations and so on. In this study, the size of training sample is 3/5 of the total available data of each data set. The set of input samples are used to produce their corresponding 1-step-ahead predicted conditional volatilities. Afterwards, the correlation among the resultant predicted series are measured to reflect the impact of overnight information. The predicted series are used as sample data for the regression analysis of volatility indexes in the second phase study.

An important concern about the calculation of realized variance is whether the return series is autocorrelated or not. If the return series is autocorrelated, the calculated result should be adjusted to offset the bias caused by the autocorrelation. In our study, the realized variance is estimated with the use of 10 minute intraday returns. Figure 5.2 contains the autocorrelation plots of two return series, S&P 500 and DJIA, with 95% confidence bounds. It indicates that the first serial correlation coefficients of both series are nonzeros and the series of DJIA has some other nonzero coefficients in lag 3, 5, 9 and

10. To tackle the autocorrelation problem, we apply the Newey-West method in equation 5.4 to estimate the realized variances for this study. Following Hansen & Lunde's practice, we select the a fixed value for the desired length of lag window for bias adjustment [46]. The autocorrelation plot (figure 5.3) for 30 minute intraday returns shows the series are not suffered from the autocorrelation problem and therefore we choose the length of lag window to cover 30 minutes (w=30).

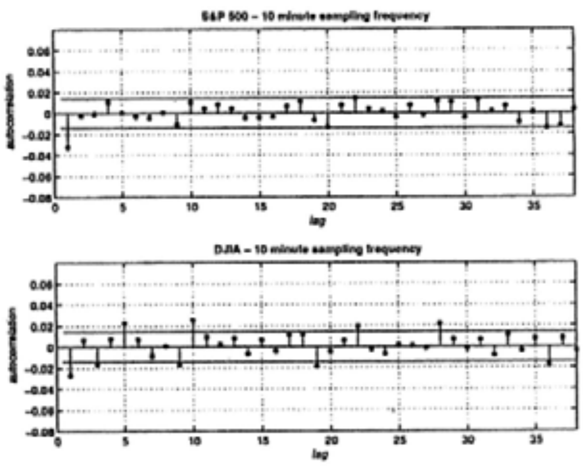


Figure 5.2: Autocorrelation of the 10 minute intraday data - S&P 500 and DJIA (with 95% confidence level bounds)

The descriptive statistics and correlation coefficients among squared overnight return (On^2), squared sigma 4 ($\hat{\sigma}_4^2$) and Newey-West realized variance ($\hat{\sigma}_{NW}^2$), are reported in Table 5.1.

The mean of On_t^2 is much smaller and its magnitude is less than 1/100 of those of $\hat{\sigma}_4^2$ or $\hat{\sigma}_{NW}^2$ for both S&P 500 and DJIA cases. Furthermore, the distribution of On^2 is more asymmetric and has thicker tails than the two variance series. Observation of a few large On_t^2 accounts for the high values of kurtosis. The low correlations of On^2 to either $\hat{\sigma}_4^2$ or $\hat{\sigma}_{NW}^2$ indicate that

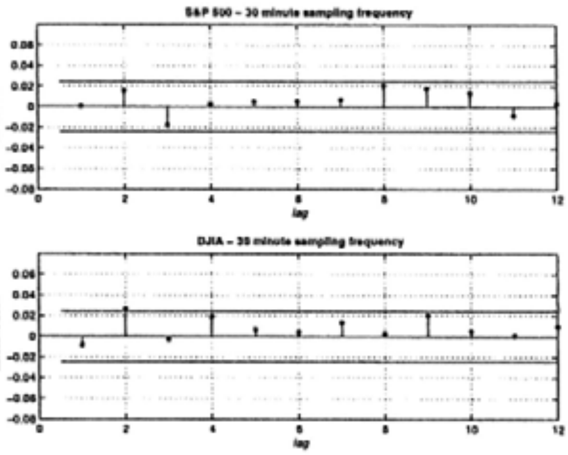


Figure 5.3: Autocorrelation of the 30 minute intraday data - S&P 500 and DJIA (with 95% confidence level bounds)

it contains information other than those in the two variance series. Besides, the correlation between $\hat{\sigma}_4^2$ and $\hat{\sigma}_{NW}^2$ is not very high (less than 0.8) and this indicates the two estimated variance series are different from each other.

5.4.3 Results from models under minimum-variance assumptions

As mentioned in Section 5.3, the estimators $\hat{\sigma}_6^2$ and $\hat{\sigma}_{wholeRV}^2$ consider overnight information in their formulation while $\hat{\sigma}_4^2$ and $\hat{\sigma}_{NW}^2$ do not. It is interesting to know whether the inclusion of overnight information in the estimators gives noticeable impacts on the predicted volatilities. MEM volatilities without overnight information (MEM-σ_4, MEM-σ_{NW}) are considered as the corresponding base series for comparison purpose. The correlation (ρ) between the base series and the testing series, which contains overnight information, are calculated to reflect the impact. The value of ρ will be 1 if the overnight returns are zeros and as a result, a large correlation demonstrates a small

S&P 500			
	On^2	$\hat{\sigma}_4^2$	$\hat{\sigma}_{NW}^2$
Mean	6.74E-07	4.55E-05	5.21E-05
Median	1.18E-09	2.58E-05	2.78E-05
Max	4.49E-05	5.33E-04	6.29E-04
Min	0	2.00E-06	3.13E-06
S.D.	3.49E-06	5.97E-05	6.54E-05
Skewness	9.07	3.86	3.31
Kurtosis	97.65	23.28	19.85
Correlation			
with On^2		0.245	0.207
with $\hat{\sigma}_4^2$			0.847
DJIA			
	On^2	$\hat{\sigma}_4^2$	$\hat{\sigma}_{NW}^2$
Mean	7.69E-07	1.31E-04	5.23E-05
Median	9.20E-09	1.04E-04	3.00E-05
Max	5.75E-05	9.09E-04	5.45E-04
Min	0	4.90E-06	2.79E-06
S.D.	3.86E-06	9.96E-05	6.10E-05
Skewness	9.24	2.86	3.10
Kurtosis	115.10	15.69	16.99
Correlation			
with On^2		-0.147	-0.009
with $\hat{\sigma}_4^2$			0.794

Table 5.1: Descriptive statistics of squared overnight return (On_t^2), Squared Sigma 4 ($\hat{\sigma}_4^2$) and Newey-West realized variance ($\hat{\sigma}_{NW}^2$)

impact.

	ARCH(2)	Q(12)
S&P 500	mean	mean
MEM-σ_4	0.476	8.529
MEM-σ_6	0.476	8.545
MEM-σ_{NW}	0.885	7.139
MEM-$\sigma_{wholeRV}$	0.885	7.139
	ARCH(2)	Q(12)
DJIA	mean	mean
MEM-σ_4	0.368	13.415
MEM-σ_6	0.368	13.439
MEM-σ_{NW}	0.994	7.095
MEM-$\sigma_{wholeRV}$	0.994	7.095

Table 5.2: Diagnostics information for the MEM models

The model fitness is checked by ARCH test on the standardized residuals and Ljung-Box test on the squared standardized residuals. Since rolling sample approach is used in this study, there are 184 MEM structures for a single input estimator and we propose to access them by the average values of the test statistics to avoid the abnormality caused by few specific samples with highly fluctuating values. The results of ARCH(2) test (5% critical value=5.99) and Q(12) (5% critical value=21.03), are reported in table 5.2 and no major specification problems are signaled by the diagnostics.

The impact of overnight information is assessed by the correlation coefficient between models with and without overnight variance. For range based estimators, the correlation coefficients between MEM-σ_4 and MEM-σ_6 are 1.000 for both S&P 500 and DJIA. It can be explained by the highly contrasting weight ratio in the formulation $\hat{\sigma}_6^2$. Under the minimum-variance situation, $\hat{\sigma}_6^2$ is calculated as $0.165^*On_t^2 + 3.249^*\hat{\sigma}_4^2$ and the ratios of the mean of On^2 to $\hat{\sigma}_6^2$ are $1:68$ for S&P 500 and $1:170$ for DJIA.[1] As a result, including On^2 in the formulation gives an extremely small impact to $\hat{\sigma}_6^2$ and thus the MEM-σ_6 should be highly correlated with MEM-σ_4.

On the other hand, for both S&P 500 and DJIA dataset, the correlation coefficients between MEM-σ_{NW} and MEM-$\sigma_{wholeRV}$ are also equal to 1.000. The weight ratios between the mean of On^2 to $\hat{\sigma}_{NW}^2$ are $1:77$ and $1:68$ for S&P 500 and DJIA respectively. [2] These results are different from Hansen & Lunde's finding [46] as the magnitude of overnight variance of market index is much smaller than those appeared in individual listed stocks. With such high correlation among the MEM volatilities, it is concluded that the effect of overnight information in either Garman's or Hansen's whole-day-based

[1]Elaboration on the calculation of the weight ratio of $\hat{\sigma}_6^2$ is documented in the supplementary note 5.7.1.

[2]Elaboration on the calculation of the weight ratio of $\hat{\sigma}_{wholeRV}^2$ is documented in the supplementary note 5.7.2.

variance is minimal under the minimum-variance situation.

5.4.4 Results from models under the generalized whole-day-based formulation

The empirical results in the previous section show that overnight return square (On^2) cannot give a significant impact on the MEM. However, the composition of the involved estimators, $\hat{\sigma}_6^2$ and $\hat{\sigma}_{wholeRV}^2$, are constrained by the minimum variance situation and therefore other possible combinations have not yet been studied. The minimum variance assumption is relaxed and an investigation under the general situation specified in equations 5.11 and 5.12 is carried out. An experiment has been conducted by adjusting λ from 0 to 1 with 0.001 for each increment and the results are recorded accordingly.

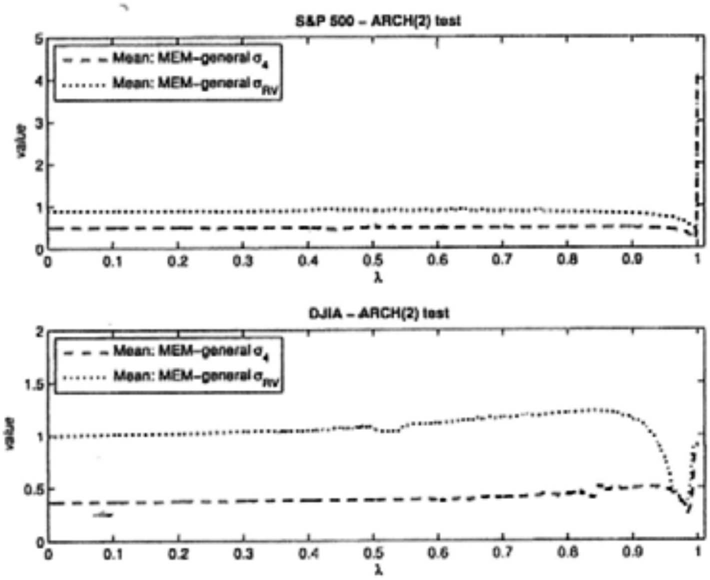


Figure 5.4: ARCH(2) model fit measure under the generalized weighted approach

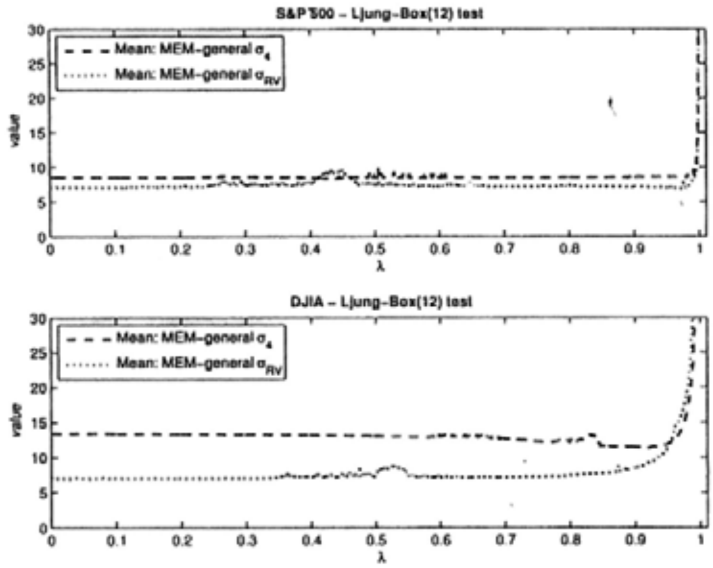


Figure 5.5: Ljung-Box(12) model fit measure under the generalized weighted approach

The model diagnostic information is visualized by figures 5.4 and 5.5. In general, MEM-general σ_4 tends to have smaller ARCH(2) values, which can be interpreted as better model fit, than MEM-general σ_{RV} over a large range of λ. However, Ljung-Box test gives an opposite result indicating MEM-general σ_{RV} always has a better fit. So, there is no consensus to differentiate which type of estimator has a better model fit. For the validation of MEM, all the mean values of ARCH(2) statistics for either MEM-general σ_4 or MEM-general σ_{RV} are smaller than the critical value (5.99) for all possible λ. For Ljung-Box(12) test, the means of MEM-general σ_4 and MEM-general σ_{RV} is larger than the critical value when $\lambda \geq$0.999 for S&P dataset. For DJIA data, the no autocorrelation hypotheses are rejected when $\lambda \geq$0.987 and $\lambda \geq$0.982 for the cases in MEM-general σ_4 and MEM-general σ_{RV} respectively. As a result, we conclude that the MEM models are valid for S&P 500 dataset under $0 \leq \lambda \leq 0.998$ and DJIA under $0 \leq \lambda \leq 0.981$.

The impact of the overnight return is reflected by the correlation graphs. Figure 5.6 depicts the correlation among the predicted volatilities for λ from 0 to 1 inclusively. To improve the visibility, the plot is zoomed to give the view for $\rho \geq 0.9$ in figure 5.7.

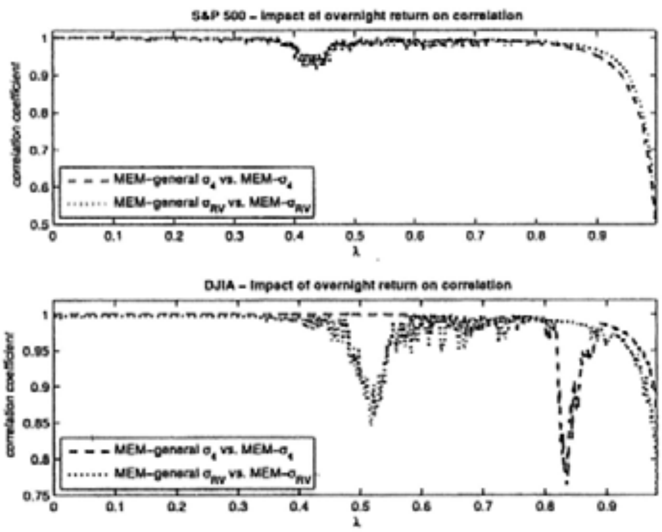


Figure 5.6: Comparison of correlation coefficient under the generalized weighted approach

For S&P 500, MEM-general σ_4 and MEM-general σ_{RV} give similar behavior. The correlation coefficients stay nearly constant (almost equal to 1) until λ reaches around 0.4 and decrease sharply to the local minimum. The ρ then bounds back to a high value and oscillates until λ attains a high value. Specifically, the ρ of MEM-general σ_4 decreases sharply from around λ=0.39 to the local minimum at λ=0.437 with ρ=0.917. The ρ increases and rebounds to the original level at λ=0.51. A similar U-shape pattern is also observed in MEM-general σ_{NW} but the region is shifted to $0.40 \leq \lambda \leq 0.52$ with the local minimum at λ=0.433 with ρ=0.927.

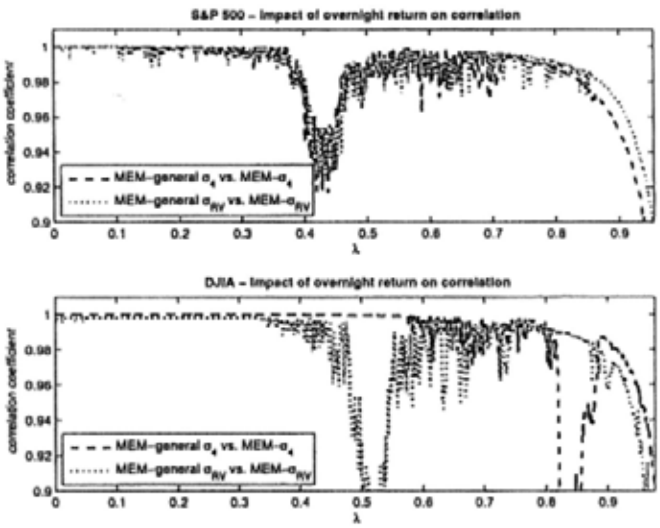


Figure 5.7: Correlation coefficient under the generalized weighted approach ($\rho \geq 0.9$)

For DJIA, the U-shape of MEM-general σ_4 appears around the range of 0.81$\leq \lambda \leq$0.88. The range for MEM-general σ_{RV} is around 0.48$\leq \lambda \leq$0.59. Their local minima are at λ=0.835 with ρ=0.765 and λ=0.517 with ρ=0.845 respectively. The shift of the U-shape can be explained by the difference of the magnitudes between the $\hat{\sigma}_4^2$ and $\hat{\sigma}_{NW}^2$.

To summarize, the appearance of a U-shape portion in every curve demonstrates the inclusion of overnight information can influence the characteristics of MEM volatility. However, the influence is not monotonic as an increase of overnight proportion does not necessary produce a larger change in correlation.

5.5 Modeling CBOE volatility index using MEM volatility

The information content of MEM predicted volatilities is assessed by their influence on the prediction of market based volatility indexes (VIX and VXD). VIX aims to capture the volatility of S&P 500 index while VXD aims at Dow Jones Industrial Average (DJIA) [75].

5.5.1 Regression models for VIX/VXD prediction

Under ideal efficient market hypothesis, risk neutral assumptions and the absence of insider information leakage, the model based predicted volatility and the market based volatility index should contain the same information content. However, these ideal situations is not valid for the real case and the information content of the two volatilities are usually different. Furthermore, their coverage on time horizon are different. The options used to computed the indexes last for around 30 calendar days covering both trading and non-trading overnight period [75] while the model based prediction may be based on estimators without any overnight component. As the market based volatility indexes cover the volatility for both overnight and active trading period, it is interesting to know whether the inclusion of overnight information in model based volatility can provide any additional predictive value. For the simplicity, the existence of any incremental information content of the overnight return is checked by simple linear regression models based on the ordinary least squares method. The general regression model is as follows :

$$\begin{aligned} MI_t &= \kappa + \gamma MI_{t-1} + \delta MemRange_t \\ &+ \zeta MemHighFreq_t \end{aligned} \tag{5.13}$$

MI is the market based index, such as VIX and VXD. κ, γ, δ and ζ are the regression coefficients, $MemRange_t$ and $MemHighFreq_t$ are the proxies that represent the MEM predicted volatilities for day t with various weight on overnight composition. The estimators in equations 5.1, 5.2 and 5.11 are substituted as the range-based proxies ($MemRange_t$) while 5.4, 5.5 and 5.12 are as high-frequency proxies ($MemHighFreq_t$) respectively.

By placing restrictions on certain parameters, we can define four different models.

1. Base model of AR(1) for the prediction of market based index by setting $\delta = \zeta = 0$
2. Model with the use of range-based information as an additional regressor: $\zeta = 0$
3. Model with the use of high-frequency information as an additional regressor: $\delta = 0$
4. Unrestricted model that incorporate both range-based and high-frequency information

The regression results are compared relatively to the base specification(i.e. simple AR(1) model). Adjusted R-square and F statistics are used as performance measures. Although the value of market indexes constructed by

CBOE represents volatilities in annual basis, the regression models are still valid as the model coefficients can absorb the scaling effect.

5.5.2 Results from models with and without overnight information

Observing that overnight information gives a noticeable impact in specific ranges, it is interesting to know whether the impact contains useful information or not. We assess the incremental information content by treating the MEM predicted conditional volatilities as exogenous variables to predict the market based volatility indexes (VIX and VXD). For the simplicity, simple linear AR(1) OLS regression models are used in this study. The base model defined in equation 5.13 is used as the baseline for comparison purposes. MEM volatilities which help to increase the adjusted R-square and pass the join zero coefficient test with 95% confidence (indication of the coefficients are non-zeros with 95% confidence) are considered to contain incremental information content in this study. The results of the adjusted R-square and F statistics are plotted in figures 5.8, 5.9 and 5.10.

For both VIX and VXD regressions, models with combination λ=0 are checked and none of them can produce a regression result with larger adjusted R-square than the base model and pass the F test at the same time. Therefore, we regard the MEM predicted volatilities without overnight information do not contain incremental information for market based indexes.

On the other hand, there are cases that give a better R-square and pass the F test when overnight information is embedded. Table 5.3 lists the combi-

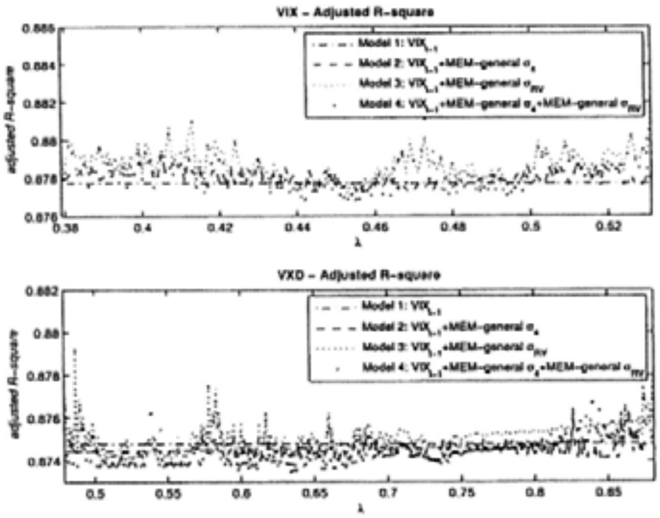


Figure 5.8: Adjusted R-square among the regression models

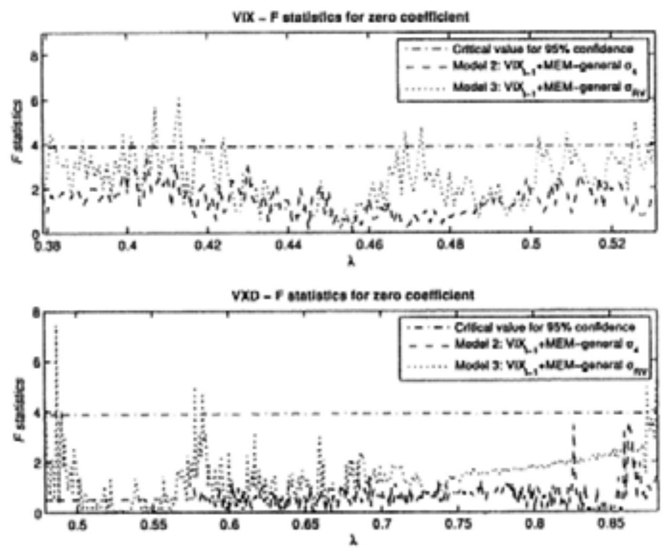


Figure 5.9: F statistics for testing zero coefficient

nations that contain incremental information for the regression. The results for VIX and VXD regressions are consistent. The inclusion of $MemRange_t$ cannot improve the prediction whereas $MemHighFreq_t$ can give favorable results for some combinations. In additional, the information content of $MemRange_t$ is not complementary to $MemHighFreq_t$ as models with the

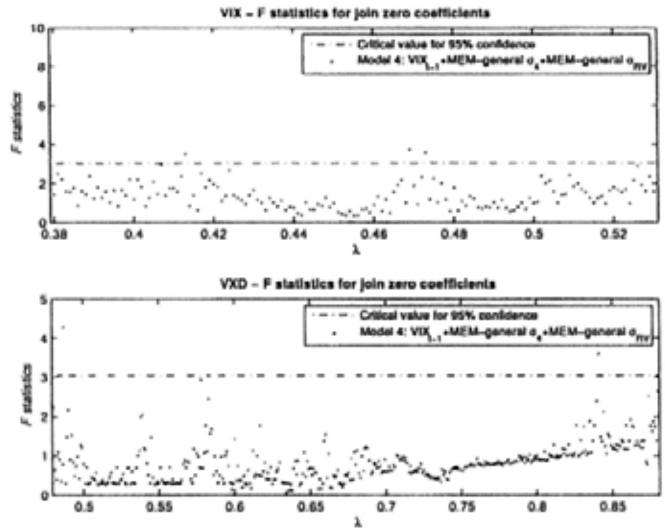


Figure 5.10: F statistics for testing join zero coefficients

use of both range-based and high-frequency information are always not as good as models that use $MemHighFreq_t$ alone. Our empirical results exemplify that overnight information can improve the prediction of the CBOE volatility indexes under specific combinations. However,

VIX regression		
Model 2	Model 3	Model 4
Nil	0.381	0.413
	0.382	0.469
	0.389	0.473
	0.399	
	0.401	
	0.407	
	0.412	
	0.413	
	0.419	
	0.424	
	0.469	
	0.473	
	0.502	
	0.509	
	0.526	
VXD regression		
Model 2	Model 3	Model 4
Nil	0.480	0.487
	0.487	0.841
	0.490	
	0.578	
	0.583	
	0.874	
	0.880	
	0.881	
	0.885	

Table 5.3: Combinations (λ) that give larger adjusted R-square and with F statistics larger than 95% confidence critical value

5.6 Concluding Remarks

The impact of overnight return on Engle's Multiplicative Error Model (MEM) is investigated in this study. Under minimum-variance situations, the overnight return has nearly no impact on models based on either Garman's or Hansen's whole-day-based estimators for both S&P 500 and DJIA data. When the minimum-variance conditions are relaxed, our general formulations demonstrate U-shape patterns in the correlation graphs among the predicted conditional volatilities with and without overnight information. However, the influence is not monotonic as an increase of the overnight proportion does not necessary produce a larger change in correlation.

On the other hand, our empirical results show that the inclusion of overnight information can improve the prediction of market based volatility indexes, VIX and VXD, under specific combinations. We demonstrate MEM volatilities resulting from linearly combined overnight variance and high frequency realized variance can help to improve the prediction of CBOE indexes under a simple linear AR(1) OLS regression. More sophisticated regression models may further exploit the hidden value from overnight information. The above findings contradict the common perception that overnight return does not contain useful information for volatility prediction.

5.7 Supplementary note

5.7.1 Analysis of the ratio of between On_t^2 and $\hat{\sigma}_4^2$

According to the specification of $\hat{\sigma}_6^2$ in the equation 5.2, the ratio of the weights between squared overnight return ($On_t^2 \equiv (O_t - C_{t-1})^2$) and $\hat{\sigma}_4^2$ can be deduced by the following approach:

Substituting the original formulation with $w_1 = \frac{a}{f}$ and $w_2 = \frac{1-a}{1-f}$:

$$\begin{aligned}\hat{\sigma}_6^2 &= a\frac{(O_t - C_{t-1})^2}{f} + (1-a)\frac{\hat{\sigma}_4^2}{(1-f)} \\ &= w_1(O_t - C_{t-1})^2 + w_2\hat{\sigma}_4^2\end{aligned}$$

The ratio $w_1 : w_2$ can be simplified as:

$$\begin{aligned} w_1 : w_2 &= \frac{a}{f} : \frac{1-a}{1-f} \\ &= 1 : 1 + \frac{f-a}{a(1-f)} \\ &= 1 : 1 + \frac{1-a}{a(1-f)} - \frac{1}{a} \end{aligned}$$

For min-variance estimator, a=0.12 and therefore $w_2 > 1$ when $\frac{1-a}{a(1-f)} - \frac{1}{a} > 0$. (i.e. $w_2 > 1$ when $f > 0.12$). Also, w_2 increases with the increase of f.

Since f represents the portion of market close and it equals to 0.729 (f=1050/1440) for US market which opens 6.5 hours a day, the ratio between w_1 and w_2 for composing $\hat{\sigma}_6^2$ will be 1 : 19.7. In addition, the ratio of $E(On_t^2)$:$E(\hat{\sigma}_4^2)$ are 1 : 68 and 1 : 170 for S&P 500 and DJIA data respectively ($E(x)$ denotes the mean of x). As a result, it is reasonable to observe that the overnight return cannot give any significant influence on the final value of $\hat{\sigma}_6^2$ and the correlations between $\hat{\sigma}_4^2$ and $\hat{\sigma}_6^2$ are very high (Our experiment data shows correlation over 0.9997).

5.7.2 Analysis of the ratio of between On_t^2 and $\hat{\sigma}_{NW,t}^2$

The specification of $\hat{\sigma}_{wholeRV}^2$ is:

$$\begin{aligned} \hat{\sigma}_{wholeRV}^2 &= \omega_1^* r_{1,t}^2 + \omega_2^* \hat{\sigma}_{NW,t}^2 \\ \omega_1^* &= (1-\varphi)\frac{\mu_0}{\mu_1} \\ \omega_2^* &= \varphi\frac{\mu_0}{\mu_2} \\ \varphi &= \frac{\mu_2^2\eta_1^2 - \mu_1\mu_2\eta_{12}}{\mu_2^2\eta_1^2 + \mu_1^2\eta_2^2 - 2\mu_1\mu_2\eta_{12}} \end{aligned}$$

The ratio of ω_1^* and ω_2^* are $1:77$ and $1:68$ for S&P 500 and DJIA data empirically. Furthermore, the magnitude of $E(On_t^2)/E(\hat{\sigma}^2_{NW,t})$ is less than $1/60$ for the two data sets. Therefore, the overnight return cannot affect the final value of $\hat{\sigma}^2_{wholeRV}$ much. The correlations between $\hat{\sigma}^2_{wholeRV}$ and $\hat{\sigma}^2_{NW,t}$ are equal to 1.00 for the two data sets.

□ **End of chapter.**

Chapter 6

Concluding remarks and further work

This research aims to explore the possibility of improving the existing practices in volatility modeling. There are three new suggestions in this study, namely 1) Utilization of the interaction effect between the intraday periodicity and the heteroskedasticity; 2) Modified normalization for the estimation of the intraday periodicity; and 3) Inclusion of overnight information for the estimation of daily volatility. The first two modifications address the limitations of the commonly used intraday volatility modeling approach - Andersen & Bollerslev's sequential estimation approach while the third modification is about the inclusion of overnight information for the estimation of daily volatility.

The first modification is based on the fact that Andersen & Bollerslev's sequential approach does not consider the interaction between the intraday periodicity and the ARCH process (heteroskedasticity). Their approach is mathematically shown to produce a sub-optimal result under a general situation.

Our proposed approach, the integrated framework approach, improves the subsequent ARCH structure in their method by integrating the filtration process and the ARCH process in a united setting and optimizing the model parameters for the raw series instead of the filtered series. The simultaneous presence of the periodicity and the ARCH volatility enables the consideration of their interaction while the use of the raw series assures the optimal fit to the target series. Besides, our framework can be re-written as a modified P-ARCH structure where the periodically varying autoregressive coefficients are represented as the product of the estimated periodicity times the ARCH parameters. (P-ARCH structure is well-known for its capability to handle the periodicity and the heteroskedasticity in a time series simultaneously.) Furthermore, the effectiveness of applying the integrated framework to different ARCH structures is also discussed in this work.

The second modification is about the estimation method of the intraday periodicity. Under Andersen & Bollerslev's method, the periodicity is estimated in two steps. The dynamics of an intraday return series is firstly approximated by a smoothing function (Flexible Fourier Form) and, secondly, the periodicity is recovered by a normalization procedure with the use of the approximation results. Their method is capable of defining the estimated periodicity to be either day-invariant or day-variant with proper adjustments. However, when the method is applied for day-variant situation, it is shown that the resultant periodicity violates the implicit constraint, which is derived from the initial modeling assumption, in some situations.

A modified normalization procedure that ensures the fulfillment of the implicit constraint is proposed in this research work. The new procedure adjusts the magnitude of the periodicity with reference to the size of its corresponding daily variance. When the proposed procedure is applied for the day-variant situation, the results turn out to be more robust to heteroskedastic errors under numeric simulations. For the day-invariant situation, our procedure is proved to give the same performance as the original normalization does mathematically.

The two proposed modifications on intraday volatility modeling are tested with different ARCH structures, including GARCH(1,1), FIGARCH(1,d,1) and HYGARCH(1,d,1), by using simulated data and market data. The performance measures include regression R^2, mean squared error, mean absolute error and Diebold-Mariano hypothesis tests on squared error and absolute error. Apart from studying ~~the~~ 1-step-ahead out-of-sample performance, several multiple-step-ahead forecasting results are also addressed. Under the same level of model flexibility (parameterized portions), our proposed modifications always outperform the original method in both in-sample fitness and out-of-sample performance on various forecasting horizons.

Early work of the two modifications has been submitted to the Journal of Applied Econometrics and reviewed by Tim Bollerslev, the original proposer of the sequential modeling approach. Although the work was rejected, Tim's responses show an interest in the first modification and an agreement on the uniqueness of the second modification. Finally, the revised writing of the first modification is accepted by the Journal of International Financial Markets, Institutions and Money, which published Andersen & Bollerslev's work on the application of their sequential method in 2000. On the other hand, the writing of the second modification was accepted for presentation in the 4th CSDA International Conference on Computational and Financial Econometrics (CFE 10) on Dec 2010.

The third modification is about the inclusion of overnight information for the estimation of daily volatility. The possibility of incorporating the overnight information is explored through the use of linearly combined daily volatility estimators. The empirical results demonstrate that the inclusion of overnight information can produce substantial influence when the minimum-variance constraints are relaxed. Besides, the influence is revealed to be not monotonic as an increase of the overnight proportion does not necessarily produce a larger influence. Furthermore, it is demonstrated that the inclusion of overnight variance can improve the prediction accuracy of the Chicago Board of options Exchange (CBOE) volatility indexes (VIX and VXD) under specific weight combinations. The findings contradict the common perception that overnight return does not contain useful information for daily volatility modeling.

The empirical findings of the third modification were initially presented in the International Symposium on Financial Engineering and Risk Management in 2007 and the revised writing is published in the Journal - 'Statistics and Its Interface' in 2008.

About the further work, it is considered that the second modification, Modified normalization for the estimation of the intraday periodicity, may be further developed. The version documented in this research is based on the key assumption that the innovations are independent of each other.

However, real situations may not follow this nice assumption and it would be good to have a method that is capable of handling correlated innovations. Some initial studies on this issue have already been carried out.

It is found out that the major obstacle for developing a workable procedure is about the mutual dependence of the periodicity and the modeled innovation. Since the two components are not directly observable, their values are estimated with the use of the other party's value and the correlation matrix of the modeled innovations. A simultaneous estimation of the two components together with the correlation matrix is infeasible due to the excessive computational requirement. Initial investigation suggests that a step-wise method may be used to provide a good approximated result for the problem. Additional effort is required for further development.

□ **End of chapter.**

Appendix A

Nomenclature

Summary

The following tables contain brief explanations of the symbols used in this thesis.

☐ **End of chapter.**

Table A.1: Nomenclature for Chapter 3 (Integrated framework approach)

$r_{t,n}$	Intraday return of day t in n-th intraday interval
$\xi_{t,n}$	Intraday innovation of day t in n-th intraday interval
$\sigma_{t,n}$	Intraday volatility (standard deviation) of day t in n-th intraday interval
$\bar{r}_{t,n}$	Deseasonalized return of day t in n-th intraday interval
$\bar{\epsilon}_{t,n}$	Deseasonalized innovation of day t in n-th intraday interval
$\bar{\sigma}_{t,n}$	Deseasonalized volatility (standard deviation) of day t in n-th intraday interval
T	Total number of days in the data sample
N	Number of intraday period per day
$\mathbf{S}$	Set that contains all the day and interval index pairs (t, n) for the in-sample data
r_t	Daily return of day t
ϵ_t	Daily innovation of day t
σ_t	Daily volatility of day t
$z_{t,n}$	Zero-mean independently and identically distributed (i.i.d.) standard normal random term
$s_{t,n}$	Intraday periodic component
$E(X)$	Unconditional expectation of the variable X
$f(\Theta\|n)$	Flexible Fourier form to approximate the U-shape periodic dynamics
θ	Parameter set for a ARCH framework
$D(x, \sigma_x^2)$	Independently and identically distributed (i.i.d.) distribution of a variable x with a variance σ_x^2
$N(x, \sigma_x^2)$	Standard normal distribution of a variable x with a variance σ_x^2
$\phi(\sigma\|x)$	Probability density function of a zero-mean variable x with its standard deviation equals to σ

Table A.2: Nomenclature for Chapter 4 (Day-varying structure)

$P_{t,n}$	Index value of day t in n-th intraday interval
$P_{t,0}$	Market open index value of day t
$r_{t,n}$	Intraday return of day t in n-th intraday interval
$\xi_{t,n}$	Intraday innovation of day t in n-th intraday interval
$\sigma_{t,n}$	Intraday volatility (standard deviation) of day t in n-th intraday interval
N	Number of intraday period per day
$\bar{\xi}_{t,n}$	Standardized intraday innovations for the n-th interval on day t
r_t	Daily return of day t
σ_t	Daily volatility of day t
$z_{t,n}$	Zero-mean independently and identically distributed (i.i.d.) standard normal random term
$u_{t,n}$	Zero-mean independently and identically distributed (i.i.d.) random term
$s_{t,n}$	Intraday periodic component
$x_{t,n}$	Intermediate proxy variable representing the intraday periodic dynamics
$E(X)$	Unconditional expectation of the variable X
$N(x, \sigma_x^2)$	Standard normal distribution of a variable x with a variance σ_x^2
$\phi(\delta_{i,j})$	Probability density function of the variable $\delta_{i,j}$
$f(\Theta\|\sigma_t, n)$	Flexible Fourier form to approximate the U-shape periodic dynamics

Table A.3: Nomenclature for Chapter 5 (Impact of the overnight information)

C_t	Log value of the closing price for day t
O_t	Log value of the opening price for day t
H_t	Log value of the highest price for day t
L_t	Log value of the lowest price for day t
u	Normalized high for day t, i.e. $H_t - O_t$
d	Normalized low for day t, i.e. $L_t - O_t$
c	Normalized close for day t, i.e. $C_t - O_t$
$\hat{\sigma}^2_{4,t}$	Garman & Klass's daily volatility estimator
$\hat{\sigma}^2_{NW,t}$	Newey-West modified realized variance estimator
On_t	Overnight return
general $\hat{\sigma}^2_{4,t}$	Proposed linearly combined range-based daily volatility estimator
general $\hat{\sigma}^2_{RV,t}$	Proposed linearly combined high-frequency daily volatility estimator
λ	Weight factor of the proposed linearly combined daily volatility estimator
MEM-general σ_4	Conditional volatility series resulting from the MEM model with general $\hat{\sigma}^2_{4,t}$ as input
MEM-general σ_{RV}	Conditional volatility series resulting from the MEM model with general $\hat{\sigma}^2_{RV,t}$ as input

Bibliography

[1] Admati Anat R, Pfleiderer Paul. 1988. A Theory of Intraday patterns: Volume and Price Variability. *The Review of Financial Studies* **1(1)** : 3-40.

[2] Andersen T G, Bollerslev Tim. 1996. Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns. *National Bureau of Economic Research, Cambridge, NBER Working Papers* **5752**.

[3] Andersen T G, Bollerslev Tim. 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* **4** : 115-158.

[4] Andersen T G, Bollerslev Tim. 1998. Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *The Journal of Finance* **53(1)** : 219-265.

[5] Andersen T G, Bollerslev Tim. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review, Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association* **39(4)** : 885-905.

[6] Andersen T G, Bollerslev Tim, Diebold FX, Labys P. 1999. Understanding, optimizing, using and forecasting realized volatility and correlation. *New York University, Leonard N. Stern School of Business, Finance Department Working Paper Series No.99-061.*

[7] Andersen T G, Bollerslev Tim, Lange Steve. 1999. Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance* **6(5)** : 457-477.

[8] Andersen T G, Bollerslev Tim, Cai J. 2000. Intraday and interday volatility in the Japanese stock market. *Journal of International Financial Markets, Institutions and Money* **10(2)** : 107-130.

[9] Andersen T G, Bollerslev Tim, Diebold Francis X, Labys Paul. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* **71(2)** : 579-625.

[10] Andersen T G, Bollerslev Tim, Christoffersen PF, Diebold FX. 2006. Volatility and correlation forecasting. In *Handbook of economic forecasting*, North-Holland : 777-878.

[11] Areal Nelson M P C, Taylor Stephen J. 2002. The realized volatility of FTSE-100 futures prices. *The Journal of Future Markets* **22(7)** : 627-648.

[12] Bai X, Russell J R, Tiao GC. 2003. Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* **114** : 349-360.

[13] Baillie Richard T, Bollerslev Tim, Mikkelsen Hans Ole. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74(1)** : 3-30.

[14] Bandi F M, Russell J R. 2004. Microstructure noise, realized volatility, and optimal sampling. *Econometric Society 2004 Latin American Meetings* Santiago, Chile.

[15] Barndorff-Nielsen Ole E, Shephard Neil. 2002. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* **17(5)** : 457-477.

[16] Bartlett M S. 1937. Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Statistical Society Series A* **160** : 268282.

[17] Beltratti Andrea, Morana Claudio. 1999. Computing value at risk with high frequency data. *Journal of Empirical Finance* **6(5)** : 431-455.

[18] Blair B J, Poon S H, Taylor S J. 2000. Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Index Returns. *Accounting and Finance Working Paper No. 99/014, Lancaster University Management School*

[19] Bollerslev Tim. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31** : 307-327.

[20] Bollerslev Tim. 1987. A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics* **69(3)** : 542-547.

[21] Bollerslev Tim, Ghysels Eric. 1996. Periodic Autoregressive Conditional Heteroscedasticity. *Journal of Business & Economic Statistics* **14(2)** : 139-151.

[22] Bollerslev Tim, Ole Mikkelsen Hans. 1996. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* **73(1)** : 151-184.

[23] Brown Morton B, Forsythe Alan B. 1974. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association* **69(346)** : 364-367.

[24] Chortareas Georgios E, Nankervis John, Jiang Ying. 2006. Forecasting Exchange Rate Volatility at High Frequency Data: Is the Euro Different ?. *Money Macro and Finance (MMF) Research Group Conference 2006 from Money Macro and Finance Research Group* **79**.

[25] Cipollini F, Engle R F, Gallo G M. 2007. A Model for Multivariate Non-negative Valued Processes in Financial Econometrics. *Econometrics Working Papers Archive, Universita' degli Studi di Firenze, Dipartimento di Statistica "G. Parenti"*.

[26] Chu CF, Lam KP. 2007. Impact of overnight information on MEM volatility prediction. *International Symposium on Financial Engineering and Risk Management 2007 (FERM 2007)*.

[27] Chu C F, Lam K P. 2008. Impact of overnight information on MEM volatility prediction. *Statistics and Its Interface* **1(2)** : 297306.

[28] Chu C F, Lam K P. 2010. Day-varying structure for modeling intraday periodicity. *4th CSDA International Conference on Computational and Financial Econometrics (CFE 10)*.

[29] Davidson J. 2004. Moment and Memory Properties of Linear Conditional Heteroskedasticity Models, and a New Model. *Journal of Business & Economic Statistics* **22(1)** : 16-29.

[30] Diebold Francis X, Mariano Roberto S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* **13(3)** :253-263.

[31] Engle R F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of Variance of United Kingdom Inflation. *Econometrica* **50(4)** : 987-1007.

[32] Engle R F, Bollerslev T. 1986. Modeling the Persistence of Conditional Variances. *Econometric Reviews* **5(1)** : 1-50.

[33] Engle R F, McFadden D L. 1994. ARCH models. In *Handbook of econometrics* **4**, North-Holland : 2959-3038.

[34] Engle R F, Patton Andrew J. 2001. What good is a volatility model ?. *Quantitative Finance* **1(2)** : 237-245.

[35] Engle R F. 2001. GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives* **15** : 157-168.

[36] Engle R F. 2002. New Frontiers for ARCH models. *Journal of Applied Econometrics* **17** : 425-446.

[37] Engle R F, Sokalska M. 2005. High Frequency Multiplicative Component GARCH. *NYU Working Paper : SC-CFE-05-05.* Available at SSRN: http://ssrn.com/abstract=1297097

[38] Engle R F, Gallo G M. 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* **131** : 3-27.

[39] Gallant Ronald A. 1984. The Fourier Flexible Form. *American Journal of Agricultural Economics* **66(2)** : 204-208.

[40] Gallo Giampiero M. 2001. Modelling the Impact of Overnight Surprises on Intra-Daily Volatility. *Australian Economic Papers* **40(4)** : 567-80.

[41] Garman M B, Klass M J. 1980. On the Estimation of Security Price Volatilities from Historical Data. *Journal of Business* **53** : 67-78.

[42] Giot Pierre. 1999. Time transformations, intraday data and volatility models. *Center for Operations Research and Econometrics, The Universit catholique de Louvain, Discussion Papers* **1999044**.

[43] Giot Pierre. 2005. Market risk models for intraday data. *European Journal of Finance* **11(4)** : 309-324.

[44] Hansen Bruce E. 1994. Autoregressive Conditional Density Estimation. *International Economic Review* **35** : 705-30.

[45] Hansen P R, Lunde A. 2003. An Optimal and Unbiased Measure of Realized Variance based on Intermittent High-frequency Data. *CIREQ-CIRANO Conference 2003*, Montreal.

[46] Hansen P R, Lunde A. 2005. A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data, *Journal of Financial Econometrics* **3** : 525-554.

[47] Hansen P R, Lunde A. 2005. A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, *Journal of Applied Econometrics* **20** : 873-889.

[48] Jarque Carlos M, Bera Anil K. 1987. A test for normality of observations and regression residuals. *International Statistical Review* **55(2)** : 163172.

[49] Kim Tae-Hwan, White Halbert. 2004. On more robust estimation of skewness and kurtosis. *Finance Research Letters* **1(1)** : 56-73.

[50] Lam K P, Ng H S. 2006. Intra-daily information of range-based volatility for MEM-GARCH, *International Conference on Time Series Econometrics, Finance, and Risk (TSEFAR)*

[51] Lanne Markku. 2006. A Mixture Multiplicative Error Model for Realized Volatility. *Journal of Financial Econometrics* **4(4)** : 594-616.

[52] Lee Sang-Won, Hansen Bruce E. 1994. Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory* **10** : 29-52.

[53] Martens Martin. 2001. Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance* **20(1)** : 1-23.

[54] Martens Martin, Chang Yuan-Chen, Taylor Stephen J. 2002. A comparison of seasonal adjustment methods when forecasting intraday volatility. *The Journal of Financial Research* **15(2)** : 283-299.

[55] McMillan David G, Speight Alan E. H. 2004. Intra-day periodicity, temporal aggregation and time-to-maturity in FTSE-100 index futures volatility. *Applied Financial Economics* **14(4)** : 253-263.

[56] McMillan David G, Speight Alan E. H. 2006. Heterogeneous information flows and intra-day volatility dynamics: evidence from the UK FTSE-100 stock index futures market. *Applied Financial Economics* **16(13)** : 959-972.

[57] McMillan David G, Garcia Raquel Quiroga. 2009. Intra-day volatility forecasts. *Applied Financial Economics* **19** : 611-623.

[58] Mehta Neelesh B, Molisch Andreas F, Wu Jingxian, Zhang Jin. 2007. Approximating a Sum of Random Variables with a Lognormal. *IEEE Trans. on Wireless Communications* **6(7)** : 2690-2699

[59] Mian Mujtaba Ç, Adam Christoper M. 2001. Volatility dynamics in high frequency financial data: an empirical investigation of the Australian equity returns. *Applied Financial Economics* **11(3)** : 341-352.

[60] J.P. Morgan/Reuters, *Part II: Statistics of financial market returns, in RiskMetrics technical document* (Fourth Edition, New York, 1996).

[61] Nelson D B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59(2)** : 347-370.

[62] Newey W K, West K D. 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **55** : 703-708.

[63] Oomen Roel C A. 2004. Modelling Realized Variance when Returns are Serially Correlated. University of Warwick. *Warwick Business School, Discussion Paper SP II 2004 - 11.*

[64] Parkinson M. 1980. The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business* **53** : 61-65.

[65] Poon Ser-Huang, Granger Clive W J. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* **41(2)** : 478-539.

[66] Rogers L C G, Satchell S E. 1991. Estimating Variance From High, Low and Closing Prices. *The Annals of Applied Probability* **1** : 504-512.

[67] Selcuk Faruk, Gencay Ramazan. 2006. Intraday dynamics of stock market returns and volatility. *Physica A: Statistical Mechanics and its Applications* **367** : 375-387.

[68] Shu J, Zhang J. 2006. Testing Range Estimators of Historical Volatility. *Journal of Futures Markets* **26** : 297-313.

[69] Taylor Stephen J, Xu Xinzhong. 1997. The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* **4(4)** : 317-340.

[70] Thompson Gary D. 1988. Choice of Flexible Functional Forms: Review and Appraisal. *Western Journal of Agricultural Economics* **13(2)** : 169-183.

[71] Wilhelmsson Anders. 2006. Garch forecasting performance under different distribution assumptions. *Journal of Forecasting* **25** : 561-578.

[72] Wood R A, McInish T H, Ord J K. 1985. An Investigation of Transactions Data for NYSE Stocks. *Journal of Finance* **40(3)** : 723-739.

[73] Yang D, Zhang Q. 2000. Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices. *Journal of Business* **73** : 477-491.

[74] Zhang J, Shu J. 2006. Testing Range Estimators of Historical Volatility, *Journal of Futures Markets* **26** : 297-313.

[75] Chicago Board Options Exchange, *Revised, more robust methodology for the VIX Index*, 2003.