國立交通大學

財務金融所

博 士 論 文

No. 02

財務市場波動及基金績效之計量分析

Essays on the Econometric Analysis of Financial Market Volatility and Mutual Fund Performance

研 究 生：孫而音

指導教授：鍾惠民 教授

中 華 民 國 九 十 六 年 十 月

財務市場波動及基金績效之計量分析

Essays on the Econometric Analysis of Financial Market Volatility and Mutual Fund Performance

研 究 生：孫而音　　Student：Erh-Yin Sun

指導教授：鍾惠民 教授　　Advisers：Dr. Huimin Chung

國 立 交 通 大 學
財務金融研究所
博 士 論 文

A Disseration
Submitted to Graduate Institute of Finance
College of Management
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Finance

October 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年十月

國　立　交　通　大　學

博碩士論文全文電子檔著作權授權書

（提供授權人裝訂於紙本論文書名頁之次頁用）

本授權書所授權之學位論文，為本人於國立交通大學 財務金融 所， 96 學年度第 1 學期取得博士學位之論文。

論文題目：財務市場波動及基金績效之計量分析

指導教授：鍾惠民 教授

■ 同意　□不同意

本人茲將本著作，以非專屬、無償授權國立交通大學與台灣聯合大學系統圖書館：基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學及台灣聯合大學系統圖書館得不限地域、時間與次數，以紙本、光碟或數位化等各種方法收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行線上檢索、閱覽、下載或列印。

論文全文上載網路公開之範圍及時間：	
本校及台灣聯合大學系統區域網路	■ 中華民國 103 年 1 月 1 日
校外網際網路	■ 中華民國 103 年 1 月 1 日

授 權 人：孫而音

親筆簽名：________________

中華民國　96　年　10　月　2　日

國　立　交　通　大　學

博碩士紙本論文著作權授權書

（提供授權人裝訂於全文電子檔授權書之次頁用）

本授權書所授權之學位論文，為本人於國立交通大學 財務金融 所， 96 學年度第 1 學期取得博士學位之論文。

論文題目：財務市場波動及基金績效之計量分析

指導教授：鍾惠民 教授

■ 同意

本人茲將本著作，以非專屬、無償授權國立交通大學，基於推動讀者間「資源共享、互惠合作」之理念，與回饋社會與學術研究之目的，國立交通大學圖書館得以紙本收錄、重製與利用；於著作權法合理使用範圍內，讀者得進行閱覽或列印。

本論文為本人向經濟部智慧局申請專利(未申請者本條款請不予理會)的附件之一，申請文號為：____________________，請將論文延至____年____月____日再公開。

授 權 人：孫而音

親筆簽名：____________________

中華民國　96　年　10　月　2　日

國家圖書館博碩士論文電子檔案上網授權書

ID:GT009331813

本授權書所授權之學位論文，為本人於國立交通大學 財務金融 所， 96 學年度第 1 學期取得博士學位之論文。

論文題目：財務市場波動及基金績效之計量分析

指導教授：鍾惠民 教授

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

※ 讀者基於非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：孫而音

親筆簽名：________________

民國 96 年 10 月 2 日

本授權書請以黑筆撰寫，並列印二份，其中一份影印裝訂於附錄三之二(博碩士紙本論文著作權授權書)之次頁；另一份於辦理離校時繳交給系所助理，由圖書館彙總寄交國家圖書館。

財務市場波動及基金績效之計量分析

學生：孫而音　　　　　　　　　　　指導教授：鍾惠民 博士

國立交通大學財務金融研究所博士班

中文摘要

本研究主要利用計量分析探討兩個財務市場的重要議題。第一個議題是財務市場波動，利用 HAR 及 MIDAS 迴歸模型探討已實現變幅波動的預測績效，研究結果發現已實現變幅波動較已實現波動有效，將波動分為連續與跳躍作為迴歸子預測未來的波動幾乎較利用其他不同變異的迴歸子其 MSE 為最小，並檢定出在已實現變幅波動下短期的跳躍會產生結構性的改變。此外，並以 HAR 及 MIDAS 迴歸模型為涵蓋迴歸，將已實現變幅波動分成連續與跳躍部份，探討此二者及隱含波動對已實現變幅波動的資訊內涵，研究結果顯示隱含波動對已實現變幅波動具有很高的資訊內涵，連續部份提供與隱含波動部份相同的資訊內涵，然而跳躍部份無法對已實現變幅波動提供任何有用的資訊；若僅考慮隱含波動與樣本外預測，則隱含波動亦提供較大的資訊內涵。第二個議題是利用門檻迴歸模型檢驗基金經理人的選股及擇時能力，實證結果顯示傳統的 Henriksson 及 Merton (1981) 模型為我們所提的門檻迴歸模型之特例，利用傳統的迴歸模型會低估擇時能力，而門檻迴歸模型能產生較為正確的推論。

關鍵字：已實現變幅波動、涵蓋迴歸、資訊內涵、共同基金、門檻模型

Essays on the Econometric Analysis of Financial Market Volatility and Mutual Fund Performance

Student: Erh-Yin Sun Advisers: Dr. Huimin Chung

Graduate Institute of Finance

National Chiao Tung University

ABSTRACT

This dissertation consists of two separate issues.

The first issue is to discuss the forecasting performance of HAR and MIDAS regression models of realized range-based volatility; we focus on the S&P 500 index. The empirical results show that the realized range-based volatility is more efficient than the realized return-based volatility; the regressors consisting of the continuous sample path and jump variability measures in the HAR and MIDAS regressions predict the future realized range volatilities, and thus dominate almost in all MSE terms. In addition, the realized range-based regressions are significant for short-run volatility forecasting, but the realized return-based regressions are almost invariant to jumps. Furthermore, we will employ the HAR and MIDAS regressions as encompassing regressions to examine the information content of the continuous and jump components of the realized range-based volatility, and the additional information content of the implied volatility as an additional regressor. We use the VIX as the measure of the implied volatility. We find that the implied volatility has a high information content and the past continuous components feature relevant information content by the implied volatility. Besides, the jump components do not contribute to future valuable information.

The second issue is to detect mutual fund market timing abilities, using the threshold regression model. The empirical results show that the traditional Henriksson and Merton (1981) model is only a special case within our model, and we demonstrate the potential bias of using the traditional model, arguing that it tends to underestimate the market-timing effect. Indeed, we find that the use of the traditional market timing test may provide misleading results in some circumstances; thus, our proposed threshold model provides more accurate inferences on the market-timing effects of mutual funds.

Keywords: realized range-based volatility; HAR regression; MIDAS regression; VIX; information content; mutual fund; threshold regression

誌謝

隨著論文口試結束，到論文的修改完成，在工作多年後，重拾書本回到校園追求另一個領域的知識，終於即將再劃上另一個句點。能夠完成這一切，心中充滿了感激。

首先，我要感謝從碩博班一路指導我的 鍾 教授 鍾惠民博士，讓我能以最短的時間完成學業。老師除了在廣闊的學術視野及對治學嚴謹的堅持是我學習的目標，待人處事上更是我學習的榜樣，老師的教誨我會銘記在心。

接下來感謝在論文口試期間，承蒙 許 副校長 許 和鈞副校長、謝 教授 謝 國文教授、 謝 教授 謝 文良教授及 王 教授 王 耀輝教授，在百忙之中撥冗指導，對於論文初稿的疏漏與不當之處，提供許多寶貴的意見，使本論文能更加完善。

感謝交通大學財務金融所所有的教授先生們，在我碩博士期間的教誨，以及文淇小姐、佳芸小姐及稚瑩小姐在行政上的協助與幫忙。另外，本論文能如此順利的完成，感謝凱秩學弟在程式等多方面提供了大量的協助與建議；而自口試至論文的製作完成感謝志強學長的指導與幫忙，使我能更順利地完成它。

感謝這幾年來，在我有挫折、低潮時，傾聽我抱怨並鼓勵我的僑光技術學院同事，淑琴、雅嫺、惠芬、惠美、媖媖、淑然等，也感謝媖媖、郁宜在英文撰寫的指導與修正。

最後，我要感謝我最親愛的家人，我的父親、大姐、二姐、二姐夫、弟弟、弟媳、妹妹、妹婿以及四位可愛的外甥、外甥女、姪女，這幾年來不斷地支持與鼓勵，感謝您們無私地奉獻，讓我能順利地完成學業。繕寫至此，心中有著些許的遺憾，謹以此論文獻給在天國的母親，永遠懷思您的愛與溫慈的呵護。

CONTENTS

LIST OF TABLES

Chapter 1 Introduction

Applied econometric methods will be used for the estimations and forecasts of important financial quantities, analyses of economic outcomes and market behaviors. This dissertation will employ econometric analyses to focus on two important issues in the financial market, including volatility forecasting and mutual fund performance evaluation.

The first issue in this dissertation is to discuss the predicting volatility. Volatility forecasting of financial asset returns is important for derivative pricing, asset allocation, and risk management. Most of the previous studies used the realized volatility to predict volatility; and Forsberg and Ghysels (2007) explored different variations as regressors to predict the realized volatility, using the heterogeneous autoregressive and the mixed data sampling regressions. Because the realized range-based estimation of integrated variance has been proved to be more efficient, we are motivated to predict this realized range-based volatility. By following Forsberg and Ghysels (2007), we use the heterogeneous autoregressive and mixed data sampling regressions to examine whether future volatility is well predicted by the past realized range-based variations, realized range bipower variations , and so on.

In addition, most of the previous studies have documented the information content of the implied volatility. They always focus on whether the implied volatility has the additional information content of historical volatility; the realized volatility is always used as the historical volatility. We will employ the heterogeneous autoregressive regressions and mixed data sampling regressions as encompassing regressions to examine the information content of the continuous and jump components of the realized range-based volatility, and the additional information content of the implied volatility as an additional regressor. Besides, we use the

Chicago Board Options Exchange volatility index new VIX as the measure of the implied volatility. The new VIX is based on S&P 500 index options and adopts the model-free volatility expectation.

The second issue examined in this dissertation is to detect mutual fund market timing abilities by the threshold regression model. Investment performance and the market timing of mutual funds continue to receive considerable attention by both academics and market practitioners, with a variety of evaluation techniques having been proposed and implemented over the years. This chapter proposes a new method to test mutual fund performance and market timing through the application of threshold regression techniques. The idea is that fund managers may adopt different trading strategies when they perceive different market conditions. As fund managers may not uniformly use the sign of the market return to capture the direction of market movement, it is natural to conjecture that a fund manager's trading behavior changes when the market return is above or below a certain threshold level, which varies across managers of different funds.

To sum up, the dissertation provides some insights into the issues of volatility forecasting and mutual fund performance evaluation. With these points in mind, the research results will provide us with the empirical evidences to comprehend the occasion of some distinctive phenomena in financial markets.

Chapter 2　On the Forecasting Performance of HAR and MIDAS Regression Models of Realized Range-Based Volatility

1 Introduction

Volatility forecasting of financial asset returns is important for derivative pricing, asset allocation, and risk management. The latent volatility is unobservable, so the observable proxies are used; in this study, we will use the realized variance (Andersen et al. (2001)) and the realized range variance (Parkinson (1980)). In recent years, with the availability of high frequency financial market data, the previous studies showed that the use of high-frequency data is beneficial in predicting volatility. The return-based estimation is the realized variance (RV), which is the sum of the squared returns over non-overlapping intervals within a sampling period. The theory states that the RV is an unbiased and efficient estimator (Andersen et al. (2001)), converges in probability to the quadratic variation (QV) as the sampling frequency tends to infinity (Barndorff-Nielsen and Shephard (2002)), and has the persistent properties (Andersen et al. (2003)). In practice, market microstructure effects pose the RV to become biased and inconsistent (Bandi et al. (2005, 2006), Aït-Sahalia et al. (2005) and Hansen et al. (2005)). An alternative way of range-based estimation is the realized range variance (RRV), which is based on the difference between the maximum and minimum prices observed during a certain period. Cristensen and Podolskij (2005) derived that the RRV has the theoretical properties similar to Barndorff-Nielsen and Shephard (2002) for RV, and it is a more efficient volatility estimator than the realized volatility (Parkinson (1980) and Martens et al. (2007)) because it is formed from the entire process. Most authors explored the univariate persistent properties, using the RV or RRV to predict volatility. The regressions of the heterogeneous autoregressive (HAR) by Corsi (2004) and of mixed data sampling (MIDAS) by Ghysels et al. (2006) were considered to be some variations as regressors to predict future increments in quadratic variations. These two models are able to reproduce the memory persistence observed in the data and to reduce the number of parameters to estimate; and their prediction powers were proved to outperform ARFIMA.

Furthermore, in high-frequency data, discrete jumps may affect returns, so the jump component appears to account for a proportion of quadratic variation. In this chapter, we consider a continuous time stochastic volatility jump-diffusion model for the asset returns. An asymptotic distribution theory for the jump test was derived by Barndorff-Nielsen and Shephard (2006), which is based on the ratio of RV and bipower variation (BPV), suitably normalized. Moreover, similar tests based on RRV were derived by Christensen and Podolskij (2006). Furthermore, Andersen et al. (2007) suggested the use of the bipower variation as the predictor of volatility, because jump components are “noise” and not helpful in predicting future volatility. The HAR and MIDAS regressions allow us to compare the forecasting abilities of different regressors and to choose the predictors with the best predicting ability. In this chapter, we will use RV, RRV, and BPV, etc. as regressors to predict latent volatilities.

In this chapter, our purpose is to forecast volatility. Most of the previous studies used the RV to predict volatility, and Forsberg and Ghysels (2007) explored different variations as regressors to predict the realized volatility, using the HAR and MIDAS regressions. Because the realized range-based estimation of integrated variance has been proved to be more efficient, we are motivated to predict this realized range-based volatility. By following Forsberg and Ghysels (2007), we use HAR and MIDAS regressions to examine whether future volatility is well predicted by past realized range-based variations, realized range bipower variations (RBV), and so on.

The results show that the realized range-based variance is more efficient and the results are the same as the previous studies. The regressors consisting of the continuous sample path and jump variability measures (CJ) in the HAR and MIDAS regressions predict the future realized range volatilities, and thus dominate almost in all mean square error (MSE) terms. In addition, in the in-sample forecasting, the relative decreasing ratio of MSE of HAR-RRV (MIDAS-RRV) regressions are almost larger than that of HAR-RV (MIDAS-RV); and in the out-of sample forecasting, the MSE of HAR-RRV and MIDAS-RRV regressions are small. Moreover, the realized range-based regressions are significant for short-run volatility forecasting, but the realized return-based regressions are almost invariant to jumps. Hence, our empirical results show that by using of the HAR and MIDAS regressions to predict latent volatility under different variations, the realized range-based variance is a good volatility proxy.

The remainder of this chapter is organized as follows. In section 2, we discuss various volatility measures and predict volatility regressions. In section 3, we present the data and the empirical results of HAR and MIDAS regressions for the S&P 500 cash index to predict RV and RRV. Section 4 concludes the article.

2 The Methodology

2.1 Construction of volatility measures

Let the logarithmic price of financial assets at time t be denoted by $p(t)$ and follow the continuous-time jump diffusion process

$$dp(t) = \mu(t)dt + \sigma(t)dw(t) + \kappa(t)dq(t), \tag{1}$$

where $\mu(t)$ and $\sigma(t)$ are the drift and instantaneous volatility, $w(t)$ is the standardized Brownian motion, $q(t)$ is a counting process with time-varying intensity $\lambda(t)$, *i.e.*

$P[dq(t)=1]=\lambda(t)dt$, and $\kappa(t)$ is the jump size. The quadratic variation process for a sequence of partitions is defined by

$$[p](t)=\text{plim}\sum_{j=0}^{n-1}(p(s_{j+1})-p(s_j))^2 ,$$

where $0=s_0<s_1<\cdots<s_n=t$ and $\sup_j\{s_{j+1}-s_j\}\to 0$ for $n\to\infty$. The quadratic variation for the cumulative return process, $QV_{t,t+H}\equiv p(t+H)-p(t)$, is then

$$QV_{t,t+H}\equiv\int_t^{t+H}\sigma^2(s)ds+\sum_{t<s\leq t+H}\kappa^2(s) . \qquad (2)$$

Hence, the quadratic variation from time t to time $t+H$ consists of the integrated volatility of the continuous sample path component ($\int_t^{t+H}\sigma^2(s)ds$) and the summation of the squared jumps between time t and time $t+H$.

Let daily returns be denoted by $r_{t,t-1}=p(t)-p(t-1)$, where the time index t refers to daily sampling. When the data is sampled at a higher frequency, M times in a day, we will denote the intraday returns and ranges as:

$$r_{t,j}=p(t-1+\tfrac{j}{M})-p(t-1+\tfrac{(j-1)}{M}),\ \ j=1,\cdots,M ;$$

and
$$s_{pi\Delta,\Delta,m}=\max_{(i-1)/M\leq s,t\leq i/M}\{p_t-p_s\} ,$$

where $s_{pi\Delta,\Delta,m}$ represents that each range is based on the corresponding m returns. The realized variance and the realized ranged-based variance over day t are defined as

$$RV_t=\sum_{j=1}^{M}r_{t,j}^2 , \qquad (3)$$

$$RRV_t^{m*}=\frac{1}{\lambda_{2,m}}\sum_{i=1}^{M}s_{pi\Delta,\Delta,m}^2 , \qquad (4)$$

where $\lambda_{j,m}=E\left(s_{W,m}^j\right)<\infty,\ \forall j\in R$ and $m\in N$ [1], and W is the Brownian motion. In

[1] There is no explicit formula for $\lambda_{j,m}$, but it is computed to any degree of accuracy from simulations.

addition, from time t to time $t+H$, the realized variance is $RV_{t,t+H} = \sum_{i=1}^{H} RV_{t+i-1,t+i}$.

As the theory of quadratic variation is applied, the realized variance converges uniformly in probability to the increment of the quadratic variation process as the sampling frequency of the underlying returns increases. That is,

$$RV_{t,t+H} \xrightarrow{P} QV_{t,t+H} . \tag{5}$$

The above equation means that the realized variation $RV_{t,t+H}$ is a consistent estimator of the quadratic variation $QV_{t,t+H}$. Other measures of realized within-day price variance are based on the realized power variation, which is denoted by

$$RPV_{t,t+1} = \sum_{j=1}^{M} \left| r_{t,j} \right| .$$

One is also called the realized bipower variation,

$$BPV_{t,t+1}(k) = \tfrac{\pi}{2}\left(\tfrac{M}{M-1-k}\right) \sum_{j=2+k}^{M} \left| r_{t,j-(1+k)} \right| \cdot \left| r_{t,j} \right| \quad , k \geq 0 , \tag{6}$$

in which two absolute returns of every other $(k+1)$ days multiply together while the other is the tripower variation,

$$TPV_{t,t+1}(k) = M \mu_{4/3}^{-3} \left(\tfrac{M}{M-2(1+k)}\right) \sum_{j=1+2(1+k)}^{M} \left(\left| r_{t,j-2(1+k)} \right| \left| r_{t,j-(1+k)} \right| \left| r_{t,j} \right| \right)^{4/3} ,$$
$$k \geq 0, \quad \mu_{4/3} = 2^{2/3} \cdot \frac{\Gamma(\frac{7}{6})}{\Gamma(\frac{1}{2})}, \tag{7}$$

in which three absolute returns of every other $(k+1)$ days multiply together and is raised to $\frac{4}{3}$ power (Barndorff-Nielsen and Shephard (2002, 2004)).

As noted by Barndorff-Nielsen and Shephard (2004) and Andersen, Bollerslev, and Diebold (2004), there are some results:

$$\lim_{M \to \infty} BPV_{t,t+1}(k) \xrightarrow{P} \int_{t}^{t+1} \sigma^2(s) ds , \tag{8}$$

$$\lim_{M \to \infty} TPV_{t,t+1}(k) \xrightarrow{P} \int_{t}^{t+1} \sigma^4(s) ds . \tag{9}$$

Under Barndorff-Nielsen and Shephard's (2006) joint asymptotic distribution, there are some results as follows,

$$\frac{RV_{t,t+1} - QV_{t,t+1}}{\sqrt{2 \cdot TPV_{t,t+1}(k)/3}} \sim N(0,1),$$

$$\frac{\sqrt{\pi/2}M^{-1/2}RPV_{t,t+1}(k) - \int_t^{t+1}\sigma(s)ds}{\sqrt{(\pi/2 - 1)M^{-1}RV_{t,t+1}(k)}} \sim N(0,1),$$

$$\frac{BPV_{t,t+1}(k) - \int_t^{t+1}\sigma^2(s)ds}{\sqrt{(\pi^2/4 + \pi - 5)TPV_{t,t+1}(k)}} \sim N(0,1).$$

As noted by the notion of Barndorff-Nielsen and Shephard (2004), combining the results of equations (5) and (8), the jumps in the underlying prices may consistently be estimated by

$$\lim_{M\to\infty}(RV_{t,t+1} - BPV_{t,t+1}) = \sum_{t<s<t+1}\kappa^2(s), \tag{10}$$

but in a given finite sample, the difference between RV and BPV may be negative. Hence, when we follow the suggestion of Andersen, Bollerslev, and Diebold (2007), they truncated the actual empirical measurements at zero,

$$J_{t,t+1} \equiv \max(RV_{t,t+1} - BPV_{t,t+1}, 0), \tag{11}$$

to ensure that all of the estimates are non-negative. However, the $J_{t,t+1}^{1/2}$ also had an unreasonably large number of non-zero small positive value. They proposed that these small jumps were treated as measurement errors, associating with only large values as jump components. Following Barndorff-Nielsen and Shephard (2004b), the joint asymptotic distribution of the realized volatility and bipower variation measures results in the following test statistics,

$$Z_{t+1} = \sqrt{M}\cdot\frac{(RV_{t,t+1}(k) - BPV_{t,t+1}(k))/RV_{t,t+1}(k)}{\sqrt{((\frac{\pi}{4})^2 + \pi - 5)\cdot\max(1, \frac{TPV_{t,t+1}(k)}{BPV_{t,t+1}^2(k)})}} \sim N(0,1). \tag{12}$$

Huang and Tauchen (2005) found that the statistics in equation (12) also had a sensible power against other empirically calibrated stochastic volatility jump diffusion models. Using equation (12), Andersen, Bollerslev, and Diebold (2007) identified the jump variation as,

$$J_{t,t+1} = I(Z_{t+1} > \Phi_{\alpha})(RV_{t,t+1} - BPV_{t,t+1}), \tag{13}$$

and the continuous component variation was estimated as the residual,

$$C_{t,t+1} = I(Z_{t+1} \le \Phi_{\alpha}) \cdot RV_{t,t+1} + I(Z_{t+1} > \Phi_{\alpha}) \cdot BPV_{t,t+1}, \tag{14}$$

where $I(\cdot)$ denotes the indicator function, $\Phi(\cdot)$ is the standard cumulated normal distribution ($\Phi_{\alpha} = \Phi^{-1}(\alpha)$), and α is a significant level of the bipower test and we will use $\alpha = 0.999$ throughout the chapter. From the definitions in equations (13) and (14), we ensure that the continuous variation and jump variation sum to the total realized variation, *i.e.* $RV_{t,t+1} = C_{t,t+1} + J_{t,t+1}$.

Similar results were shown by Christensen and Podolskij (2006a, 2006b). As noted by Christensen and Podolskij (2006a, 2006b),

$$RRV_t^{m*} \xrightarrow{p} \int_0^t \sigma^2(s)ds + \frac{1}{\lambda_{2,m}} \cdot \sum_{i=1}^{M} J_i^2, \text{ as } M \to \infty,$$

i.e., RRV_t^{m*} is inconsistent. Hence, they modified the intraday high-low statistic to make it consistent with the quadratic variation. The realized range-based bipower variation with parameter $(r,s) \in R_+^2$ is defined as[2]:

$$RBV_{(r,s),t}^{m} = M^{(r+s)/2-1} \frac{1}{\lambda_{r,m}} \frac{1}{\lambda_{s,m}} \sum_{i=1}^{M-1} s_{pi\Delta,\Delta,m}^{r} s_{p(i+1)\Delta,\Delta,m}^{s},$$

then $RBV_{(1,1),t}^{m} \equiv RBV_t^m \xrightarrow{p} \int_0^t \sigma^2(s)ds$, as $M \to \infty$. Hence, the new range based estimator, $RRV_t^m \equiv \lambda_{2,m} RRV_t^{m*} + (1-\lambda_{2,m}) RBV_t^m$, is consistent with quadratic variation, *i.e.*, $RRV_t^m \xrightarrow{p} \int_0^t \sigma^2(s)ds + \sum_{i=1}^{M} J_i^2$.

Furthermore, using the bivariate distribution of (RRV_t^m, RBV_t^m) and the delta-method, Christensen and Podolskij (2006b) found the jump detection statistic,

[2] I maintain some notations used by Christensen and Podolskij (2006b) throughout the chapter.

$$\frac{\sqrt{M}\,(RRV_t^m - RBV_t^m)}{\sqrt{\nu_m \int_0^t \sigma(s)_u^4 ds}} \xrightarrow{\ d\ } N(0,1)\,,$$

where $\nu_m = \lambda_{2,m}^2(\Lambda_m^R + \Lambda_m^B - 2\Lambda_m^{RB})$, $\Lambda_m^R = \dfrac{\lambda_{4,m} - \lambda_{2,m}^2}{\lambda_{2,m}^2}$, $\Lambda_m^B = \dfrac{\lambda_{2,m}^2 + 2\lambda_{1,m}^2\lambda_{2,m} - 3\lambda_{1,m}^4}{\lambda_{1,m}^4}$, and

$\Lambda_c^{RB} = \dfrac{2\lambda_{3,c}\lambda_{1,c} - 2\lambda_{2,c}\lambda_{1,c}^2}{\lambda_{2,c}\lambda_{1,c}^2}$. In addition, they adopted the modified ratio-statistic to improve the size properties in finite samples. The modified ratio-statistic is

$$Z_{t+1} = \frac{\sqrt{M}\left(1 - {}^{RBV_t^m}\!/\!_{RRV_t^{*m}}\right)}{\sqrt{\nu_m \max\left\{{}^{RQQ_t^m}\!/\!_{(RBV_t^{*m})^2}\,,\, {}^{1}\!/\!_{t}\right\}}} \xrightarrow{\ d\ } N(0,1)\,, \tag{15}$$

where $RQQ_t^m = \frac{M}{\lambda_{1,m}^4}\sum_{i=1}^{M-3} s_{pi\Delta,\Delta,m}\, s_{p(i+1)\Delta,\Delta,m}\, s_{p(i+2)\Delta,\Delta,m}\, s_{p(i+3)\Delta,\Delta,m}$ and $RQQ_t^m \xrightarrow{\ p\ } \int_0^t \sigma^4(s)ds$.

2.2 Predicting volatility models

In this study, we will explore MIDAS regression model, which was introduced by Ghysels et al. (2002, 2006) and HAR regression model, which was suggested by Corsi (2004), to predict volatility from t to $t+H$, where H is the prediction horizon in days.

The multi-period realized variances, which were constructed by Andersen et al. (2007), were defined as the normalized sum of the one-period realized variances,

$$RV_{t,t+H} = H^{-1}(RV_{t+1} + RV_{t+2} + \cdots + RV_{t+H}),\ \ RV_{t,t+1} \equiv RV_{t+1}\,,$$

where $H = 1,\ 5,\ 10,\ 15,\ \text{and}\ 20$. H is the prediction horizon in days, in the empirical analysis, as one day, weekly, bi-weekly, tri-weekly, and monthly. In addition, we will follow Forsberg et al. (2007), using RV and its standard deviation and log form to predict volatility.

The HAR-RV models have been introduced by Corsi (2004), and they can capture the long memory property of the realized variance. The model is defined as

$$RV_{t,t+1} = \alpha_0 + \alpha_D X_{t-1,t} + \alpha_W X_{t-5,t} + \alpha_M X_{t-20,t} + \varepsilon_{t,t+1}, \tag{16}$$

in which $X = RV, BPV, C,$ and CJ. Similarly, the HAR-RRV model is as follows

$$RRV_{t,t+1} = \alpha_0 + \alpha_D X_{t-1,t} + \alpha_W X_{t-5,t} + \alpha_M X_{t-20,t} + \varepsilon_{t,t+1}, \tag{17}$$

where $X = RRV, RBV, C,$ and CJ. Andersen et al. (2007) defined the HAR-RV-CJ model, which explores the separation of RV_t into the continuous part C_t and jump part J_t. This separation was suggested by Barndorff-Nielsen et al. (2004). The HAR-RV-CJ model and the HAR-RRV-CJ are shown below

$$\begin{aligned} RV_{t,t+H} &= \alpha_0 + \alpha_{CD} C_{t-1,t} + \alpha_{CW} C_{t-5,t} + \alpha_{CM} C_{t-20,t} \\ &\quad + \alpha_{JD} J_{t-1,t} + \alpha_{JW} J_{t-5,t} + \alpha_{JM} J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned}, \tag{18}$$

$$\begin{aligned} RRV_{t,t+H} &= \alpha_0 + \alpha_{CD} C_{t-1,t} + \alpha C_{CW} C_{t-5,t} + \alpha_{CM} C_{t-20,t} \\ &\quad + \alpha_{JD} J_{t-1,t} + \alpha_{JW} J_{t-5,t} + \alpha_{JM} J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned}. \tag{19}$$

Besides, following Forsberg and Ghysels (2007), we will consider the HAR-RV^{½}-X^{½} (HAR-RRV^{½}-X^{½}) model, and the HAR-ln RV-ln X (HAR-ln RRV-ln X) model, with $X = RV(RRV), BPV(RBV), C,$ and CJ. The models are as follows,

$$RV^{½}_{t,t+1}(RRV^{½}_{t,t+1}) = \alpha_0 + \alpha_D X^{½}_{t-1,t} + \alpha_W X^{½}_{t-5,t} + \alpha_M X^{½}_{t-20,t} + \varepsilon_{t,t+1}, \tag{20}$$

$$\begin{aligned} &\ln RV_{t,t+H} (\ln RRV_{t,t+H}) \\ &\quad = \alpha_0 + \alpha_D \ln X_{t-1,t} + \alpha_W \ln X_{t-5,t} + \alpha_M \ln X_{t-20,t} + \varepsilon_{t,t+H} \end{aligned}, \tag{21}$$

$$\begin{aligned} RV^{½}_{t,t+1}(RRV^{½}_{t,t+1}) &= \alpha_0 + \alpha_{CD} C^{½}_{t-1,t} + \alpha_{CW} C^{½}_{t-5,t} + \alpha_{CM} C^{½}_{t-20,t} \\ &\quad + \alpha_{JD} J^{½}_{t-1,t} + \alpha_{JW} J^{½}_{t-5,t} + \alpha_{JM} J^{½}_{t-20,t} + \varepsilon_{t,t+1} \end{aligned}, \tag{22}$$

$$\begin{aligned} &\ln RV_{t,t+H} (\ln RRV_{t,t+H}) \\ &\quad = \alpha_0 + \alpha_{CD} \ln C_{t-1,t} + \alpha_{CW} \ln C_{t-5,t} + \alpha_{CM} \ln C_{t-20,t} \\ &\quad + \alpha_{JD} \ln J_{t-1,t} + \alpha_{JW} \ln J_{t-5,t} + \alpha_{JM} \ln J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned}, \tag{23}$$

where $X^{1/2}_{t,t+H} = H^{-1}\left(X^{1/2}_{t+1} + X^{1/2}_{t+2} + ... + X^{1/2}_{t+H}\right)$ and $\ln X_{t,t+H} = H^{-1}\left(\ln X_{t+1} + ... + \ln X_{t+H}\right)$.

The differences between MIDAS and HAR regressions models are the lagged regressors and their weights. The MIDAS regression is introduced by Ghysels et al.

(2002, 2005). MIDAS regressions can run parsimoniously parameterized regressors of the data observed at different frequencies. Ghysels et al. (2006) used the MIDAS regressions to predict volatility. In my present chapter, I followed their method. The MIDAS-RV and MIDAS-RRV models can be written as

$$RV_{t,t+H} = \mu_H + \phi_H \sum_{k=0}^{k_{\max}} b(k,\theta_1,\theta_2) X_{t-k-1,t-k} + \varepsilon_{t+H}\,, \tag{24}$$

$$RRV_{t,t+H} = \mu_H + \phi_H \sum_{k=0}^{k_{\max}} b(k,\theta_1,\theta_2) X_{t-k-1,t-k} + \varepsilon_{t+H}\,, \tag{25}$$

where $b(k;\theta_1,\theta_2) = \dfrac{f(k/k_{\max};\theta_1,\theta_2)}{\sum_{k=1}^{k_{\max}} f(k/k_{\max};\theta_1,\theta_2)}$, $f(x;\theta_1,\theta_2) = \dfrac{x^{\theta_1-1}(1-x)^{\theta_2-1}}{\beta(\theta_1,\theta_2)}$, and

$\beta(\theta_1,\theta_2) = \dfrac{\Gamma(\theta_1)\,\Gamma(\theta_2)}{\Gamma(\theta_1+\theta_1)}$, with $X = RV(RRV),\ BPV(RBV),\ C,$ and CJ . For all the MIDAS regressions, we use $k_{\max} = 50$ [3], and we fix $\theta_1 = 1$[4]. As noted by Forsberg et al. (2007), we will consider the MIDAS-RV^{½}-X^{½} model, and the MIDAS-lnRV-lnX model, with $X = RV,\ BPV,\ C,$ and CJ . The models are as follows,

$$RV_{t,t+H}^{½} = \mu_H + \phi_H \sum_{k=0}^{k_{\max}} b(k,\theta_1,\theta_2) X_{t-k-1,t-k}^{½} + \varepsilon_{t+H}\,, \tag{26}$$

$$\ln(RV_{t,t+H}) = \mu_H + \phi_H \sum_{k=0}^{k_{\max}} b(k,\theta_1,\theta_2) \ln(X_{t-k-1,t-k}) + \varepsilon_{t+H}\,; \tag{27}$$

and under the MIDAS-RRV^{½}-X^{½} and MIDAS-lnRRV-lnX, the $RV_{t,t+H}$ is replaced by $RRV_{t,t+H}$, and the $BPV_{t,t+H}$ is replaced by $RBV_{t,t+H}$ in the equations (26), (27).

2.3 Conditional models

We will show whether the jump will make the structural breaks in conditional variance process or not. The structural change is modeled by allowing the dynamics to be different, if the bipower jump test indicates a jump at time t. The conditional

---

[3] Ghysels (2006) showed that using longer lags (*i.e.*, $k > 50$) resulted in little effect over the results.

[4] This will give us declining weights in the lag polynominal.

model is given by

$$
\begin{aligned}
RV_{t,t+H} &= (\alpha_0 + \alpha_0^J D_t) + (\alpha_D + \alpha_D^J D_t) X_{t-1,t} + (\alpha_W + \alpha_W^J D_t) X_{t-5,t} \\
&\quad + (\alpha_M + \alpha_M^J D_t) X_{t-20,t} + \varepsilon_{t,t+H}
\end{aligned}, \qquad (28)
$$

where D_t is the indicated variable, in which $D_t = 1$, if the bipower jump test indicates a jump at time t. $X = RV, BPV$, and C. We employ the similar approach for its standard deviation and log form regressions. In the realized range-based variance, the RV replaces the RRV, and the BPV is replaced by RBV. We will test the null $H_0 : \alpha_0^J = \alpha_D^J = \alpha_W^J = \alpha_M^J = 0$, using the Chow Test. If the null H_0 is rejected, it represents that the jump will result in the structural breaks.

2.4 Evaluation measures

Patton (2006) showed that the mean square error (MSE) loss function is robust with regards to the volatility proxy used. In this chapter, we will use the MSE as a comparing criterion. In the realized return-based volatility, let $RV_{t,t+H}$ denote the true value of the RV for the H days, and let $\widehat{RV}_{t,t+H}$ denote the predicted value of the dependent variable. In the realized range-based volatility, RV is replaced by RRV. Then the MSE is given by

$$
MSE = N^{-1} \sum_{i=1}^{N} (RV_{i,i+H} - \widehat{RV}_{i,i+H})^2 ,
$$

where N is the number of forecasts. As denoted by Forsberg and Ghysels (2007), in order to be able to compare the MSE from the regressions, we undo the transformation, *i.e.*, the MSE of the RV in standard deviation and log form are respectively given by

$$
MSE = N^{-1} \sum_{i=1}^{N} (RV_{i,i+H} - (\widehat{RV}^{1/2}_{i,i+H})^2)^2 ,
$$

$$
MSE = N^{-1} \sum_{i=1}^{N} (RV_{i,i+H} - \exp(\ln \widehat{RV}_{i,i+H}))^2 .
$$

Using these measures, we can compare the models with different transformations of the dependent variable.

Moreover, we will compare both the HAR and MIDAS models. Because the dependent variables are not identical, we will compare their relative decreasing ratio of MSE.

3 Data and empirical results

3.1 Data descriptions

This study employs 5-minute intra-day data of the S&P 500 index securities. The intraday data are obtained from Tick Data Inc., covering the period from January 1, 1995 to March 31, 2005, and consisting of 2535 days with 78 intra-day 5-minute observations. Table 1 shows that the descriptive statistics of the data. Panel A represents the descriptive statistics of the realized return-based variation. Panel B represents the descriptive statistics of the realized range-based variation. LB_{10} reports the Liung-Box test statistics for up to the tenth order serial correlation. RV_t denotes the realized variance, RRV_t is the realized range-based variance; C_t is the continuous part, and J_t is the jump part of RV_t (or RRV_t) as separated by the bipower jump test of Barndorff-Nielsen and Shephard (2004a). For the bipower jump test, a significance level $\alpha = 0.999$ is used, and the critical value of LB_{10} is 18.3070. BPV_t (RBV_t) denotes the realized bipower (realized range bipower) variation.

Table 1 shows that the realized range-based variation is a more efficient volatility estimator than the realized return-based variation because the standard deviation of RRV_t is smaller than the standard deviation of RV_t. But note that the means of the realized range-based variations are smaller than the realized variations, *i.e.*, using the realized range-based variations to estimate the latent volatility will induce a downward bias. This is because the price path is not observed continuously.

The observed minimum and maximum price over- and underestimates the true minimum and maximum, respectively. Studying the LB statistics, the RV_t (RRV_t) exhibits the highest degree of serial correlation for all the transformations, which denote that these volatility measures have higher persistence.

3.2 In-sample empirical results

In Table 2, we examine the in-sample fit for the HAR regressions of the realized variance (HAR-RV) and the realized range-based variance (HAR-RRV), using the S&P 500 cash index data set. Panel B and C denote the standard deviation and log transformation of the variances respectively. We focus on five different prediction horizons, one day, one, two, three and four weeks, corresponding to $RV_{t,t+H}$ and $RRV_{t,t+H}$ for $H = 1$, 5, 10, 15, and 20 respectively. For both the HAR-RV and the HAR-RRV regressions, comparing across prediction horizons, the MSE is always the lowest when the horizon is two weeks, thus indicating that the MSE for the two weeks prediction horizon is more precise than the other horizons. Furthermore, the relative decreasing ratios of the mean square errors of HAR-RRV regressions are always larger than those of the HAR-RV regressions. That is to say, using the realized range-based volatility to predict the latent volatility proxy is more precise than the realized volatility.

In Table 3, we examine the in-sample fit for the MIDAS regressions of the realized variance (MIDAS-RV) and the realized range-based variance (MIDAS-RRV), using the S&P 500 cash index data set. We can obtain similar results as the above ones. For both the MIDAS-RV and the MIDAS-RRV regressions, when we predict volatility for two or three weeks horizons, we can get more accurate predictions. The relative decreasing ratios of the mean square errors of MIDAS-RRV regressions are always larger than those of the MIDAS-RV regressions, *i.e.*, using the realized range-based volatility to predict the latent volatility is more precise than the realized volatility.

3.3 Conditional HAR models

If there is a jump at time t, the jump would imply a different dynamic of the continuous part of the price process. Table 4 reports the result of the Chow test of the corresponding F statistics and $P-values$. The realized range-based regressions are significant at one day and one week horizons, *i.e.*, the price process is not invariant to jumps at short-run volatility forecasting, but the realized return-based regressions are almost invariant to jumps. The results are similar when we model the realized return and range-based variances in the standard deviation and the log form, as reported in Table 4.

Besides, by looking at the results in Table 5, there are similar results as the above ones. For both the conditional HAR-RV and the conditional HAR-RRV regressions, when we predict volatility for two or three weeks horizons, we can get more accurate predictions. In addition, the relative decreasing ratios of the mean square errors of HAR-RRV regressions are almost larger than those of the HAR-RV regressions, *i.e.*, using the realized range-based volatility of the conditional HAR regressions to predict the latent volatility, which is more precise than the realized volatility.

3.4 Out-of-sample empirical results

Table 6 and Table 7 report the out-of-sample results both for the HAR and MIDAS regressions of the realized volatility and the realized range-based volatility. We split the data into two parts: a sample part to estimate these two models and an out-of-sample part for forecasting. In this study, we explore the rolling window analysis to predict the out-of-sample forecasting. The rolling window's width is 2400 observations and the windows are rolled through the sample once for at a time; there will be 135 rolling estimates for each parameter. Similarly, there will be 131 (one week), 126 (two weeks), 121 (three weeks), and 116 (four weeks) rolling estimates for each parameter respectively. Hence, the first in-sample period covers January 1, 1995 to September 16, 2004, a total of 2400 days. For both the HAR and the MIDAS regressions, the mean square errors of the RRV are almost smaller than the RV regressions, and the results are the same as the above.

4. Conclusions

In this chapter, we employ the mixed data sampling regression and the heterogeneous autoregressive regression that are able to reproduce the memory persistence observed in the data, and easy to estimate; and the realized volatility and the realized range-based volatility measures have higher persistence. Using the MIDAS models and HAR models to predict volatility with the dependent variables, which are the realized volatility and the realized range-based volatility, we expect to find the most accurate way for forecasting volatility. From the empirical results, the realized range-based variance is more efficient; in the in-sample forecasting, the relative decreasing ratio of MSE of HAR-RRV (MIDAS-RRV) regressions are almost larger than that of HAR-RV (MIDAS-RV); and in the out-of sample forecasting, the MSE of HAR-RRV and MIDAS-RRV regressions are small. The regressors consisting of the continuous sample path and jump variability measures (CJ) in the HAR and MIDAS regressions predict the future realized range volatilities, and such dominates almost in all mean square error (MSE) terms. Furthermore, the realized range-based regressions are significant for short-run volatility forecasting, but the realized return-based regressions are almost invariant to jumps. Hence, for our empirical results, using the HAR and MIDAS regressions to predict latent volatility, under different variations, and the realized range-based variance is a good volatility proxy.

Table 2.1 Descriptive statistics of S&P 500 1995/01/01 ~ 2005/03/31

Panel A	Realized return-based						
	Mean	St. Dev	Skew	Kurt	Min	Max	LB_{10}
RV_t	0.8792	1.2114	9.5654	197.2392	0.0277	32.3951	4199.6178
C_t	0.8035	0.9845	4.5023	35.4207	0.0277	12.2312	6848.0977
J_t	0.0757	0.5616	34.7844	1497.8044	0.0000	24.8401	4.3412
BPV_t	0.7463	0.9574	4.9252	41.7276	0.0281	12.7705	6856.5760
$RV_t^{1/2}$	0.8336	0.4295	2.0464	13.1757	0.1665	5.6917	10051.2788
$C_t^{1/2}$	0.7980	0.4084	1.6319	7.8198	0.1665	3.4973	10702.3322
$J_t^{1/2}$	0.0931	0.2589	5.3464	62.9823	0.0000	4.9840	28.8587
$BPV_t^{1/2}$	0.7654	0.4008	1.7462	8.6065	0.1678	3.5736	10980.4914
$\ln RV_t$	-0.5945	0.9596	-0.0090	3.0608	-3.5861	3.4780	12603.5280
$\ln C_t$	-0.6873	0.9765	-0.0575	2.9382	-3.5861	2.5040	12677.0609
$\ln(J_t+1)$	0.0489	0.1675	6.7823	79.1892	0.0000	3.2519	59.7342
$\ln BPV_t$	-0.7779	0.9883	-0.0251	2.9318	-3.5703	2.5471	13156.5850
Panel B	**Realized range-based**						
	Mean	St. Dev	Skew	Kurt	Min	Max	LB_{10}
RRV_t	0.7589	1.0791	14.6190	422.6160	0.0365	35.2861	3476.9151
C_t	0.6740	0.8816	8.7098	167.0704	0.0365	22.5290	5505.4818
J_t	0.0849	0.3104	28.1328	1104.4405	0.0000	12.7571	58.2870
RBV_t	0.6477	0.8513	9.2516	188.7055	0.0374	22.5290	5375.3518
$RRV_t^{1/2}$	0.7785	0.3911	2.1817	17.5788	0.1911	5.9402	11308.8325
$C_t^{1/2}$	0.7326	0.3706	1.8761	11.7264	0.1911	4.7465	12130.6048
$J_t^{1/2}$	0.1759	0.2323	2.6193	23.8600	0.0000	3.5717	300.0182
$RBV_t^{1/2}$	0.7186	0.3623	1.9040	12.2177	0.1933	4.7465	12177.7777
$\ln RRV_t$	-0.7210	0.9379	0.0094	2.8980	-3.3103	3.5635	14373.9498
$\ln C_t$	-0.8478	0.9473	0.0517	2.7497	-3.3103	3.1148	14842.8362
$\ln(J_t+1)$	0.0686	0.1348	5.7459	69.4079	0.0000	2.6216	335.6110
$\ln RBV_t$	-0.8842	0.9429	0.0535	2.7622	-3.2871	3.1148	14924.9842

Note: The table shows that the S&P 500 cash index securities cover the period from January 1, 1995 to March 31, 2005, consisting of 2535 days with 78 intra-day 5-minute observations. Panel A represents the descriptive statistics of the realized return-based variations. Panel B represents the descriptive statistics of the realized range-based variations. LB_{10} reports the Liung-Box test statistic for up to the tenth order serial correlation. RV_t denotes the realized variance, RRV_t is the realized range-based variance; C_t is the continuous part, and J_t is the jump part of RV_t (or RRV_t) as separated by the bipower jump test of Barndorff-Nielsen and Shephard (2004a). The bipower jump tests a significant level at $\alpha = 0.999$; and the critical value of LB_{10} is 18.3070. BVP_t (RBV_t) denotes the realized bipower (realized range bipower) variation. In Panel A, the first part describes the RV_t, the next describes the square root transformation, and the last describes the log transformations of the variables. Panel B replaces RRV_t to RV_t.

Table 2.2 In-sample results S&P 500 1995/01/01 ~ 2005/03/31: HAR

Panel A	HAR-RV				HAR-RRV			
Horizon	RV	BPV	C	CJ	RRV	RBV	C	CJ
MSE								
1 day	0.9258	0.8394	0.8429	0.8389	0.7902	0.7545	0.7478	0.7140
	(0.2358)	(0.3071)	(0.3042)	(0.3075)	(0.2677)	(0.3008)	(0.3070)	(0.3383)
1 week	0.4256	0.4178	0.4091	0.4047	0.3198	0.3167	0.3151	0.3077
	(0.6487)	(0.6551)	(0.6623)	(0.6659)	(0.7036)	(0.7065)	(0.7080)	(0.7149)
2 weeks	0.2930	0.2684	0.2736	0.2681	0.2033	0.1897	0.1895	0.1861
	(0.7581)	(0.7784)	(0.7741)	(0.7787)	(0.8116)	(0.8242)	(0.8244)	(0.8275)
3 weeks	0.3002	0.2831	0.2931	0.2899	0.2032	0.1982	0.1979	0.1945
	(0.7522)	(0.7663)	(0.7580)	(0.7607)	(0.8117)	(0.8163)	(0.8166)	(0.8198)
4 weeks	0.3155	0.3028	0.3014	0.2855	0.2114	0.2043	0.2022	0.1997
	(0.7396)	(0.7500)	(0.7512)	(0.7643)	(0.8041)	(0.8107)	(0.8126)	(0.8149)
Panel B	HAR-$RV^{1/2}$				HAR-$RRV^{1/2}$			
Horizon	$RV^{1/2}$	$BPV^{1/2}$	$C^{1/2}$	$(CJ)^{1/2}$	$RRV^{1/2}$	$RBV^{1/2}$	$C^{1/2}$	$(CJ)^{1/2}$
MSE								
1 day	0.8964	0.8432	0.8503	0.8495	0.7560	0.7338	0.7295	0.7192
	(0.2600)	(0.3039)	(0.2981)	(0.2987)	(0.2994)	(0.3200)	(0.3240)	(0.3335)
1 week	0.4059	0.4026	0.3980	0.3960	0.3041	0.3036	0.3031	0.3004
	(0.6649)	(0.6677)	(0.6715)	(0.6731)	(0.7182)	(0.7187)	(0.7191)	(0.7216)
2 weeks	0.2878	0.2655	0.2678	0.2639	0.2022	0.1903	0.1900	0.1885
	(0.7624)	(0.7808)	(0.7789)	(0.7822)	(0.8126)	(0.8236)	(0.8239)	(0.8253)
3 weeks	0.2944	0.2810	0.2896	0.2869	0.1998	0.1972	0.1960	0.1945
	(0.7570)	(0.7680)	(0.7609)	(0.7632)	(0.8148)	(0.8173)	(0.8184)	(0.8198)
4 weeks	0.3121	0.2990	0.2975	0.2841	0.2091	0.2023	0.2006	0.1952
	(0.7424)	(0.7532)	(0.7544)	(0.7655)	(0.8062)	(0.8125)	(0.8141)	(0.8191)
Panel C	HAR- $\ln RV$				HAR- $\ln RRV$			
Horizon	$\ln RV$	$\ln BPV$	$\ln C$	$\ln(CJ)$	$\ln RRV$	$\ln RBV$	$\ln C$	$\ln(CJ)$
MSE								
1 day	0.9112	0.8797	0.8898	0.8887	0.7507	0.7364	0.7351	0.7302
	(0.2478)	(0.2738)	(0.2655)	(0.2664)	(0.3043)	(0.3176)	(0.3188)	(0.3233)
1 week	0.4068	0.3981	0.3998	0.3952	0.3023	0.2998	0.3000	0.2958
	(0.6642)	(0.6714)	(0.6700)	(0.6738)	(0.7199)	(0.7222)	(0.7220)	(0.7259)
2 weeks	0.3040	0.2893	0.2875	0.2818	0.2129	0.2035	0.2029	0.2040
	(0.7491)	(0.7612)	(0.7627)	(0.7674)	(0.8027)	(0.8114)	(0.8120)	(0.8110)
3 weeks	0.3026	0.2928	0.2998	0.2949	0.2034	0.2020	0.2005	0.1983
	(0.7502)	(0.7583)	(0.7525)	(0.7566)	(0.8115)	(0.8128)	(0.8142)	(0.8162)
4 weeks	0.3222	0.3118	0.3080	0.2866	0.2168	0.2111	0.2091	0.2084
	(0.7340)	(0.7426)	(0.7457)	(0.7634)	(0.7991)	(0.8044)	(0.8062)	(0.8069)

Note: The table represents MSE of the equations (17) – (24) for one day, one week through four weeks in-sample predictions of the HAR regressions of RV and RRV of S&P 500 cash index from 1995/01/01 to 2005/03/31. The different columns represent the use of different regressors. RV denotes the realized variance, RRV denotes the realized range-based variance, BVP_t (RBV_t) denotes the realized bipower (realized range bipower) variation, C denotes the continuous part of RV (RRV) as determined by the bipower test. (CJ) denotes the continuous part and the square root of the jump part that are used as separate regressors. Panel B is the model of the standard deviation. Panel C is the model of the log form. In the bipower test to separate RV (RRV) into C and J, the significant level $\alpha = 0.999$ was used. On the left side, the dependent variable is the RV for all horizons; and on the right side, the dependent variable is the RRV. The related decreasing ratios of MSE are in parenthesis.

Table 2.3 In-sample results S&P 500 1995/01/01 ~ 2005/03/31: MIDAS

Panel A	MIDAS-RV				MIDAS-RRV			
Horizon	RV	BPV	C	CJ	RRV	RBV	C	CJ
MSE								
1 day	0.9306	0.8463	0.8491	0.8449	0.7945	0.7593	0.7530	0.7197
	(0.2318)	(0.3014)	(0.2991)	(0.3025)	(0.2637)	(0.2964)	(0.3022)	(0.3331)
1 week	0.4248	0.4104	0.4026	0.4020	0.3207	0.3157	0.3141	0.3120
	(0.6493)	(0.6612)	(0.6677)	(0.6682)	(0.7028)	(0.7074)	(0.7089)	(0.7109)
2 weeks	0.2832	0.2611	0.2635	0.2583	0.1995	0.1883	0.1874	0.1849
	(0.7662)	(0.7845)	(0.7825)	(0.7868)	(0.8151)	(0.8255)	(0.8263)	(0.8287)
3 weeks	0.2905	0.2752	0.2837	0.2836	0.1994	0.1958	0.1950	0.1947
	(0.7602)	(0.7728)	(0.7658)	(0.7659)	(0.8152)	(0.8186)	(0.8193)	(0.8196)
4 weeks	0.3075	0.2932	0.2929	0.2925	0.2053	0.1976	0.1958	0.1956
	(0.7462)	(0.7580)	(0.7582)	(0.7585)	(0.8097)	(0.8169)	(0.8186)	(0.8187)
Panel B	MIDAS-RV$^{1/2}$				MIDAS-RRV$^{1/2}$			
Horizon	RV$^{1/2}$	BPV$^{1/2}$	C$^{1/2}$	(CJ)$^{1/2}$	RRV$^{1/2}$	RBV$^{1/2}$	C$^{1/2}$	(CJ)$^{1/2}$
MSE								
1 day	0.9007	0.8485	0.8545	0.8536	0.7599	0.7382	0.7342	0.7258
	(0.2565)	(0.2996)	(0.2946)	(0.2954)	(0.2958)	(0.3159)	(0.3196)	(0.3274)
1 week	0.4035	0.3970	0.3933	0.3931	0.3030	0.3015	0.3012	0.3009
	(0.6669)	(0.6723)	(0.6753)	(0.6755)	(0.7192)	(0.7206)	(0.7209)	(0.7212)
2 weeks	0.2806	0.2600	0.2604	0.2555	0.1977	0.1884	0.1877	0.1869
	(0.7684)	(0.7854)	(0.7850)	(0.7891)	(0.8168)	(0.8254)	(0.8261)	(0.8268)
3 weeks	0.2902	0.2773	0.2854	0.2855	0.1977	0.1954	0.1938	0.1902
	(0.7604)	(0.7711)	(0.7644)	(0.7643)	(0.8168)	(0.8189)	(0.8204)	(0.8237)
4 weeks	0.3057	0.2883	0.2890	0.2771	0.2044	0.1976	0.1965	0.1949
	(0.7476)	(0.7620)	(0.7614)	(0.7713)	(0.8106)	(0.8169)	(0.8179)	(0.8194)
Panel C	MIDAS- $\ln RV$				MIDAS- $\ln RRV$			
Horizon	$\ln RV$	$\ln BPV$	$\ln C$	$\ln(CJ)$	$\ln RRV$	$\ln RBV$	$\ln C$	$\ln(CJ)$
MSE								
1 day	0.9130	0.8814	0.8907	0.8893	0.7539	0.7399	0.7387	0.7351
	(0.2463)	(0.2724)	(0.2647)	(0.2659)	(0.3014)	(0.3143)	(0.3154)	(0.3188)
1 week	0.4039	0.3935	0.3958	0.3898	0.3010	0.2982	0.2984	0.2958
	(0.6666)	(0.6752)	(0.6733)	(0.6782)	(0.7211)	(0.7237)	(0.7235)	(0.7259)
2 weeks	0.3048	0.2900	0.2887	0.2835	0.2118	0.2034	0.2029	0.1986
	(0.7484)	(0.7606)	(0.7617)	(0.7660)	(0.8037)	(0.8115)	(0.8120)	(0.8160)
3 weeks	0.3016	0.2932	0.3014	0.3023	0.2041	0.2035	0.2025	0.1997
	(0.7510)	(0.7580)	(0.7512)	(0.7505)	(0.8109)	(0.8114)	(0.8123)	(0.8149)
4 weeks	0.3288	0.3189	0.3152	0.2804	0.2245	0.2179	0.2164	0.1953
	(0.7286)	(0.7368)	(0.7398)	(0.7685)	(0.7920)	(0.7981)	(0.7995)	(0.8190)

Note: The table represents MSE of the equations (17) – (24) for one day, one week through four weeks in-sample predictions of the MIDAS regressions of RV and RRV of S&P 500 cash index from 1995/01/01 to 2005/03/31. The different columns represent the use of different regressors. The related decreasing ratios of MSE are in parenthesis. See Table 2 for further details.

Table 2.4 In-sample results S&P 500 1995/01/01 ~ 2005/03/31
Chow test for conditional HAR regressions

Panel A	HAR-RV			HAR-RRV		
Horizon	RV	BPV	C	RRV	RBV	C
$F-stat$ and $p-value$						
1 day	52.6680	1.7142	1.0365	59.9671	32.2231	30.0457
	(0.0000)	(0.1440)	(0.3868)	(0.0000)	(0.0000)	(0.0000)
1 week	1.8256	2.3656	2.1403	14.3689	15.5310	15.2141
	(0.1226)	(0.0520)	(0.0747)	(0.0000)	(0.0000)	(0.0000)
2 weeks	2.7435	0.5548	0.5413	3.6889	1.8788	1.6905
	(0.0292)	(0.6957)	(0.7055)	(0.0062)	(0.1148)	(0.1528)
3 weeks	0.7533	0.2970	0.3210	1.0382	0.6336	0.4908
	(0.5572)	(0.8796)	(0.8636)	(0.3893)	(0.6393)	(0.7425)
4 weeks	0.6198	0.1298	0.1385	1.7564	0.9276	0.8743
	(0.6493)	(0.9713)	(0.9677)	(0.1423)	(0.4505)	(0.4817)
Panel B	HAR-RV$^{1/2}$			HAR-RRV$^{1/2}$		
Horizon	RV$^{1/2}$	BPV$^{1/2}$	C$^{1/2}$	RRV$^{1/2}$	RBV$^{1/2}$	C$^{1/2}$
$F-stat$ and $p-value$						
1 day	26.6570	1.3136	0.5917	32.6609	10.9828	8.2993
	(0.0000)	(0.2625)	(0.6687)	(0.0000)	(0.0000)	(0.0000)
1 week	1.6470	0.9922	0.9285	7.2790	6.3360	6.1991
	(0.1612)	(0.4113)	(0.4470)	(0.0000)	(0.0001)	(0.0001)
2 weeks	2.4738	0.4077	0.3580	4.1786	1.7955	1.3640
	(0.0451)	(0.8031)	(0.8384)	(0.0027)	(0.1303)	(0.2470)
3 weeks	0.9541	0.2664	0.2282	0.6741	0.0904	0.0475
	(0.4345)	(0.8992)	(0.9223)	(0.6109)	(0.9854)	(0.9957)
4 weeks	0.7671	0.1276	0.1057	1.9653	0.9361	0.8101
	(0.5487)	(0.9722)	(0.9803)	(0.1043)	(0.4457)	(0.5211)
Panel C	HAR- $\ln RV$			HAR- $\ln RRV$		
Horizon	$\ln RV$	$\ln BPV$	$\ln C$	$\ln RRV$	$\ln RBV$	$\ln C$
$F-stat$ and $p-value$						
1 day	16.4395	1.1144	0.9940	15.3766	2.8791	1.7581
	(0.0000)	(0.3479)	(0.4095)	(0.0000)	(0.0215)	(0.1345)
1 week	1.5409	0.0340	0.0445	4.5171	2.3127	2.0388
	(0.1891)	(0.9978)	(0.9963)	(0.0014)	(0.0567)	(0.0878)
2 weeks	2.0553	0.4506	0.4645	3.6827	1.4817	0.9904
	(0.0873)	(0.7719)	(0.7618)	(0.0062)	(0.2083)	(0.4134)
3 weeks	1.3271	0.2565	0.1589	1.5545	0.6619	0.5296
	(0.2622)	(0.9053)	(0.9587)	(0.1891)	(0.6194)	(0.7142)
4 weeks	0.6590	0.0862	0.0473	1.6982	0.8170	0.6134
	(0.6217)	(0.9866)	(0.9957)	(0.1551)	(0.5168)	(0.6538)

Note: The table is the Chow test of the F -statistics and $p-value$ for the test of the hypothesis that the jump dummies are zero for one day, one week through four weeks in-sample predictions of the conditional HAR regression for RV and RRV of S&P 500 cash index from 1995/01/01 to 2005/03/31. The $p-value$ are in parenthesis.
See Table 2 for further details.

Table 2.5 In-sample results S&P 500 1995/01/01 ~ 2005/03/31
Using conditional HAR regressions

Panel A	HAR-RV			HAR-RRV		
Horizon	RV	BPV	C	RRV	RBV	C
MSE						
1 day	0.8540	0.8371	0.8415	0.7902	0.7545	0.7478
	(0.2950)	(0.3090)	(0.3053)	(0.2677)	(0.3008)	(0.3070)
1 week	0.4194	0.4100	0.4021	0.3198	0.3167	0.3151
	(0.6538)	(0.6615)	(0.6681)	(0.7036)	(0.7065)	(0.7080)
2 weeks	0.2803	0.2660	0.2712	0.2033	0.1897	0.1895
	(0.7686)	(0.7804)	(0.7761)	(0.8116)	(0.8242)	(0.8244)
3 weeks	0.2946	0.2810	0.2908	0.2032	0.1982	0.1979
	(0.7568)	(0.7680)	(0.7599)	(0.8117)	(0.8163)	(0.8166)
4 weeks	0.3089	0.3015	0.2999	0.2114	0.2043	0.2022
	(0.7450)	(0.7511)	(0.7524)	(0.8041)	(0.8107)	(0.8126)
Panel B	HAR-$RV^{1/2}$			HAR-$RRV^{1/2}$		
Horizon	$RV^{1/2}$	$BPV^{1/2}$	$C^{1/2}$	$RRV^{1/2}$	$RBV^{1/2}$	$C^{1/2}$
MSE						
1 day	0.8964	0.8432	0.8503	0.7560	0.7338	0.7295
	(0.2600)	(0.3039)	(0.2981)	(0.2994)	(0.3200)	(0.3240)
1 week	0.4059	0.4026	0.3980	0.3041	0.3036	0.3031
	(0.6649)	(0.6677)	(0.6715)	(0.7182)	(0.7187)	(0.7191)
2 weeks	0.2878	0.2655	0.2678	0.2022	0.1903	0.1900
	(0.7624)	(0.7808)	(0.7789)	(0.8126)	(0.8236)	(0.8239)
3 weeks	0.2944	0.2810	0.2896	0.1998	0.1972	0.1960
	(0.7570)	(0.7680)	(0.7609)	(0.8148)	(0.8173)	(0.8184)
4 weeks	0.3121	0.2990	0.2975	0.2091	0.2023	0.2006
	(0.7424)	(0.7532)	(0.7544)	(0.8062)	(0.8125)	(0.8141)
Panel C	HAR- $\ln RV$			HAR- $\ln RRV$		
Horizon	$\ln RV$	$\ln BPV$	$\ln C$	$\ln RRV$	$\ln RBV$	$\ln C$
MSE						
1 day	0.9112	0.8797	0.8898	0.7507	0.7364	0.7351
	(0.2478)	(0.2738)	(0.2655)	(0.3043)	(0.3176)	(0.3188)
1 week	0.4068	0.3981	0.3998	0.3023	0.2998	0.3000
	(0.6642)	(0.6714)	(0.6700)	(0.7199)	(0.7222)	(0.7220)
2 weeks	0.3040	0.2893	0.2875	0.2129	0.2035	0.2029
	(0.7491)	(0.7612)	(0.7627)	(0.8027)	(0.8114)	(0.8120)
3 weeks	0.3026	0.2928	0.2998	0.2034	0.2020	0.2005
	(0.7502)	(0.7583)	(0.7525)	(0.8115)	(0.8128)	(0.8142)
4 weeks	0.3222	0.3118	0.3080	0.2168	0.2111	0.2091
	(0.7340)	(0.7426)	(0.7457)	(0.7991)	(0.8044)	(0.8062)

Note: The table represents MSE of the equations (29) for one day, one week through four weeks in-sample predictions of the HAR regressions of RV and RRV of S&P 500 cash index from 1995/01/01 to 2005/03/31. The related decreasing ratios of MSE are in parenthesis.

See Table 2 for further details.

Table 2.6 Out-of-sample forecasts of S&P 500 1995/01/01 ~ 2005/03/31: HAR

Panel A	HAR-RV				HAR-RRV			
Horizon	RV	BPV	C	CJ	RRV	RBV	C	CJ
MSE								
1 day	0.0295	0.0247	0.0221	0.0217	0.0187	0.0169	0.0167	0.0187
1 week	0.0246	0.0203	0.0157	0.0162	0.0155	0.0149	0.0147	0.0171
2 weeks	0.0295	0.0246	0.0203	0.0194	0.0176	0.0169	0.0164	0.0183
3 weeks	0.0470	0.0374	0.0313	0.0293	0.0275	0.0253	0.0248	0.0304
4 weeks	0.0502	0.0396	0.0358	0.0321	0.0287	0.0270	0.0263	0.0274

Panel B	HAR-RV$^{1/2}$				HAR-RRV$^{1/2}$			
Horizon	RV$^{1/2}$	BPV$^{1/2}$	C$^{1/2}$	(CJ)$^{1/2}$	RRV$^{1/2}$	RBV$^{1/2}$	C$^{1/2}$	(CJ)$^{1/2}$
MSE								
1 day	0.0214	0.0198	0.0193	0.0193	0.0128	0.0122	0.0122	0.0126
1 week	0.0103	0.0088	0.0079	0.0083	0.0078	0.0078	0.0078	0.0080
2 weeks	0.0104	0.0088	0.0078	0.0079	0.0081	0.0084	0.0084	0.0085
3 weeks	0.0168	0.0124	0.0111	0.0113	0.0120	0.0113	0.0113	0.0123
4 weeks	0.0167	0.0144	0.0134	0.0132	0.0118	0.0122	0.0122	0.0131

Panel C	HAR- $\ln RV$				HAR- $\ln RRV$			
Horizon	$\ln RV$	$\ln BPV$	$\ln C$	$\ln(CJ)$	$\ln RRV$	$\ln RBV$	$\ln C$	$\ln(CJ)$
MSE								
1 day	0.0210	0.0199	0.0195	0.0195	0.0128	0.0122	0.0122	0.0122
1 week	0.0083	0.0078	0.0072	0.0074	0.0068	0.0070	0.0071	0.0072
2 weeks	0.0072	0.0071	0.0065	0.0064	0.0065	0.0071	0.0073	0.0076
3 weeks	0.0089	0.0081	0.0074	0.0072	0.0077	0.0081	0.0081	0.0088
4 weeks	0.0097	0.0097	0.0090	0.0090	0.0081	0.0090	0.0092	0.0100

Note: The table represents MSE, for one day, one week through four weeks in-sample predictions of the HAR regressions of RV and RRV, of the out-of-sample forecasts of the S&P 500 cash index from September 16, 2004 to March 31, 2005. Data from January 1, 1995 to March 31, 2005 was used to estimate the parameters of the models. See Table 2 for further details.

Table 2.7 Out-of-sample forecasts of S&P 500 1995/01/01 ~ 2005/03/31: MIDAS

Panel A	MIDAS-RV				MIDAS-RRV			
Horizon	RV	BPV	C	CJ	RRV	RBV	C	CJ
MSE								
1 day	0.0256	0.0273	0.0325	0.0219	0.0191	0.0217	0.0214	0.0186
1 week	0.0093	0.0064	0.0083	0.0169	0.0091	0.0091	0.0088	0.0168
2 weeks	0.0097	0.0051	0.0056	0.0187	0.0078	0.0071	0.0070	0.0193
3 weeks	0.0254	0.0062	0.0044	0.0298	0.0097	0.0040	0.0040	0.0297
4 weeks	0.0146	0.0066	0.0049	0.0331	0.0048	0.0039	0.0039	0.0270
Panel B	MIDAS-$RV^{1/2}$				MIDAS-$RRV^{1/2}$			
Horizon	$RV^{1/2}$	$BPV^{1/2}$	$C^{1/2}$	$(CJ)^{1/2}$	$RRV^{1/2}$	$RBV^{1/2}$	$C^{1/2}$	$(CJ)^{1/2}$
MSE								
1 day	0.0215	0.0198	0.0193	0.0193	0.0129	0.0122	0.0123	0.0127
1 week	0.0106	0.0091	0.0082	0.0082	0.0079	0.0081	0.0081	0.0085
2 weeks	0.0113	0.0096	0.0087	0.0086	0.0086	0.0090	0.0090	0.0095
3 weeks	0.0193	0.0137	0.0125	0.0124	0.0136	0.0124	0.0124	0.0131
4 weeks	0.0183	0.0162	0.0155	0.0150	0.0129	0.0132	0.0132	0.0153
Panel C	MIDAS- $\ln RV$				MIDAS- $\ln RRV$			
Horizon	$\ln RV$	$\ln BPV$	$\ln C$	$\ln(CJ)$	$\ln RRV$	$\ln RBV$	$\ln C$	$\ln(CJ)$
MSE								
1 day	0.0209	0.0199	0.0194	0.0195	0.0128	0.0122	0.0123	0.0122
1 week	0.0085	0.0080	0.0074	0.0075	0.0070	0.0073	0.0074	0.0076
2 weeks	0.0078	0.0076	0.0071	0.0069	0.0071	0.0078	0.0080	0.0084
3 weeks	0.0096	0.0088	0.0082	0.0077	0.0084	0.0089	0.0090	0.0097
4 weeks	0.0106	0.0103	0.0099	0.0087	0.0089	0.0096	0.0097	0.0106

Note: The table represents MSE, for one day, one week through four weeks in-sample predictions of the MIDAS regressions of RV and RRV, of the out-of-sample forecasts of the S&P 500 cash index from September 16, 2004 to March 31, 2005. Data from January 1, 1995 to March 31, 2005 was used to estimate the parameters of the models. See Tables 2 for further details.

Chapter 3 The Information Content of Implied Volatility in the presence of the Continuous Components, and the Jump Components of Realized Range Volatility

1 Introduction

Most of the previous studies have documented the information content of the implied volatility. They always focus on whether the implied volatility has the additional information content of historical volatility; the realized volatility (RV) is always used as the historical volatility. Giot and Laurent (2007) considered the information content of implied volatility in the continuous and jump components of the realized volatility, whose decomposition was suggested by Barndroff-Nielsen and Shephard (2004), using the encompassing regressions. Because the realized range-based estimation of the integrated variance has been proved to be more efficient, we will use the realized range-based volatility to measure the historical volatility. We will employ the heterogeneous autoregressive (HAR) regressions by Corsi (2004) and mixed data sampling (MIDAS) regressions by Ghysels et al. (2006) as encompassing regressions to examine the information content of the continuous and jump components of the realized range-based volatility (RRV), and the additional information content of the implied volatility as an additional regressor. In addition, this study focuses on the S&P 500 index, hence, we use the Chicago Board Options Exchange (CBOE) volatility index new VIX as the measure of the implied volatility. The new VIX is based on S&P 500 index options and adopts the model-free volatility expectation.

The results show that the implied volatility and almost all continuous components are statistically significant, while the jump components are almost not significant. The implied volatility has a high information content and the continuous components of the past realized range-based volatility feature relevant information content by the implied volatility. Beside, the jump components do not contribute to future valuable information.

In addition, except for $h=1$ horizon, the implied volatility and the out-of-sample volatility have information contents but the implied volatility has more powerful explanation abilities than the out-of-sample volatility for the future realized range volatility.

The remainder of this chapter is organized as follows. In section 2, we discuss the volatility measure, predict volatility regressions and cover the models we are going to use. In section 3, we present the data and the empirical results. Section 4 concludes the article.

2 The Methodology

2.1 Construction of volatility measures

Let the logarithmic price of financial assets at time t be denoted by $p(t)$ and follow the continuous-time jump diffusion process

$$dp(t)=\mu(t)dt+\sigma(t)dw(t)+\kappa(t)dq(t)\,, \qquad (1)$$

where $\mu(t)$ and $\sigma(t)$ are the drift and instantaneous volatility, $w(t)$ is the standardized Brownian motion, $q(t)$ is a counting process with time-varying intensity $\lambda(t)$, *i.e.* $P[dq(t)=1]=\lambda(t)dt$, and $\kappa(t)$ is the jump size. The quadratic variation process for a sequence of partitions is defined by

$$[p](t)=\text{plim}\sum_{j=0}^{n-1}(p(s_{j+1})-p(s_j))^2\,,$$

where $0=s_0<s_1<\cdots<s_n=t$ and $\sup_j\{s_{j+1}-s_j\}\to 0$ for $n\to\infty$.

When the data is sampled at a higher frequency, M times in a day, we will denote the intraday ranges as:

$$s_{pi\Delta,\Delta,m} = \max_{(i-1)/M \le s,t \le i/M} \{p_t - p_s\},$$

where $s_{pi\Delta,\Delta,m}$ represents that each range is based on the corresponding m returns. The realized ranged-based variance over day t is defined as

$$RRV_t^{m*} = \frac{1}{\lambda_{2,m}} \sum_{i=1}^{M} s_{pi\Delta,\Delta,m}^2, \tag{2}$$

where $\lambda_{j,m} = E\left(s_{W,m}^j\right) < \infty$, $\forall j \in R$ and $m \in N$ [5], and W is the Brownian motion. As noted by Christensen and Podolskij (2006a, 2006b),

$$RRV_t^{m*} \xrightarrow{P} \int_0^t \sigma^2(s)ds + \frac{1}{\lambda_{2,m}} \cdot \sum_{i=1}^{M} J_i^2, \text{ as } M \to \infty,$$

i.e., RRV_t^{m*} is inconsistent. Hence, they modified the intraday high-low statistic to make it consistent with the quadratic variation. The realized range-based bipower variation with parameter $(r,s) \in R_+^2$ is defined as[6]:

$$RBV_{(r,s),t}^m = M^{(r+s)/2-1} \frac{1}{\lambda_{r,m}} \frac{1}{\lambda_{s,m}} \sum_{i=1}^{M-1} s_{pi\Delta,\Delta,m}^r s_{p(i+1)\Delta,\Delta,m}^s,$$

then $RBV_{(1,1),t}^m \equiv RBV_t^m \xrightarrow{p} \int_0^t \sigma^2(s)ds$, as $M \to \infty$. Hence, the new range based estimator, $RRV_t^m \equiv \lambda_{2,m} RRV_t^{m*} + (1-\lambda_{2,m}) RBV_t^m$, is consistent for quadratic variation, *i.e.*, $RRV_t^m \xrightarrow{P} \int_0^t \sigma^2(s)ds + \sum_{i=1}^{M} J_i^2$.

Furthermore, using the bivariate distribution of (RRV_t^m, RBV_t^m) and the delta-method, Christensen and Podolskij (2006b) found the jump detection statistic,

$$\frac{\sqrt{M}(RRV_t^m - RBV_t^m)}{\sqrt{\nu_m \int_0^t \sigma(s)_u^4 ds}} \xrightarrow{d} N(0,1),$$

5 There is no explicit formula for $\lambda_{r,m}$, but it is computed to any degree of accuracy from simulations.

6 I maintain some notations used by Christensen and Podolskij (2006b) throughout the chapter.

where $\nu_m = \lambda_{2,m}^2(\Lambda_m^R + \Lambda_m^B - 2\Lambda_m^{RB})$, $\Lambda_m^R = \dfrac{\lambda_{4,m} - \lambda_{2,m}^2}{\lambda_{2,m}^2}$, $\Lambda_m^B = \dfrac{\lambda_{2,m}^2 + 2\lambda_{1,m}^2\lambda_{2,m} - 3\lambda_{1,m}^4}{\lambda_{1,m}^4}$, and

$\Lambda_c^{RB} = \dfrac{2\lambda_{3,c}\lambda_{1,c} - 2\lambda_{2.c}\lambda_{1,c}^2}{\lambda_{2,c}\lambda_{1,c}^2}$. In addition, they adopted the modified ratio-statistic to improve the size properties in finite samples. The modified ratio-statistic is

$$Z_{t+1} = \frac{\sqrt{M}\left(1 - {RBV_t^m}/{RRV_t^{*m}}\right)}{\sqrt{\nu_m \max\left\{{RQQ_t^m}/{(RBV_t^{*m})^2}, {1}/{t}\right\}}} \xrightarrow{d} N(0,1), \tag{3}$$

where $RQQ_t^m = \frac{M}{\lambda_{1,m}^4}\sum_{i=1}^{M-3} s_{pi\Delta,\Delta,m}s_{p(i+1)\Delta,\Delta,m}s_{p(i+2)\Delta,\Delta,m}s_{p(i+3)\Delta,\Delta,m}$ and $RQQ_t^m \xrightarrow{p} \int_0^t \sigma^4(s)ds$.

Huang and Tauchen (2005) found that the statistics in equation (3) also had a sensible power against other empirically calibrated stochastic volatility jump diffusion models. Using equation (3), Andersen, Bollerslev, and Diebold (2007) identified the jump variation as,

$$J_{t,t+1} = I(Z_{t+1} > \Phi_\alpha)(RRV_{t,t+1}^m - RBV_{t,t+1}^m), \tag{4}$$

and the continuous component variation was estimated as the residual,

$$C_{t,t+1} = I(Z_{t+1} \le \Phi_\alpha)\cdot RRV_{t,t+1}^m + I(Z_{t+1} > \Phi_\alpha)\cdot RBV_{t,t+1}^m, \tag{5}$$

where $I(\cdot)$ denotes the indicator function, $\Phi(\cdot)$ is the standard cumulated normal distribution ($\Phi_\alpha = \Phi^{-1}(\alpha)$), and α is a significant level of the bipower test and we will use $\alpha = 0.999$ throughout the article. From the definitions in equations (4) and (5), we ensure that the continuous variation and jump variation sum to the total realized variation, *i.e.* $RRV_{t,t+1}^m = C_{t,t+1} + J_{t,t+1}$.

2.2 Predicting volatility models

In this study, we will exploit MIDAS regression model, which was introduced by Ghysels et al. (2002, 2006), and HAR regression model, which was suggested by

Corsi (2004), to predict volatility from t to $t+H$, where H is the predicting horizon in days.

The multi-period realized variances that were constructed by Andersen et al. (2007). Similarly, the multi-period realized range-based varinaces were defined as the normalized sum of the one-period realized range variances,

$$RRV_{t,t+H} = H^{-1}(RRV_{t,t+1} + RRV_{t,t+2} + \cdots + RRV_{t,t+H}),$$

where $H = 1,\ 5,\ 10,\ 15,\ \text{and}\ 20$. H is the prediction horizon in days, in the empirical analysis, as one day, weekly, bi-weekly, tri-weekly, and monthly.

The HAR-RV models were introduced by Corsi (2004) and they can capture the long memory property of the realized variance. Similarly, the HAR-RRV model is written as follows

$$RRV_{t,t+H} = \alpha_0 + \alpha_D X_{t-1,t} + \alpha_W X_{t-5,t} + \alpha_M X_{t-20,t} + \varepsilon_{t,t+1},$$

where $RRV_{t,t+H}$ represents the future RRV, using the HAR regressions.

Andersen et al. (2007) defined the HAR-RV-CJ model, which explores the separation of RV_t into the continuous part C_t and jump part J_t. This separation was suggested by Barndorff-Nielsen et al. (2004). Following Andersen et al.'s model, the HAR-RRV-CJ is shown as below

$$\begin{aligned} RRV_{t,t+H} &= \alpha_0 + \alpha_{CD} C_{t-1,t} + \alpha C_{CW} C_{t-5,t} + \alpha_{CM} C_{t-20,t} \\ &\quad + \alpha_{JD} J_{t-1,t} + \alpha_{JW} J_{t-5,t} + \alpha_{JM} J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned}.$$

As noted by Andersen et al. (2001) , the log form probability density of the error term is close to the normal density, we will consider the HAR-ln RRV-ln CJ model in this chapter. The models are as follows,

$$\begin{aligned} \ln RRV_{t,t+H} &= \alpha_0 + \alpha_{CD} \ln C_{t-1,t} + \alpha_{CW} \ln C_{t-5,t} + \alpha_{CM} \ln C_{t-20,t} \\ &\quad + \alpha_{JD} \ln J_{t-1,t} + \alpha_{JW} \ln J_{t-5,t} + \alpha_{JM} \ln J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned}, \tag{6}$$

where $\ln X_{t,t+H} = H^{-1}\left(\ln X_{t+1} + \ldots + \ln X_{t+H}\right)$.

The differences between MIDAS and HAR regressions models are the lagged regressors and their weights. The MIDAS regression was introduced by Ghysels et al. (2002, 2005). MIDAS regressions can run parsimoniously parameterized regressors of data the observed at different frequencies. Ghysels et al. (2006) used the MIDAS regressions to predict volatility and we follow him. The MIDAS-RRV models can be written as

$$RRV_{t,t+H} = \mu_H + \phi_H \sum_{k=0}^{k_{\max}} b(k,\theta_1,\theta_2) X_{t-k-1,t-k} + \varepsilon_{t+H} \,,$$

where $RRV_{t,t+H}$ represents the future RRV, using the MIDAS regressions.

$$b(k;\theta_1,\theta_2) = \frac{f(k/k_{\max};\theta_1,\theta_2)}{\sum_{k=1}^{k_{\max}} f(k/k_{\max};\theta_1,\theta_2)},\ f(x;\theta_1,\theta_2) = \frac{x^{\theta_1-1}(1-x)^{\theta_2-1}}{\beta(\theta_1,\theta_2)} \quad ,$$

and $\beta(\theta_1,\theta_2) = \dfrac{\Gamma(\theta_1)\,\Gamma(\theta_2)}{\Gamma(\theta_1+\theta_1)}$. For all the MIDAS regressions, we use $k_{\max} = 50$ [7], and we fix θ_1 =1[8]. Similarly as the above, we will consider the MIDAS- ln RV- ln CJ model, which is written as follows,

$$\begin{aligned} \ln(RRV_{t,t+H}) = \mu_H + \phi_1 \sum_{k=0}^{k_{\max}} b_C(k,1,\theta_{2C}) \ln(C_{t-k-1,t-k}) \\ + \phi_2 \sum_{k=0}^{k_{\max}} b_J(k,1,\theta_{2J}) \ln(J_{t-k-1,t-k}) + \varepsilon_{t+H} \end{aligned} . \qquad (7)$$

2.3 The models

We will use the HAR and MIDAS regressions as encompassing regressions to examine the information content of the continuous and jump components of the realized range volatility and the additional information content of the implied volatility as an additional regressor. The models are shown as follows,

Model 1: Equation (6) and (7).

---

[7] Ghysels (2006) showed that using longer lags (*i.e.*, $k > 50$) has little effect on the results.

[8] This will give us declining weights in the lag polynominal.

Model 2:

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_{IV} \ln(IV_{t-1,t}) + \varepsilon_{t,t+H} \ , \tag{8}$$

Model 3:

$$\begin{aligned} \ln RRV_{t,t+H} &= \alpha_0 + \alpha_{IV} \ln(IV_{t-1,t}) + \alpha_{CD} \ln C_{t-1,t} + \alpha_{CW} \ln C_{t-5,t} + \alpha_{CM} \ln C_{t-20,t} \\ &+ \alpha_{JD} \ln J_{t-1,t} + \alpha_{JW} \ln J_{t-5,t} + \alpha_{JM} \ln J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned} , \tag{9}$$

$$\begin{aligned} \ln RRV_{t,t+H} &= \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) + \phi_1 \sum_{k=0}^{k_{\max}} b_C(k,1,\theta_{2C}) \ln(C_{t-k-1,t-k}) \\ &+ \phi_2 \sum_{k=0}^{k_{\max}} b_J(k,1,\theta_{2J}) \ln(J_{t-k-1,t-k}) + \varepsilon_{t+H} \end{aligned} . \tag{10}$$

Model 4:

$$\begin{aligned} \ln RRV_{t,t+H} &= \alpha_0 + \alpha_{IV} \ln(IV_{t-1,t}) \\ &+ \alpha_{JD} \ln J_{t-1,t} + \alpha_{JW} \ln J_{t-5,t} + \alpha_{JM} \ln J_{t-20,t} + \varepsilon_{t,t+H} \end{aligned} , \tag{11}$$

$$\begin{aligned} \ln RRV_{t,t+H} &= \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) \\ &+ \phi_2 \sum_{k=0}^{k_{\max}} b_J(k,1,\theta_{2J}) \ln(J_{t-k-1,t-k}) + \varepsilon_{t+H} \end{aligned} . \tag{12}$$

Model 5:

$$\begin{aligned} \ln RRV_{t,t+H} &= \alpha_0 + \alpha_{IV} \ln(IV_{t-1,t}) \\ &+ \alpha_{CD} \ln C_{t-1,t} + \alpha_{CW} \ln C_{t-5,t} + \alpha_{CM} \ln C_{t-20,t} + \varepsilon_{t,t+H} \end{aligned} , \tag{13}$$

$$\begin{aligned} \ln RRV_{t,t+H} &= \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) \\ &+ \phi_1 \sum_{k=0}^{k_{\max}} b_C(k,1,\theta_{2C}) \ln(C_{t-k-1,t-k}) + \varepsilon_{t+H} \end{aligned} . \tag{14}$$

2.4 Implied volatility and out-of-sample volatility

Regarding the HAR-RRV-CJ and MIDAS-RRV-CJ[9] volatility forecasts, we employ the out-of-sample estimation approach to generate the H-day volatility forecasts. We explore the rolling window analysis to predict the out-of-sample forecasts. The rolling window width is 1236 observations (1995/01/01 ~ 1999/12/31) and the windows are rolled through the sample once at a time; there will be 1299 rolling estimates for each

[9] Using HAR (MIDAS) regressions, we model RRV with C (the continuous components) and J (the jump components) regressors.

parameter. Similarly, there will be 1295 (one week), 1285 (two weeks), and 1280 (one month) rolling estimates for each parameter respectively.

Then, we explore the information content of the implied volatility and the out-of-sample volatility in the light of the future realized range volatility. The model is shown as follows,

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_1 \ln IV_{t-1,t} + \varepsilon_{t,t+H} , \tag{15}$$

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_2 \ln \widehat{RRV}_{t,t+H} + \varepsilon_{t,t+H} , \tag{16}$$

and

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_1 \ln IV_{t-1,t} + \alpha_2 \ln \widehat{RRV}_{t,t+H} + \varepsilon_{t,t+H} , \tag{17}$$

where $\ln \widehat{RRV}_{t+1,t+H}$ is the out-of-sample estimation of the HAR-RRV-CJ (or MIDAS-RRV-CJ) volatility forecast.

3 Data and empirical results

3.1 Data descriptions

This study employs 5-minute intra-day data of the S&P 500 index securities. The intraday data are obtained from Tick Data Inc., covering the period from January 1, 1995 to March 31, 2005 and consisting of 2535 days with 78 intra-day 5-minute observations. The daily data of the VIX implied volatility index are supplied by the CBOE. The time period is from January 1, 1995 to March 31, 2005.

3.2 Empirical results

Table 1 and Table 2 report the results of encompassing regressions, using the HAR and MIDAS regressions respectively. These two encompassing regressions provide almost the same results. Besides, in our empirical results, when we use the HAR regressions, the adjusted R^2 is larger, *i.e.*, their regressors have more powerful explanation abilities.

Results of Model 1, without the implied volatility, are that the continuous components of the past realized range volatility provide major information for the future realized range volatility (except for $h = 20$ horizon, the weekly continuous component is not significant), and the jump components are almost not significant (except for $h = 1$ horizon, the daily jump component is significant).

In Model 2, the sole regressor of implied volatility is concerned. The coefficients of the log implied volatility are statistically significant and almost close to 1, *i.e.*, the future realized range-based volatility co-moves almost one-to-one with the implied volatility. Because the adjusted R^2s are large, the implied volatility has a high information content.

In Model 4, the jump components are added to Model 2. There are barely increases with the adjusted R^2s and the jump components are almost not significant. This means that the implied volatility has a high information content and the jump components do not contribute to future valuable information.

In Model 5, the continuous components are added to Model 2; the implied volatility and almost all continuous components are statistically significant. Since there are few increases with the adjusted R^2s, the total explanation powers of the encompassing regression are scarcely enhanced. Hence, the implied volatility subsumes more information than the continuous components of the past realized range volatility. Similarly, different from Model 2, Model 3 includes all our possible regressors: implied volatility, the jump/continuous components of the past realized range volatility. The implied volatility and almost all continuous components are statistically significant, but the jump components are almost not significant and there are little increases with the adjusted R^2s. Therefore, the implied volatility has a high information content. Moreover, from Model 2 to Model 5 (or 3), the coefficients of the implied volatility (α_{IV}) are decreasing and the continuous components are statistically significant. This means the past continuous components feature relevant information content by the implied volatility.

Table 3 reports the information content of the implied volatility and the out-of-sample volatility, which are the HAR-RRV-CJ (or MIDAS-RRV-CJ) volatility forecast in the light of the future realized range volatility. We split the data into two parts: a sample part to estimate these two models and an out-of-sample part for forecasting. In this study, we explore the rolling window analysis to predict the out-of-sample forecasting. The rolling window's width is 1236 observations and the windows are rolled through the sample once at a time. Hence, the first in-sample period covers from January 1, 1995 to December 31, 1999. We find that all coefficients are statistically significant for all horizons. Only $h = 1$ horizon, the coefficients of the out-of-sample volatilities are larger than the implied volatilities. The smaller adjusted R^2 represents that the regression has noise. Besides, when the implied volatility is added to the out-of-sample volatility, there are few increases with the adjusted R^2s. The total explanation powers of the encompassing regression are scarcely enhanced. Hence, the out-of-sample volatility subsumes more information than the implied volatility. For other horizons, the coefficients of the implied volatilities are larger than the out-of-sample volatilities and the adjusted R^2s are larger. In contrast, the adjusted R^2s of the sole regressor of the implied volatility are greater than the sole regressor of the out-of-sample volatility. When the out-of-sample volatility is added to the implied volatility, there are few increases with the adjusted R^2s. The total explanation powers of the encompassing regression are scarcely enhanced. Hence, both regressors have information contents but the implied volatility has more powerful explanation abilities than the out-of-sample volatility for the future realized range volatility.

4 Conclusions

The purpose of this chapter is to examine the information content of the continuous and jump components of the realized range-based volatility, and the additional information content of the implied volatility as an additional regressor. We use the model-free implied volatility, VIX, as the measure of the implied volatility. For our empirical results and based on the HAR and MIDAS regressions as an encompassing regression analysis, we find that the implied volatility has a high information content and the past continuous components feature relevant information content by the implied volatility. Besides, the jump components do not contribute to future valuable information. Furthermore, the implied volatility and the out-of-sample volatility have information contents for the future realized volatility, but the implied volatility has more powerful explanation abilities than the out-of-sample volatility for the future realized range volatility (except for $h=1$ horizon).

Table 3.1 Encompassing Regression: HAR

S&P 500 index 1995/01/01 ~ 2005/03/31

Model 1: $\ln RRV_{t,t+H} = \alpha_0 + \alpha_{CD}\ln C_{t-1,t} + \alpha_{CW}\ln C_{t-5,t} + \alpha_{CM}\ln C_{t-20,t} + \alpha_{JD}\ln J_{t-1,t} + \alpha_{JW}\ln J_{t-5,t} + \alpha_{JM}\ln J_{t-20,t} + \varepsilon_{t,t+H}$

Model 2: $\ln(RRV_{t,t+H}) = \alpha_0 + \alpha_{IV}\ln(IV_{t-1,t}) + \varepsilon_{t,t+H}$

Model 3: $\ln RRV_{t,t+H} = \alpha_0 + \alpha_{IV}\ln(IV_{t-1,t}) + \alpha_{CD}\ln C_{t-1,t} + \alpha_{CW}\ln C_{t-5,t} + \alpha_{CM}\ln C_{t-20,t} + \alpha_{JD}\ln J_{t-1,t} + \alpha_{JW}\ln J_{t-5,t} + \alpha_{JM}\ln J_{t-20,t} + \varepsilon_{t,t+H}$

Model 4: $\ln RRV_{t,t+H} = \alpha_0 + \alpha_{IV}\ln(IV_{t-1,t}) + \alpha_{JD}\ln J_{t-1,t} + \alpha_{JW}\ln J_{t-5,t} + \alpha_{JM}\ln J_{t-20,t} + \varepsilon_{t,t+H}$

Model 5: $\ln RRV_{t,t+H} = \alpha_0 + \alpha_{IV}\ln(IV_{t-1,t}) + \alpha_{CD}\ln C_{t-1,t} + \alpha_{CW}\ln C_{t-5,t} + \alpha_{CM}\ln C_{t-20,t} + \varepsilon_{t,t+H}$

	Model 1	Model 2	Model 3	Model 4	Model 5
$h=1$					
α_0	0.0797* (0.0287)	-1.3285* (0.0293)	-0.1865* (0.0380)	-1.2579* (0.0395)	-0.2408* (0.0354)
α_{IV}		1.2331* (0.0407)	0.3915* (0.0383)	1.3442* (0.0544)	0.3142* (0.0346)
α_{CD}	0.3647* (0.0320)		0.3145* (0.0317)		0.3158* (0.0297)
α_{CW}	0.3425* (0.0447)		0.2586* (0.0426)		0.2954* (0.0439)
α_{CM}	0.2360* (0.0405)		0.1789* (0.0403)		0.1363* (0.0356)
α_{JD}	-0.2581* (0.0754)		-0.2450* (0.0700)	-0.0856 (0.0845)	
α_{JW}	0.2725 (0.2264)		0.1931 (0.2147)	0.4080 (0.3848)	
α_{JM}	0.0080 (0.3080)		-1.2423* (0.3684)	-2.1564* (0.8089)	
$adjR^2$	0.7538	0.6350	0.7663	0.6407	0.7632
$h=5$					
α_0	0.1421* (0.0529)	-1.2301* (0.0452)	-0.1241 (0.0640)	-1.1716* (0.0608)	-0.1902* (0.0594)
α_{IV}		1.1687* (0.0646)	0.3824* (0.0610)	1.2576* (0.0807)	0.2951* (0.0572)
α_{CD}	0.2418* (0.0490)		0.1646* (0.0492)		0.1738* (0.0485)
α_{CW}	0.3116* (0.0713)		0.2601* (0.0694)		0.3195* (0.0697)
α_{CM}	0.3665* (0.0710)		0.3036* (0.0677)		0.2288* (0.0633)
α_{JD}	-0.0513 (0.1665)		-0.0229 (0.1429)	-0.0185 (0.1939)	
α_{JW}	0.5209 (0.4147)		0.4001 (0.3946)	0.4850 (0.4544)	
α_{JM}	-0.7640 (0.6233)		-1.8934* (0.7015)	-1.9682 (1.1285)	
$adjR^2$	0.7867	0.6745	0.8005	0.6781	0.7971
$h=10$					
α_0	0.1807* (0.0744)	-1.1845* (0.0544)	-0.0600 (0.0921)	-1.1311* (0.0770)	-0.1351 (0.0827)
α_{IV}		1.1253* (0.0814)	0.3367* (0.0780)	1.2066* (0.0991)	0.2568* (0.0773)
α_{CD}	0.2835* (0.0630)		0.2149* (0.0652)		0.2161* (0.0639)
α_{CW}	0.2400* (0.1074)		0.2001 (0.1076)		0.2507* (0.1097)
α_{CM}	0.3961* (0.0909)		0.3315* (0.0902)		0.2684* (0.0875)
α_{JD}	-0.4050 (0.2219)		-0.2903 (0.2181)	-0.1192 (0.3392)	
α_{JW}	0.4893 (0.4997)		0.2886 (0.5183)	0.4843 (0.6744)	
α_{JM}	-0.5655 (0.6968)		-1.5346 (0.8174)	-1.7351 (1.2479)	
$adjR^2$	0.7759	0.6551	0.7871	0.6561	0.7835
$h=20$					
α_0	0.2199 (0.1042)	-1.1228* (0.0698)	-0.0210 (0.1228)	-1.0652* (0.0943)	-0.1153 (0.1165)
α_{IV}		1.0339* (0.1061)	0.3267* (0.1061)	1.1235* (0.1119)	0.2112* (0.1006)
α_{CD}	0.2313* (0.0797)		0.1688* (0.0838)		0.1832* (0.0838)
α_{CW}	0.1622 (0.1741)		0.1175 (0.1795)		0.1439 (0.1904)
α_{CM}	0.4956* (0.1525)		0.4319* (0.1423)		0.3806* (0.1327)
α_{JD}	-0.1930 (0.4462)		-0.1789 (0.4077)	-0.5166 (0.5033)	
α_{JW}	-0.6509 (1.0094)		-0.7153 (1.0525)	-0.2530 (1.3807)	

α_{JM}	-0.2759 (1.2454)		-1.2065 (1.4330)	-0.7573 (1.7804)	
$adjR^2$	0.7166	0.6010	0.7275	0.6011	0.7223

Note: Newey-West standard errors are given in parenthesis.

* denotes statistical significance at the 1% level.

Table 3.2 Encompassing Regression: MIDAS

S&P 500 index 1995/01/01 ~ 2005/03/31

Model 1: $\ln(RRV_{t,t+H}) = \mu_H + \phi_1 \sum_{k=0}^{k_{\max}} b_C(k,1,\theta_{2C}) \ln(C_{t-k-1,t-k}) + \phi_2 \sum_{k=0}^{k_{\max}} b_J(k,1,\theta_{2J}) \ln(J_{t-k-1,t-k}) + \varepsilon_{t+H}$

Model 2: $\ln(RRV_{t,t+H}) = \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) + \varepsilon_{t,t+H}$

Model 3: $\ln(RRV_{t,t+H}) = \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) + \phi_1 \sum_{k=0}^{k_{\max}} b_C(k,1,\theta_{2C}) \ln(C_{t-k-1,t-k}) + \phi_2 \sum_{k=0}^{k_{\max}} b_J(k,1,\theta_{2J}) \ln(J_{t-k-1,t-k}) + \varepsilon_{t+H}$

Model 4: $\ln(RRV_{t,t+H}) = \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) + \phi_2 \sum_{k=0}^{k_{\max}} b_J(k,1,\theta_{2J}) \ln(J_{t-k-1,t-k}) + \varepsilon_{t+H}$

Model 5: $\ln(RRV_{t,t+H}) = \mu_H + \alpha_{IV} \ln(IV_{t-1,t}) + \phi_1 \sum_{k=0}^{k_{\max}} b_C(k,1,\theta_{2C}) \ln(C_{t-k-1,t-k}) + \varepsilon_{t+H}$

	Model 1	Model 2	Model 3	Model 4	Model 5
$h=1$					
μ_H	0.0015 (0.0182)	-1.3198* (0.0303)	-0.4987* (0.0377)	-1.3151* (0.0303)	-0.5232* (0.0359)
α_{IV}		1.2229* (0.0418)	0.5581* (0.0359)	1.2327* (0.0434)	0.5423* (0.0353)
ϕ_1	0.8291* (0.0137)		0.5538* (0.0225)		0.5456* (0.0217)
ϕ_2	-0.1906 (0.1257)		-0.3723* (0.0786)	-0.1401 (0.0830)	
$adjR^2$	0.6913	0.6237	0.7383	0.6239	0.7356
$h=5$					
μ_H	0.1086* (0.0242)	-1.2172* (0.0458)	-0.1765* (0.0598)	-1.2144* (0.0464)	-0.1807* (0.0594)
α_{IV}		1.1540* (0.0654)	0.2898* (0.0564)	1.1587* (0.0655)	0.2838* (0.0565)
ϕ_1	0.9009* (0.0191)		0.7251* (0.0389)		0.7248* (0.0389)
ϕ_2	0.0028 (0.1359)		-0.1101 (0.1284)	-0.0820 (0.1743)	
$adjR^2$	0.7819	0.6632	0.7919	0.6626	0.7922
$h=10$					
μ_H	-0.0704 (0.0716)	-1.1684* (0.0541)	-0.7626* (0.0688)	-1.1619* (0.0560)	-0.7739* (0.0669)
α_{IV}		1.1072* (0.0815)	0.8165* (0.0818)	1.1180* (0.0801)	0.8036* (0.0825)
ϕ_1	0.6600* (0.0481)		0.2888* (0.0439)		0.2871* (0.0440)
ϕ_2	0.3168 (0.2471)		-0.2344 (0.2279)	-0.1727 (0.2589)	
$adjR^2$	0.5228	0.6438	0.6902	0.6430	0.6904
$h=20$					
μ_H	0.0007 (0.0820)	-1.1130* (0.0705)	-0.5583* (0.0901)	-1.0861* (0.0747)	-0.5694* (0.0893)
α_{IV}		1.0225* (0.1066)	0.6369* (0.0898)	1.0612* (0.1086)	0.5989* (0.0887)
ϕ_1	0.7295* (0.0602)		0.3965* (0.0699)		0.4039* (0.0693)
ϕ_2	0.0669 (0.3532)		-0.5527 (0.3030)	-0.7062 (0.4406)	
$adjR^2$	0.5766	0.5912	0.6672	0.5959	0.6650

Note: Newey-West standard errors are given in parenthesis.

* denotes statistical significance at the 1% level.

Table 3.3 Encompassing Regression: Implied Volatility and Out-of-Sample Volatility

S&P 500 Index

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_1 \ln IV_{t-1,t} + \varepsilon_{t,t+H}$$

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_2 \ln \widehat{RRV}_{t,t+H} + \varepsilon_{t,t+H}$$

$$\ln RRV_{t,t+H} = \alpha_0 + \alpha_1 \ln IV_{t-1,t} + \alpha_2 \ln \widehat{RRV}_{t,t+H} + \varepsilon_{t,t+H}$$

where $\ln \widehat{RRV}_{t+1,t+H}$ is the out-of-sample estimation of the HAR-RRV-CJ (or MIDAS-RRV-CJ) volatility forecast.

	$h=1$			$h=5$			$h=10$			$h=20$		
HAR	α_0	α_1	α_2	α_0	α_1	α_2	α_0	α_1	α_2	α_0	α_1	α_2
Coefficient	0.3060*	1.1587*		0.3205*	1.1371*		0.3405*	1.1063*		0.3724*	1.0586*	
Std. Error	0.0477	0.1085		0.0432	0.0996		0.0397	0.0927		0.0351	0.0816	
$adjR^2$	0.4476			0.5958			0.6282			0.6488		
Coefficient	1.3219*		1.0222*	1.2406*		0.9928*	1.2101*		0.9740*	1.1642*		0.8992*
Std. Error	0.0559		0.0898	0.0537		0.0888	0.0538		0.0869	0.0518		0.0775
$adjR^2$	0.4848			0.5626			0.5470			0.4599		
Coefficient	0.9969*	0.3875*	0.7267*	0.6745*	0.7203*	0.4326*	0.5969*	0.8016*	0.3465*	0.5296*	0.8723*	0.2577*
Std. Error	0.1276	0.1363	0.1409	0.0882	0.1141	0.1057	0.0769	0.1115	0.0932	0.0648	0.1020	0.0877
$adjR^2$	0.4940			0.6223			0.6495			0.6663		

MIDAS	α_0	α_1	α_2	α_0	α_1	α_2	α_0	α_1	α_2	α_0	α_1	α_2
Coefficient	0.3060*	1.1587*		0.3205*	1.1371*		0.3405*	1.1063*		0.3724*	1.0586*	
Std. Error	0.0477	0.1085		0.0432	0.0996		0.0397	0.0927		0.0351	0.0816	
$adjR^2$	0.4476			0.5958			0.6282			0.6488		
Coefficient	1.3289*		1.0218*	1.2538*		1.0059*	1.2233*		0.9794*	1.1650*		0.8974*
Std. Error	0.0560		0.0896	0.0537		0.0889	0.0546		0.0875	0.0526		0.0784
$adjR^2$	0.4864			0.5736			0.5479			0.4593		
Coefficient	1.0116*	0.3760*	0.7354*	0.7095*	0.6844*	0.4693*	0.6014*	0.8015*	0.3473*	0.5275*	0.8749*	0.2530*
Std. Error	0.1219	0.1273	0.1356	0.0893	0.1119	0.1075	0.0743	0.1078	0.0896	0.0611	0.0989	0.0788
$adjR^2$	0.4950			0.6259			0.6492			0.6655		

Note: The standard errors are computed following a robust procedure, Neway-West standard errors, taking into account of the heteroscedastic and autocorrelated error structure.

* denotes statistical significance at the 1% level.

Chapter 4 Detecting Mutual Fund Timing Ability Using the Threshold Model

1. INTRODUCTION

Investment performance and the market timing of mutual funds continue to receive considerable attention by both academics and market practitioners alike, with a variety of evaluation techniques having been proposed and implemented over the years. Treynor (1965), Sharpe (1966) and Jensen (1968), for example, measured the excess returns for systematic risk,[10] whilst more recently, Bollen and Busse (2001) and Chance and Helmer (2001) have stressed the importance of daily tests for performance measurement.

This chapter proposes a new method of testing mutual fund performance and market timing through the application of threshold regression techniques. The idea is that fund managers may adopt different trading strategies when they perceive different market conditions. As fund managers may not uniformly use the sign of the market return to capture the direction of market movement, it is natural to conjecture that a fund manager’s trading behavior changes when the market return is above or below a certain threshold level, which varies across managers of different funds.

Threshold models have been widely applied in the econometric analysis; the threshold autoregressive model (TAR), for example, remains popular in the examination of nonlinear time-series data. Hansen (2000) presented a statistical theory for threshold estimation, in a regression context, proposing least squares estimation of the regression parameters and concluding with the asymptotic distribution theory for the regression estimates.

---

[10] Treynor and Mazuy (1966), Henriksson and Merton (1981) and Chang and Lewellen (1984) noted that investment managers have superior information and forecasting skills.

This chapter aims to contribute to this field through the introduction of the threshold model into the testing of mutual fund market-timing effects. The traditional Henriksson and Merton (1981) model is shown to represent only a special case within our model, and we demonstrate the potential bias of using the traditional model, arguing that it tends to underestimate the market-timing effect. Indeed, we find that the use of the traditional market timing test may provide misleading results in some circumstances; thus, our proposed threshold model provides more accurate inferences on the market-timing effects of mutual funds.

2. THRESHOLD MODEL AND MARKET TIMING

2.1 Models for Mutual Fund Performance and Market-timing Effects

We begin by using the threshold regression model developed by Hansen (1996) to propose a model for testing mutual fund performance and market-timing effects. The threshold regression model takes the form:

$$\begin{aligned} R_i - R_f &= \alpha_i^1 + \beta_i^1 (R_m - R_f) + e_i \qquad R_m - R_f \le q_i \\ R_i - R_f &= \alpha_i^2 + \beta_i^2 (R_m - R_f) + e_i \qquad R_m - R_f > q_i \end{aligned} , \tag{1}$$

where R_i is the rate of return on the i^{th} mutual fund; R_m is the rate of return on the market portfolio; R_f is the riskless rate; q_i is the threshold parameter; α_i^1 (α_i^2) is the abnormal return of the i^{th} mutual fund when the excess return rate on the market portfolio is smaller (larger) than the threshold variable; and β_i^1 (β_i^2) is the systematic risk of the i^{th} mutual fund when the excess return on the market portfolio is smaller (larger) than the threshold variable. If there is any significant increase in systematic risk, ($\beta_i^2 > \beta_i^1$), fund managers will have market-timing ability.

The Henriksson and Merton (1981) model can be written as follows:

$$R_i - R_f = \alpha_i + \beta_{i1}(R_m - R_f) - \beta_{i2} \cdot d_m(0) \cdot (R_m - R_f) + e_i \,, \qquad (2)$$

where d_m (0) = $I\{R_m - R_f < 0\}$ is the dummy variable with $I\{\cdot\}$ as the indicator function; α_i is the abnormal return of the i^{th} mutual fund; β_{i1} and (β_{i2}) are beta regression coefficients; and the fund manager's market-timing ability is expressed as β_{i2}. It is clear that the traditional Henriksson and Merton (1981) model is a special case of the threshold regression model in equation (1) where q to the value of *0*,

The above threshold regression model (1) can be rewritten as follows:

$$r_i = \theta_1' \cdot r_m^* + \lambda' \cdot r_m^*(q) + e_i \,, \quad ,$$

where $r_m^* = [1^* \; r_m]$, $r_m^*(q) = [1^* \; (r_m d_m(q))]$, r_m is the $n \times 1$ vector of excess return rate on the market portfolio; and 1^* is a column vector of ones. r_m^* and $r_m^*(q)$ are both $n \times 2$ matrices; n represents the number of observations on the i^{th} mutual fund; d_m (q) = $I\{r_m > q\}$ is the dummy variable with $I\{\cdot\}$ as the indicator function; r_i is the $n \times 1$ vector of excess return rate on the i^{th} mutual fund; θ_1 is the vector of coefficients of the model when the excess return on the market portfolio is smaller than the threshold variable; θ_2 is the vector of coefficients of the model when the excess return on the market portfolio is greater than the threshold variable; $\lambda = \theta_2 - \theta_1$ denotes the 'threshold effect'; and e_i is the $n \times 1$ vector of error. If the results of the test on λ are significantly different from zero, this will indicate that the manager possesses market-timing ability.

The regression parameters are estimated by the least squares method, with the sum of the squared errors function being shown as:

$$S_n(\theta_1, \lambda, q) = (r_i - \theta_1' \cdot r_m^* - \lambda' \cdot r_m^*(q))' \cdot (r_i - \theta_1' \cdot r_m^* - \lambda' \cdot r_m^*(q)) \,.$$

Conditional on q yielding the OLS estimators $\hat{\theta}_1(q)$ and $\hat{\lambda}(q)$, by regression of r_i on $(r_m^*, r_m^*(q))$, the concentrated sum of the squared errors function is:

$$S_n(q) = S_n(\hat{\theta}_1(q), \hat{\lambda}(q), q) = r_i' \cdot r_i - r_i' \cdot r_m^{q^*} \left(r_m^{q^{*\prime}} r_m^{q^*} \right)^{-1} r_m^{q^{*\prime}} r_i \,,$$

where $r_m^{q^*}$ is the excess return on the market portfolio under the threshold condition. For the minimization of the sum of the squared errors, q is assumed to be restricted to a bounded set (empirically, it usually uses the 15 per cent quartile of the sample to the 85 per cent quartile of the sample); the least-squares estimate $\hat{q}$ of the threshold parameter q is the value which minimizes $S_n(q)$. The consistency threshold estimate $\hat{q}$ is defined as:

$$\hat{q} = \arg\min S_n(q) \,.$$

Note that the LS estimator is also the MLE when e_i is *i.i.d.* $N(0,\sigma^2)$. Hansen (2000) provided the asymptotic distribution of the consistent threshold estimate $\hat{q}$, and suggested the use of the likelihood ratio statistic to test the hypothesis $H_0 : q = q_0$ under the condition of e_i being *i.i.d.* $N(0,\sigma^2)$. The likelihood ratio statistic under homoskedasticity is different from that under heteroskedasticity. The test proposed by White (1980) can be employed to examine the homoskedastic disturbances.

Under the assumption of homoskedasticity, the likelihood ratio statistic for $q = q_0$ is defined as:

$$LR(q_0) = n \cdot \frac{S_n(q_0) - S_n(\hat{q})}{S_n(\hat{q})} \tag{3}$$

The likelihood ratio test of H_0 is rejected for large values of $LR_n(q_0)$. If heteroskedasticity exists, the likelihood ratio statistic under $q = q_0$ is defined as:

$$LR^*(q_0) = \frac{LR(q_0)}{\hat{\eta}^2} = n \cdot \frac{S_n(q_0) - S_n(\hat{q})}{S_n(\hat{q}) \cdot \hat{\eta}^2} \tag{4}$$

where $\hat{\eta}^2$ is an estimator of $\eta^2 = \dfrac{c' E(r_m r_m' e_i^2 \mid q = q_0) c}{\sigma^2 \cdot c' E(r_m r_m' \mid q = q_0) c}$.

As demonstrated in both Henriksson and Merton (1981) and Chang and Lewellen (1984), we can use the excess return on the market portfolio to determine whether or not a bull market exists. Our aim is to test whether the market managers are able to adjust their investment principles according to the market index; that is, to test the hypothesis $H_0 : q = 0$.

2.2 Testing for Threshold Effects

Using the changes in the regression coefficients of the threshold estimate allows us to evaluate the mutual fund manager's stock-selection and market-timing abilities. We construct the hypothesis $H_0 : \lambda = 0$ to test for the threshold effect.

If the fund manager does not exhibit market timing behavior, the conditional sum of the squared errors $S_n(q_0)$ of (3) and (4) will be equal to the sum of the squared errors $(e_i' e_i)$ in the traditional one-regime CAPM (i.e., $r_i = \theta' \cdot r_m + e_i$).

In the presence of homoskedasticity, the likelihood ratio statistic is defined as:

$$LR = n \cdot \frac{e_i' e_i - S_n(\hat{q})}{S_n(\hat{q})} \tag{5}$$

Under H_0 the threshold q remains unidentified; therefore, the classical tests have non-standard distribution. Hansen (1996) suggested the adoption of a bootstrap to simulate the asymptotic distribution of the likelihood ratio test, showing that a bootstrap procedure attains the first-order asymptotic distribution; thus, the *p-values* constructed for the bootstrap are asymptotically valid. We use bootstrap replication to generate a bootstrap sample of size 1,000 so that the residual features are the same as those of an individual mutual fund. The small sample distribution and the *p-value* of the likelihood ratio test estimator are then obtained.

2.3 Test for the Source of the Threshold Effect

In order to test whether the threshold effect stems from manager's stock-selection ability or market-timing ability, we use the threshold estimate as the dummy variable, thereby dividing the mutual fund samples into two sample sets. We then construct a test which can determine whether the threshold effect comes from manager's stock-selection ability or market-timing ability. The model constructed is similar to the Fabozzi and Francis (1979) model, as follows:

$$r_i = \alpha_i + \lambda_1 d_m^*(\hat{q}) + \beta_i r_m + \lambda_2 d_m^*(\hat{q}) r_m + e_i \, , \tag{6}$$

where $d_m^*(\hat{q}) = I\{r_m > \hat{q}\}$ is the dummy variable with $I\{\cdot\}$ as the indicator function; $\hat{q}$ is the threshold estimator; α_i is the excess return rate on the i^{th} mutual fund without threshold effect; β_i is the systematic risk of the i^{th} mutual fund without threshold effect, λ_1 is the abnormal return disparity under $(r_m > \hat{q})$; λ_2 is the systematic risk disparity of the i^{th} mutual fund under $(r_m > \hat{q})$; and e_i is a regression error. The aim of constructing the hypothesis test is to determine whether the threshold effect stems from manager's stock-selection ability or market-timing ability; this is undertaken by testing to see whether the corresponding differential coefficient is statistically different from zero. A positive value of λ_1 represents that the fund manager presents sufficient stock-selection ability in anticipation of a bull market, while a positive λ_2 indicates that the fund manager has market-timing ability.

3. Data and Empirical Results

Bollen and Busse (2001) demonstrated that daily tests are more forceful than monthly tests, with mutual funds more often displaying significant timing ability from such daily tests; hence, our analysis of the market-timing effect is based upon the daily returns of 30 randomly-selected mutual funds. The sample is taken from the aggressive growth mutual fund of the Center for Research in Security Prices (CRSP) mutual fund database, with the sample period running from 1 January 2000 to 31 January 2003. We employ the net asset value and dividends to form a daily return series for each fund. We use the CRSP value-weighted index, including NYSE, AMEX and NASDAQ stocks, as an overall market benchmark. Three-month Treasury Bills rates, drawn from the Federal Reserve Board, are used as the risk-free rates. Our results show that half of the mutual funds beat the market.[11]

The results in Table 1 demonstrates that 17 of the funds have threshold effects and that the abnormal returns of 16 of the 17 funds are both significant and positive ($\alpha_i^2 > \alpha_i^1$), which indicates that the managers have stock-selection abilities. Only four of the 17 fund managers have market-timing ability because there is a significant increase in their systematic risks ($\beta_i^2 > \beta_i^1$); four of the 17 funds possess both stock-selection and market-timing abilities. Furthermore, superior fund managers will increase the systematic risk of a portfolio in anticipation of a bull market, so as to raise the risk premium and reduce the systematic risk of the portfolio, thus reducing losses when a bear market is forecasted.

The traditional Henriksson and Merton Model is the threshold regression model, with the restriction that q = 0. The results of Table 2 reveal that 11 of the 17 funds show a rejection of the hull hypothesis that q = 0; therefore, the traditional Henriksson

[11] The results are omitted to save space. However, they are available upon request.

and Merton (1981) model is rejected. Hence, we demonstrate that there is potential bias in the use of the traditional model.[12]

The model employed in this study essentially explores the assumption of the existence of a threshold effect. This assumption is important because it does affect our evaluation of the investment performance of mutual fund managers. For example, as demonstrated in Table 3, under the traditional model of Henriksson and Merton (1981), four of the funds indicate that the fund managers do not possess any market-timing or stock-selection ability; however, the fund managers not only achieved more abnormal returns, but also increased the systematic risk so as to earn higher market risk premiums once the market excess return was larger than the threshold estimate.

4. CONCLUSIONS

This study has proposed the use of the threshold regression model to evaluate the market-timing abilities of mutual fund managers. The empirical results for a set of randomly-selected US mutual funds indicate that the threshold values of market timing are different from 0 for more than 50% of the mutual funds. Our results indicate potential bias in the use of the traditional Henriksson and Merton (1981) model with regard to its evaluation of the ability of fund managers to select stocks, and we find that the traditional model also tends to underestimate the market-timing effect under the use of the capital asset pricing model with threshold effects.

[12] The regression results of the threshold effect from equation (6) are omitted for saving space. Sixteen of the mutual funds exhibited a positive and significant value for $\lambda_1(\hat{q})$, indicating that the fund manager has stock-selection ability based upon the threshold effect. Four of the mutual funds also exhibited a positive and significant value for $\lambda_2(\hat{q})$, indicating that the fund manager has market-timing ability based upon the threshold effect.

Table 4.1 Estimation results of mutual fund market-timing effect using the threshold model

Fund name	Threshold Variables	α_i^1 [b]	β_i^1 [c]	α_i^2 [b]	β_i^2 [c]	p-value [d]
Bear Stearns Small Cap Value Portfolio/C	-0.0087	0.005	1.174	0.000	0.825	0.065 *
		(3.013)	(12.363)	(0.110)	(20.553)	
Dreyfus Founders Funds: Discovery Fund/T	0.0057	-0.001	1.003	0.003	0.856	0.080 *
		(-2.322)	(25.305)	(2.588)	(13.733)	
Oppenheimer Discovery Fund/A	-0.0025	-0.002	0.874	0.001	0.844	0.005**
		(-2.603)	(16.867)	(2.691)	(22.401)	
INVESCO Dynamics Fund/Instl	-0.0061	-0.007	1.091	0.000	1.391	0.004**
		(-4.423)	(11.966)	(0.396)	(29.474)	
NI Numeric Investors Growth Fund	0.0072	0.000	1.070	0.004	0.890	0.097 *
		(-0.699)	(30.497)	(3.054)	(14.336)	
Quaker Aggressive Growth Fund	0.0055	-0.001	0.227	0.004	-0.021	0.000**
		(-2.337)	(8.951)	(5.569)	(-0.536)	
Smith Barney Small Cap Core Fund/B	-0.0086	-0.008	0.839	-0.001	1.240	0.006**
		(-4.845)	(8.050)	(-3.370)	(33.872)	
Royce Fund: Opportunity/ Instl Serv	-0.0025	0.000	0.836	0.002	0.652	0.012**
		(0.458)	(15.681)	(3.798)	(16.728)	
TD Waterhouse Extended Market Index Fund	0.0075	0.000	1.003	0.003	0.821	0.058 *
		(-0.056)	(39.384)	(3.027)	(17.373)	
Aetna Index Plus Small Cap Fund/I	0.0074	0.000	0.864	0.004	0.703	0.057 *
		(0.05)	(30.017)	(3.588)	(13.249)	
AIM Small Cap Opportunities	-0.0029	-0.003	0.603	0.000	0.752	0.038**
		(-3.590)	(10.035)	(0.782)	(17.947)	
Analysts Aggressive Stock Fund	-0.0061	-0.004	0.997	0.001	1.162	0.052 *
		(-3.099)	(12.970)	(1.412)	(29.191)	
J Hancock Small Cap Growth Fund/I	-0.0025	-0.002	1.009	0.001	1.006	0.076 *
		(-2.563)	(17.960)	(1.772)	(24.495)	
Undiscovered Managers Small Cap Growth/Instl	0.0079	-0.001	1.377	0.009	0.955	0.003**
		(-0.712)	(22.378)	(3.933)	(8.298)	
Merrill Lynch Master Small Cap Vl Tr Fund/B	-0.0025	-0.001	0.868	0.002	0.828	0.046**
		(-1.237)	(18.434)	(3.442)	(24.024)	
Lord Abbett Developing Growth Fund/A	-0.0037	-0.002	0.955	0.001	0.910	0.055 *
		(-1.836)	(16.560)	(2.151)	(24.286)	
State Street Research: Emerging Growth Fund/B1	0.0072	0.000	1.061	0.006	0.782	0.003**
		(-0.502)	(26.615)	(4.300)	(11.090)	

Notes:

[a] This table presents the estimation results for the model: as $R_i - R_f = \alpha_i^1 + \beta_i^1(R_m - R_f) + e_i \quad R_m - R_f \le q$; $R_i - R_f = \alpha_i^2 + \beta_i^2(R_m - R_f) + e_i \quad R_m - R_f > q$; where q is the threshold parameter; R_i is the return rate of the i^{th} mutual fund; R_m is the return rate on the market portfolio; and the mutual funds have threshold effect.

[b] α_i^1 (α_i^2) is an abnormal return of the i^{th} mutual fund when the excess return rate on the market portfolio is smaller (larger) than the threshold estimate.

[c] β_i^1 (β_i^2) is the systematic risk of the i^{th} mutual fund when the excess return rate on the market portfolio is smaller (larger) than the threshold estimate.

[d] The null hypothesis of the test is $\lambda = 0$.

[e] Figures in parentheses are t-ratios.

[f] * indicates significance at the 10 per cent level; ** indicates significance at the 5 per cent level.

Table 4.2 Results of tests for the threshold variable of market timing being equal to zero

Fund name	LR [a]	p-value [b]
Bear Stearns Small Cap Value Portfolio/C	8.744	0.438
Dreyfus Founders Funds: Discovery Fund/T	8.413	0.073*
Oppenheimer Discovery Fund/A	9.100	0.042
INVESCO Dynamics Fund/Instl	8.570	0.009**
NI Numeric Investors Growth Fund	8.717	0.086*
Quaker Aggressive Growth Fund	8.774	0.007**
Smith Barney Small Cap Core Fund/B	8.790	0.023**
Royce Fund:Opportunity/Instl Serv	8.700	0.530
TD Waterhouse Extended Market Index Fund	8.872	0.288
Aetna Index Plus Small Cap Fund/I	8.552	0.026**
AIM Small Cap Opportunities	8.001	0.035**
Analysts Aggressive Stock Fund	8.560	0.022**
J Hancock Small Cap Growth Fund/I	8.602	0.332
Undiscovered Managers Small Cap Growth/Instl	8.547	0.003**
Merrill Lynch Master Sm Cp Vl Tr Fund/B	8.851	0.167
Lord Abbett Developing Growth Fund/A	8.658	0.089*
State Street Research:Emerging Growth Fund/B1	8.609	0.010**

Notes:

[a] The null hypothesis of LR is $q = 0$.

[b] * indicates significance at the 10 per cent level; ** indicates significance at the 5 per cent level.

Table 4.3 Mutual fund market timing and performance test, threshold model vs. Henriksson and Merton model

Fund name	Threshold Regression Model [a]				Henriksson and Merton Model [b]		
	α	β	$\lambda_1(\hat{q})$	$\lambda_2(\hat{q})$	α	β_1	β_2
INVESCO Dynamics Fund/Instl	-0.007	1.091	0.007	0.300	-0.001	1.438	0.027
	(-4.42)	(11.96)	(4.28)**	(2.92)**	(-1.19)	(28.05)	(0.30)
Smith Barney Small Cap Core Fund/B	-0.008	0.839	0.007	0.401	-0.001	1.259	0.041
	(-4.845)	(8.050)	(4.075)**	(3.631)**	(-3.424)	(27.046)	(0.502)
AIM Small Cap Opportunities	-0.003	0.603	0.004	0.150	-0.001	0.806	0.056
	(-3.59)	(10.04)	(3.48)**	(2.04)**	(-1.25)	(19.75)	(0.80)
Analysts Aggressive Stock Fund	-0.004	0.997	0.005	0.165	-0.001	1.202	0.001
	(-3.10)	(12.97)	(3.40)**	(1.90)*	(-0.08)	(27.98)	(0.01)

Notes:

[a] This table presents threshold regression results for the model: $r_i = \alpha_i + \lambda_1 d_m^*(\hat{q}) + \beta_i r_m + \lambda_2 d_m^*(\hat{q}) r_m + e_i$,

where $d_m^*(\hat{q}) = I\{r_m > \hat{q}\}$ is the dummy variable with $I\{\cdot\}$ as the indicator function; $\hat{q}$ is the threshold estimator.

[b] α_i is alpha regression intercept for the i^{th} mutual fund, β_{i1} and β_{i2} are beta regression coefficients.

[c] Figures in parentheses are t-ratios.

[d] * indicates significance at the 10% level; **indicates significance at the 5% level.

Chapter 5 Conclusions

This dissertation will employ econometrics analyses to focus on two important issues in the financial market, including volatility forecasting and mutual fund performance evaluation.

The first issue in this dissertation is to discuss the predicting volatility. The empirical results show that the realized range-based variance is more efficient and the results are the same as the previous studies. The regressors consisting of the continuous sample path and jump variability measures in the HAR and MIDAS regressions predict the future realized range volatilities, and such dominates almost in all MSE terms. In addition, in the in-sample forecasting, the relative decreasing ratio of MSE of HAR-RRV (MIDAS-RRV) regressions are almost larger than that of HAR-RV (MIDAS-RV); and in the out-of sample forecasting, the MSE of HAR-RRV and MIDAS-RRV regressions are small. Moreover, the realized range-based regressions are significant for short-run volatility forecasting, but the realized return-based regressions are almost invariant to jumps. Hence, our empirical results show that by using of the HAR and MIDAS regressions to predict latent volatility under different variations, the realized range-based variance is a good volatility proxy.

Furthermore, the implied volatility and almost all continuous components are statistically significant, while the jump components are almost not significant, and there are little increases with the adjusted R^2 s. Hence, when the continuous components are included in the encompassing regression, they will share the explanatory power with the implied volatility. That is to say, the implied volatility subsumes most of the relevant volatility information. Beside, the jump components do not contribute to future valuable information. In addition, the implied volatility and the out-of-sample volatility have information contents but the implied volatility has more powerful explanation abilities than the out-of-sample volatility for the future realized range-based volatility.

The second issue examined in this dissertation is to detect mutual fund market timing abilities using the threshold regression model. Our results indicate potential bias in the use of the traditional Henriksson and Merton (1981) model with regard to its evaluation of the ability of fund managers to select stocks, and we find that the traditional model also tends to underestimate the market-timing effect under the use of the capital asset pricing model with threshold effects.

To conclude, the essays of this dissertation provide some insights into the issues of volatility forecasting and mutual fund performance evaluation. The results of this study provide us with empirical evidences to comprehend the occasion of some distinctive phenomena in financial markets.

References

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

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

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003), "Modeling and Forecasting Realized Volatility", Econometrica 71, 579-625.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2004), "Some Like it Smooth, and Some Like it Rough: Untangling Continuous and Jump Components in Measuring, Modeling and Forecasting Asset Return Volatility", Working Paper, Duke University.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007), "Roughing It Up: Including Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility", Review of Economics and Statistics 98(4).

Bandi, F. M. and J. R. Russell (2005), "Microstructure Noise, Realized Variance, and Optimal Sampling", Graduate School of Business, University of Chicago, Working Paper.

Bandi, F. M. and J. R. Russell (2006), "Separating Microstructure Noise from Volatility", Journal of Financial Economics 79, 655-692.

Barndorff-Nielsen, O. E., and N. Shephard (2002), "Estimating Quadratic Variation Using Realized Variance", Journal of Applied Econometrics 17, 457-477.

Barndorff-Nielsen, O. E., and N. Shephard (2004a), "Power and Bipower Variation with Stochastic Volatility and Jumps", Journal of Financial Econometrics 2, 1-37.

Barndorff-Nielsen, O. E., and N. Shephard (2004b), "How Accurate is the Asymptotic

Approximation to the Distribution of Realised Volatility", in Identification and Inference for Econometric Models. A Festschrift in Honors of T. J. Rothenberg (D. Andrews, J. Powell, P.A. Ruud, and J. Stock, eds.). Cambridge, UK: Cambridge University Press.

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

Bollen, N.P.B. and J.A. Busse (2001), "On the Timing Ability of Mutual Fund Managers", Journal of Finance, 41, 715--30.

Chance, D.M. and M.L. Hemler (2001), "The Performance of Professional Market Timers: Daily Evidence from Executed Strategies", Journal of Financial Economics, 62, 377--411.

Chang, E.C. and W.G. Lewellen (1984), "Market Timing and Mutual Fund Investment Performance", Journal of Business, 57, 57--72.

Christensen, K. and M. Podolskij (2006a), "Realized Range-Based Estimation of Integrated Variance", Journal of Econometrics (Forthcoming).

Christensen, K. and M. Podolskij (2006b), "Range-Based Estimation of Quadratic Variance", Aarhus School of Business.

Corsi, F. (2004), "A Simple Long Memory Model of Realized Volatility", Unpublished manuscript, University of Southern Switzerland.

Fabozzi, F.J. and J.C. Francis (1979), "Mutual Fund Systemic Risk for Bull and Bear Markets: An Empirical Examination", Journal of Finance, 34, 1243--50.

Forsberg, L. and E. Ghysels (2007), "Why Do Absolute Returns Predict Volatility So well?", Journal of Financial Econometrics 5, 31-67.

Ghysels, E., P. Santa-Clara, and R. Valkanoy (2002), "The MIDAS Touch: Mixed Data Sampling Regression Models", Chapel Hill, NC: University of North Carolina.

Ghysels, E., P. Santa-Clara, and R. Valkanoy (2005), "There is a Risk-Return Trade-Off After All", Journal of Financial Economics 76, 509-548.

Ghysels, E., P. Santa-Clara, and R. Valkanoy (2006), "Predicting Volatility: Getting the most out of Return Data Sampled at Different Frequencies", Journal of Econometrics 131, 59-95.

Giot, P. and S. Laurent (2007), "The Information Content of Implied Volatility in Light of the Jump/Continuous Decomposition of Realized Volatility", The Journal of Futures Markets 27, 337-359.

Hansen, B.E. (1996), "Inference when a Nuisance Parameter is Not Identified under the Null Hypothesis", Econometrica, 64, 413--30.

Hansen, B.E. (2000), "Sample Splitting and Threshold Estimation", Econometrica, 68, 575--603.

Hansen, P. R. and A. Lunde (2006), "Realized Variance and Market Microstructure Noise", Journal of Business and Economic Statistics 24, 127-218.

Henriksson, R.D. and R.C. Merton (1981), "On Market Timing and Investment Performance, II. Statistical Procedures for Evaluating Forecasting Skills", Journal of Business, 54, 513--33.

Huang, X. and G. Tauchen. (2005), "The Relative Contribution of Jumps to Total Price Variance", Journal of Financial Econometrics 3, 456-499.

Jensen, M.C. (1968), "The Performance of Mutual Funds in the Period 1945-1964", Journal of Finance, 23, 389--416.

Martens, M. and Dick van Dijk. (2007), "Measuring Volatility with the Realized

Range”, Journal of Econometrics 138, 181-207.

Merton, R.C. (1981), “On Market Timing and Investment Performance, I. An Equilibrium Theory of Value for Market Forecasts”, Journal of Business, 54, 363--406.

Patton A. (2006), “Volatility Forecast Comparison Using Imperfect Volatility Proxies”, Unpublished manuscript, London School of Economics.

Parkinson, M. (1980), “The Extreme Value Method for Estimating the Variance of the Rate of Return”, Journal of Business 53, 61-65.

Sharpe, W.F. (1966), “Mutual Fund Performance”, Journal of Business, 39, 119--38.

Treynor, J.L. (1965), “How to Rate Management of Investment Funds”, Harvard Business Review, 43, 63--75.

Treynor, J.L. and K.K. Mazuy (1966), “Can Mutual Funds Outguess the Market?”, Harvard Business Review, 44, 131--6.

White, H. (1980), “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”, Econometrica, 48, 817--38.