國立臺灣大學社會科學院經濟學系

碩士論文

Department of Economics

College of Social Sciences

National Taiwan University

Master Thesis

異質性自我迴歸模型於實現波動率之預測

—台灣加權股票指數的實證研究

Forecasting Realized volatility using the Heterogeneous Autoregressive Model：Evidence from Taiwan Stock Exchange Index

林朕陞

Jhen-Sheng Lin

指導教授：林金龍 博士

Advisor: Jin-Lung Lin, Ph.D.

中華民國 99 年 6 月

June, 2010

誌謝

這篇論文的誕生是我投注許多心力得以完成，這之中遇到許多的困難卻都能順利解決，這都要感謝幫助過我的老師及同學，所以要趁此機會表達由衷的謝意。首先，感謝指導教授林金龍老師對於論文內容提供建議與想法，也因我是以英文撰寫，在許多字句文法的問題，老師都很耐心的協助修改；此外，對老師感到抱歉，因為早在碩二上的期中就找了老師當指導教授，卻延到碩三才把碩士論文完成，所以再次謝謝老師過去 2 年的等待與耐心指導。

另外，也要感謝 Elijah Dickhaus 對於英文部分的潤稿，Elijah 在英文部分幫助我讓文法和句子通順度臻至完美，而他學的是語文所以在校稿上也遇到一些非屬他專業的領域，所以 Elijah 花了很多時間仔細修改字句。接著要感謝兩位口試委員—吳中書院長與林建甫教授提出的評論與建議，力求論文正確無誤，使我在學術涵養上增進不少。

最後，特別謝謝爸媽一路的辛苦栽培以及友人的支持，自碩二開始他們就很擔心我不能畢業的問題，如今論文順利高分通過，他們總算是放下心中的大石，希望未來的人生道路能永遠有家人的支持與關心。

林朕陞

謹誌於台北五股家中

民國 99 年 7 月

摘要

本文採用異質性自我迴歸(HAR-RV)模型，以精簡原則來配適具有緩長記憶性的實現波動率序列，來預測波動率並且能提供統計上的結果去解釋台股指數波動率的特性。我們希冀能配適一個能夠充分預測價格變動的波動模型，進而幫助投資人在風險管理或交易策略上提供有效的決策依據。

實證結果發現，加入了槓桿效果與交易量作為解釋變數的 LHAR-RV-cum-Vol 模型提供了最佳的預測模型。其中，槓桿效果顯示了具有異質性結構，而且槓桿效果在短期是由日跳躍所引起，但長期下卻不由日跳躍所引起。此外，研究結果顯示只有日交易量對於未來波動率具有顯著的解釋能力，特別是以成交筆數當做訊息流動之替代變數時，才能提供最好的預測能力。

再者，我們發現使用 Corsi, Pirino and Reno(2009)的 CTBPV 方法來分離連續與跳躍並加入 HAR-RV 模型其所提供的預測能力優於 Barndorff-Nielsen and Shephard(2004)的 BPV 方法，但兩者的差異不具有統計顯著性。最後，本研究將市場區分成多頭市場與空頭市場時，研究結果發現門檻連續與跳躍(TCJ)做為解釋變數不論在多頭或是空頭市場皆提供最好的預測能力，但是已實現冪次變異(RPV)做為解釋變數時只有在空頭市場提供最佳的預測能力。當市場處於空頭市場時，其所隱含的交易資訊會上升，進而提升了實現波度率的預測績效。

關鍵字： 波動度預測、跳躍、槓桿效果、修正後門檻估計量、異質結構、高頻率資料

Abstract

This paper employs a 'Heterogeneous Autoregressive' (HAR) model which is suitable to parsimoniously model long memory in realized volatility time series. The purpose is to use this model to predict the future volatility and provide some statistical results to explain volatility behavior in Taiwan stock index market. We hope to provide an accurate predictive model on the volatility and then help investors with regards to risk management or trading strategies.

The empirical results verify that the "best" model for volatility prediction is the LHAR-RV-cum-Vol model which includes the leverage effect and trading volume as regressors. Particularly, the leverage effect unveils a heterogeneous structure and this effect is induced by jumps for short-run prediction horizons but not for long-run prediction horizon. Besides, results reveal only daily trading volume has significant effect on future volatility, especially the number of transactions as a proxy for information flows provides the best predictive ability on the volatility.

The empirical results also reveal that the HAR model adds continuous components (C) and jump components (J) extracted by Corrected Threshold Bi-power Variation (Corsi et al. 2009) to predict volatility better than Bi-power Variation (Barndorff-Nielsen and Shephard, 2004). However, we do not get significant gain derived by dividing the continuous and jump components. Lastly, this study separates the market into up-market days and down-market days. We find that the threshold continuous and jump (TCJ) as a regressor is the top forecaster in both markets, while realized power variation (RPV) only performs best on down-market days. When the market is down the amount of market information increases, the predictive ability of future volatility also increases.

Keywords: volatility forecasting; jumps; leverage effects; corrected threshold estimator; heterogeneous structure; high-frequency data.

Content

Figure Index

Table Index

1 Introduction

Volatility forecasting plays a central role in a number of financial issues, such as asset pricing, asset allocation, and risk management. A major difficulty in this kind of forecasting is that volatility can only be observed after the fact. The traditional approach to estimating volatility is to use a parametric framework, such as the ARCH and GARCH and stochastic volatility models. In recent years, the Taiwan financial market has grown to a mature level, coinciding with an increased availability of high-frequency data on asset returns. The availability of this data suggests a non-parametric approach to modeling volatility dynamics using improved measures of ex post, or *integrated* volatility, constructed from high frequency data. This method is known as 'realized volatility', or RV.

The RV method has been advocated by such reputable financial analysts as Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold and Labys (henceforth ABDL) (2001), Barndorff-Nielsen and Shephard (henceforth BN-S) (2002a, b), and Meddahi (2002), among others. The main idea is to sum up the corresponding intra-daily squared returns; this is a consistent estimator of integrated volatility, as well as a jump component for a broad class of continuous time models. Specifically, models based on realized volatility have been found to produce forecasts of volatility that are superior to traditional measures of volatility, such as squared returns. As an example, empirical results produced by ABDL (2003) powerfully indicated that the simple linear models of realized volatility outperformed the popular GARCH and related stochastic volatility models in out-of-sample forecasting.

There are, however, two alternative measures to realized volatility: realized power variation (RPV) and realized bi-power variation (BPV). Both of these alternatives were introduced by BN-S (2004a, 2004b). The first, RPV, is based on summing powers of intraday absolute returns, while BPV is the sum of products of consecutive intraday absolute returns. Authors such as Ding et al. (1993), BN-S (2004b), Forsberg and Ghysels (2007), Ghysels et al. (2006), Ghysels and Sinko (2006), have demonstrated that RPV indeed improved the volatility forecasting. Not only this, but RPV and BPV are also immune to jumps. This indicates, then, that both RPV and BPV are excellent methods for predicting future volatility.

BN-S (2004a, 2006), have shown that RV can be decomposed into one continuous component, known as 'realized bi-power variation (BPV)', as well as a discontinuous jump component. Incorporating a measure of jumps is important because their relative contribution to the total variation is about 7% as noted by Huang and Tauchen (2005). The recent studies on jump issue include test specification, (Lee and Mykland, 2007), as well as nonparametric estimation in the presence of jumps.(Mancini and Reno, 2006).

In Corsi (2009) a simple Heterogeneous Auto-Regressive (HAR-RV) model has been introduced for realized volatility in order to capture the long memory of volatility in a parsimonious manner. The HAR-RV model provides an additive cascade of various volatility components, each of which is generated by the actions of different types of market participants. The main idea is that agents with different time horizons perceive, react to, and cause, different types of volatility components. Typically, three primary volatility components are used: the short-term traders with daily or higher trading frequency, the medium-term investors, who typically rebalance their positions weekly, and the long-term agents with a characteristic time of one or more months. The idea of heterogeneity of volatility component stems from the so-called Heterogeneous Market Hypothesis presented by Miller et al. (1993), which recognizes the presence of heterogeneity across traders.

Andersen et al. (henceforth ABD) (2007), first incorporated the jumps into the HAR-RV model to obtain a non-parametric HAR-RV-CJ model, using related bi-power variation measures and adopting the jump test of BN-S (2004a, 2006). In doing so, it was found that the jumps were not useful in predicting future volatility. However, Corsi et al. (2009) provided an alternative intraday volatility estimator, the 'corrected threshold bipower variation', or CTBPV. It was demonstrated that the apparent puzzle found in ABD was due to a measurement bias, introduced by the bi-power variation in finite samples. Specifically, this happened when two jumps occurred in the same daily trajectory. In contrast, the CTBPV estimator was nearly unbiased in the presence of jumps. Empirical analysis (on the S&P500 index, single stocks and US bond yields) has shown that the newly proposed techniques significantly improved the accuracy of volatility forecasts, especially during periods following the occurrence of a jump.

Equity and stock-index volatilities exhibit a significant asymmetric response to past return. Early studies by Black (1976) and Christie (1982) found that volatility rises when stocks prices go down, and decreases when stock prices go up. This asymmetry in the relationship between equity market returns and volatility is known as the ‘leverage effect’. Figlewski and Wang (2001) found that the magnitude of the effect of negative return realization on future volatilities is too significant to be explained on the basis of changes in firm's financial leverage alone. Thus, the ’leverage effect’ should more properly be termed a ‘down market effect’. Moreover, French, Schwert, and Stambaugh (1987), Engle and Ng (1993), Zakoian (1994), Bekaert and Wu (2000), Wu (2001), and more recently Bollerslev et al (2006) have all pointed out that volatility changes are negatively correlated with returns. In view of the demonstrated importance of volatility leverage in explaining the negative relation between return and volatility, Corsi and Reno (2009) developed the LHAR-RV model. This model incorporates not only daily negative returns, but also their weekly and monthly aggregation, into the HAR-RV model.

Typically, there is a positive correlation between volatility and trading volume, a relationship that has been examined extensively. Two theoretical hypotheses designed to explain this connection exist. One, the ‘Mixture of Distributions Hypothesis’ (MDH) literature (see Clark, 1973; Epps and Epps, 1976; Tauchen and Pitts, 1983 and Harris, 1987), which is based on the tenet that volatility and trading volume are jointly driven by the unobservable information flow. Indeed, MDH helps to explain the high degree of positive relationship between volumes and volatility (see Karpoff, 1987). Lamoureux and Lastrapes (1990) and Andersen (1996) have suggested that trading volume can serve as a proxy measure of the latent information flowing into the market. More recently, Manganelli (2005) and Bowe et al. (2007) maintain trading volume conveys relevant information relating to market conditions and may have a direct effect upon prices. The second hypothesis is 'Sequential Information Arrival Hypothesis' (SIAH), developed by Copeland (1976), Morse (1981), and Jennings and Barry (1983). In SIAH, new information flows into the market to generate both trading volume and price movement in a sequential manner. Thus, the SIAH suggests that lagged trading volume may have explanatory power for predicting current volatility, and vice versa.. Gallant et al. (1992) and Bessembinder and Seguin (1993) documented evidence which also supports a positive relationship between volume and volatility.

As described above, the main contribution of this paper is to incorporate lagged trading volume into the LHAR-RV model and attempt to examine the role of trading volume as well as to improve the forecasting performance of realized volatility. This study refers to the modified model as LHAR-RV-cum-Vol model. To the best of our knowledge, no published study has yet modeled and forecasted realized volatility with adding the lagged trading volume into the LHAR-RV model.

In addition to investigating the impact of trading volume on future volatility for Taiwan's stock market, this paper also investigates whether average trade size or number of transaction provide the best explanation of price volatility. Admati and Pfleiderer (1988), and Foster and Vishwanathan (1990) suggested that informed traders may strategically break a large trade into many trades of smaller sizes. Thus, the number of transactions may actually carry more information than trade size. Easley and O'Hara (1990) demonstrated that market makers can also learn from a lack of transactions and the length of no-trading periods. Therefore, the number of transactions is a crucial variable in understanding the process of price formation.

Meanwhile, Harris and Raviv (1993) have argued that transactions occur due to traders' different assessments of the impact of information on stock prices. They predicted that the number of transactions would have a positive impact on the absolute value of price changes. Jones, Kaul and Lipson (1994), using Nasdaq data, found empirical evidence in support of Harris and Raviv's prediction. They also suggested that the impact of average trade size on price volatility was dominated by the impact of the number of transactions on price volatility. Therefore, they concluded that it was the number of transactions, and not trading volume, that possessed the most informational content. Moreover, using Nasdaq data, Gopinath and Krishnamurti (2001) reported that the number of trades had a larger impact on volatility than the average trade size. This finding was further corroborated for the Taiwan OTC market data by Chiang et al. (2006). These publications all concluded that the number of trades indeed produced more information than average trade size.

Furthermore, this paper also investigates whether bid-ask frequency or bid-ask volume provide the most information for explaining price volatility. From a supply and demand point of view, bid-ask frequency (volume) represents supply and demand of the stock market. Since that bid information and ask information can serve as a proxy measure of the quantity demanded and quantity supplied in the stock market, bid-ask information corresponds to excess supply or excess demand. Using Taiwan Stock Exchange (TSE) market data, Chen (2005) provided empirical evidence that bid-ask volume has a much larger impact on volatility than the impact of bid-ask frequency.

Finally, there is an interesting yet unaddressed issue within different market conditions in the context of volatility forecasting using the HAR-RV model and its new variant models. In particular, this paper examines whether different market conditions (e.g. up market day and down market day) generate different empirical results for value-added intraday data information in volatility forecasting. Using GARCH models, Fuertes et al. (2008) showed that the additional use of intraday data for day *t-1* to forecast volatility on day *t* is more advantageous when *t-1* is an up market day. The GARCH-RPV model ranks top in both regimes, based upon 14 NYSE stocks.

The remainder of this paper is organized as follows. Section 2 briefly reviews the theoretical framework behind the concept of realized volatility and methodology used for developing and testing the forecasting models. Section 3 presents a brief description of the data and illustrates the empirical in-sample and out-of-sample results on a long series of high frequency TAIEX data. Section 4 contains some concluding remarks.

2. Research Methodology

2.1 Modeling volatility

Assume that the state variable *p(t)*, for example the logarithmic price of a stock , is driven by the continuous time stochastic volatility jump-diffusion process:

$$dp(t) = \mu(t)dt + \sigma(t)dw(t) + \kappa(t)dq(t), 0 \le t \le T \quad (1)$$

where $\mu(t)$ is a continuous and locally bounded variation process, $\sigma(t)$ is the stochastic volatility process, $w(t)$ denotes a standard Brownian motion, $dq(t)$ is a counting process with $dq(t) = 1$ corresponding to a jump at time t and $dq(t) = 0$ corresponding to no jump, a jump intensity $\lambda(t)$, and $\kappa(t)$ refers to the jump size. The quadratic variation (QV) process of $p(t)$ can be defined by

$$[p](t) = p \to \lim_{n\to\infty} \sum_{j=0}^{n} (p(s_{j+1}) - p(s_j))^2 \quad (2)$$

for any sequence of partitions $0 = s_0 < s_1 < ... < s_n = t$ with $\sup_j \{s_j - s_{j-1}\} \to 0 \ for \ n \to \infty$.

The most important aim is attempt to predict the increments in quadratic variation over certain horizons, H, is then:

$$QV_{t,t+H} \equiv \int_t^{t+H} \sigma^2(s)ds + \sum_{t \le s \le t+H} \kappa^2(s) \quad (3)$$

where the first component, referred to as integrated volatility, is from the continuous component of (3), and the second term is the contribution derived from discrete jumps.

This paper employs the intraday data on the Taiwan stock index to predict future volatility using the trading hours between 9:00 a.m. and 13:30 p.m., Monday-Friday. Let the discrete daily returns be denoted by $r_{t,t-1} = 100(\ln P(t) - \ln P(t-1))$, where the time index t refers to the day of sampling. This study normalizes the daily time interval to unity and divides it into M periods. Each period has length $\Delta = 1/M$. Then define the Δ period return as $r_{t,j}^{M} = 100(p(t - j/M) - p(t - (j-1)/M))$, $j = 1,2,...,M$, where M is the sampling frequency. This paper sets M = 54 since this corresponds to the five-minute sampling frequency as is adopted by Andersen et al. (2001, 2005) and BN-S (2004b). ABDL (2001) claimed that sampling at five-minute intervals is sufficient to ensure that there is minimal measurement error in the daily realized volatilities, while also preventing microstructure biases from becoming a concern. This paper also defines daily 'realized volatility' (RV), or quadratic variation, which can be estimated by the sum of the corresponding M intra-daily squared returns, as follows:

$$RV_{t,t+1}^{M} = \sum_{j=1}^{M} (r_{t,j}^{M})^2 \tag{4}$$

This is a consistent estimator of $QV_{t,t+1}$, as M $\rightarrow$ ∞, see BN-S (2002a, b) and ABDL (2003) for a review. In this case, realized volatility consists of integrated volatility plus the jump component. Other measures of realized volatility, introduced by BN-S (2004a), are realized power variation (RPV), and realized bi-power variation (BPV), which this study defines as:

$$RPV_{t,t+1}^{M} = \mu_1^{-1} M^{-1/2} \sum_{j=1}^{M} \left| r_{t,j}^{M} \right| \tag{5}$$

$$BPV_{t,t+1}^{M} = \mu_1^{-2} \sum_{j=2}^{M} \left| r_{t,j}^{M} \right| \left| r_{t,j-1}^{M} \right| \tag{6}$$

where $\mu_1 \equiv \sqrt{2/\pi} = E(|Z|)$ denotes the mean of the absolute value of the standard Gaussian random variable, Z. In particular, this study also defines the standardized realized tri-power quarticity (TQ), adopted in the bi-power jump test (BN-S, 2004b; 2006), as follows:

$$TQ_{t,t+1}^{M} = M\ \mu_{4/3}^{-3} \sum_{j=3}^{M} (\left|r_{t,j}^{M}\right| \left|r_{t,j-1}^{M}\right| \left|r_{t,j-2}^{M}\right|)^{4/3} \tag{7}$$

where $\mu_{4/3}^{-3} \equiv 2^{2/3}\Gamma(7/6)\Gamma(1/2) = E(|Z|^{4/3})$. It is straightforward to drive that for M→∞,

$$\lim_{M\to\infty} RPV_{t,t+1}^{M} \to \int_{t}^{t+1} \sigma^2(s)ds \equiv \sigma_{t,t+1}, \tag{8}$$

$$\lim_{M\to\infty} BPV_{t,t+1}^{M} \to \int_{t}^{t+1} \sigma^2(s)ds \equiv \sigma_{t,t+1}^{2}, \tag{9}$$

and

$$\lim_{M\to\infty} TQ_{t,t+1}^{M} \to \int_{t}^{t+1} \sigma^4(s)ds \equiv \sigma_{t,t+1}^{4}. \tag{10}$$

Hence, BPV provides a consistent estimator of the integrated variance and it is also immune to jumps.

Corsi et al. (2009) provides an alternative estimator of the continuous part of volatility, the Corrected Threshold Bipower Variation (CTBPV):

$$CTBPV_t = \frac{2}{\pi}\sum_{j=2}^{M} Z_1(r_{t,j},\theta_j) Z_1(r_{t,j-1},\theta_{j-1}) \qquad t = 1,\ldots,T \tag{11}$$

where $Z_1(r_{t,j},\theta_j)$ is a special function equal to $|r_{t,j}|$ when $r_{t,j} < \theta_j$, and equal to $1.094\sqrt{\theta_j}$ when $r_{t,j} \geq \theta_j$, and θ_j is the threshold that is a multiple of local variance, $\hat{V}_j$, that is chosen according to an iterative procedure, that is:

$$\theta_j = c_\theta \cdot \hat{V}_j \tag{12}$$

A typical value of c_θ is $c_\theta = 3$. Even if $CTBPV_t$ converges to IV_t when M → ∞, it is possible to show that $CTBPV_t \to BPV\ as\ c_\theta \to \infty$. As will be shown in the next sub-section, this correction is essential for building test statistics. Finally, this study also defines the standardized corrected realized threshold tri-power quarticity (CTTQ), adopted in the corrected threshold bi-power jump test (Corsi et al., 2009), as follows:

$$CTTQ_{t,t+1}^{M} = M\ \mu_{4/3}^{-3} \sum_{j=3}^{M} Z_{4/3}(r_{t,j},\theta_j) Z_{4/3}(r_{t,j},\theta_{j-1}) Z_{4/3}(r_{t,j},\theta_{j-2}) \tag{13}$$

where $\mu_{4/3}^{-3} \equiv 2^{2/3}\Gamma(7/6)\Gamma(1/2) = E(|Z|^{4/3})$, and $Z_{4/3}(r_{t,j},\theta_j)$ is a special function equal to $|r_{t,j}|^{4/3}$ when $r_{t,j} < \theta_j$, and equal to $1.129 \cdot \theta_j^{2/3}$ when $r_{t,j} \geq \theta_j$. It is also easy to show that for $\mathrm{M} \rightarrow \infty$,

$$\lim_{M\to\infty} CTTQ_{t,t+1}^{M} \rightarrow \int_t^{t+1} \sigma^4(s)ds \equiv \sigma_{t,t+1}^4 . \tag{14}$$

2.2 Modeling Realized Jumps

In this sub-section, I employ two tests for jump detection to separate the jump and continuous sample path components of QV. The first jump test, introduced by BN-S (2004a, 2005), utilizes bipower variation to estimate the continuous integrated volatility and, by difference, the jump contribution to the total quadratic variation, defined as:

$$RV_{t,t+1}^{M} - BPV_{t,t+1}^{M} \rightarrow \sum_{t<s<t+1} \kappa^2(s) . \tag{15}$$

Following Huang and Tauchen (2005) and ABD (2007), this paper identifies significant jumps by using $Z_{t,t+1}$ given by (17) and Φ_α as the quantile function of the Normal distribution at confidence level α, thus, the jump component is given by:

$$J_{t,t+1}^{\alpha} = I(Z_{t,t+1} > \Phi_\alpha)(RV_{t,t+1} - BV_{t,t+1})^{+} \tag{16}$$

$$Z_{t,t+1} = \sqrt{M}\ \frac{(RV_{t,t+1} - BPV_{t,t+1}) / RV_{t,t+1}}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2} - 5)\max(1, TQ_{t,t+1} / BRV_{t,t+1}{}^{2})}} \tag{17}$$

where $x^{+} = \max(x, 0)$, and $I(*)$ denotes the indicator function which equal equals 1 if jumps are detected on day t, and equals to 0 elsewhere, and $Z_{t,t+1}$ is normally distributed if there are no jumps (BN-S, 2004b). In order to correctly separate $RV_{t,t+1}$ into continuous and jump component, the continuous component is naturally defined as:

$$C_{t,t+1}^{\alpha} = RV_{t,t+1} - J_{t,t+1}^{\alpha} \tag{18}$$

This ensures that $C_{t,t+1}^{\alpha}$ and $J_{t,t+1}^{\alpha}$ add up to $RV_{t,t+1}$. In the empirical study I set $\alpha = 0.999$ throughout the paper. Huang and Tauchen (2005) use Monte Carlo simulation to demonstrate that the z-statistic shown above has appropriate size, power, and jump detection ability.

The second jump test proposed by Corsi et al. (2009), employs corrected threshold bipower variation to estimate continuous integrated volatility. The residual jump component is then calculated as the difference between the realized volatility and the CTBPV:

$$RV_{t,t+1}^{M} - CTBPV_{t,t+1}^{M} \rightarrow \sum_{t<s<t+1} \kappa^2(s) \tag{19}$$

Corsi also introduced a correction of the z statistics of BN-S (2006), based on corrected threshold multipower variation, which identifies significant jumps as:

$$TJ_{t,t+1}^{\alpha} = I(C - Tz_{t,t+1} > \Phi_{\alpha})\,(RV_{t,t+1} - CTBPV_{t,t+1})^{+} \tag{20}$$

$$C - Tz_{t,t+1} = \sqrt{M}\,\frac{(RV_{t,t+1} - CTBPV_{t,t+1})/RV_{t,t+1}}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2} - 5)\max(1, CCTQ_{t,t+1}/CTBPV_{t,t+1}{}^{2})}} \tag{21}$$

It is possible to show that $C - Tz \rightarrow N(0,1)$ stably in law as M → ∞. The corresponding continuous component is defined as:

$$TC_{t,t+1}^{\alpha} = RV_{t,t+1} - TJ_{t,t+1}^{\alpha} \tag{22}$$

which is equal to $RV_{t,t+1}$ if $I(*) = 0$ and to $CTBPV_{t,t+1}$ if $I(*) = 1$. Corsi et al. (2009)

used Monte Carlo simulation to show that the C-Tz test has significantly more power than the z test, especially when jumps are consecutive, which is quite common in high-frequency data. Therefore, this paper employs the concept of BPV and corrected threshold BPV to estimate both components.

2.3 Modeling leverage effects

Financial asset volatilities often exhibit significant asymmetric response to past returns. In other words, volatility tends to increase more after a negative shock than after a positive shock of the same magnitude. This asymmetric return-volatility phenomenon is known as the 'leverage effect'. This sub-section is inspired by Corsi and Reno (2009) who found that not only daily but also weekly and monthly negative past returns (e.g. leverage effect) have a high forecasting power for future volatility. Hence, this paper defines daily returns $r_t = 100\,(p_t - p_{t-1})$ and past aggregated negative and positive returns as:

$$r^{(+)}_{t-H,t} = \frac{1}{n}(r_t + \ldots + r_{t-H+1}) I\,((r_t + \ldots + r_{t-H+1}) \geq 0\,) \tag{23}$$

$$r^{(-)}_{t-H,t} = \frac{1}{n}(r_t + \ldots + r_{t-H+1}) I\,((r_t + \ldots + r_{t-H+1}) < 0\,) \tag{24}$$

where $I(*)$ denotes the indicator function.

2.4 Modeling Trading Volume

Research on the importance of trading volume in financial markets is rapidly growing. Much of this research has been focused on the positive correlation between both trading volume and volatility, such as Gallant et al. (1992), and Bessembinder and Seguin (1993). Tseng (2009) suggested that including trading volume in the HAR-RV model provided more accurate predictions. Total trading volume can also be jointly determined by number of transactions, and average size per trade. From the viewpoint of the market, bid-ask frequency (volume) can represent supply and demand of the stock market. As a result, any discrepancy between bid-ask frequency (volume) carries information on excess supply or excess demand in the stock market.

In this sub-section, various measures of volume are employed, and daily volume is measured in five ways:

$$Trading\ Volume\ value\,(VOL_t\,)\text{-}\,type = \sum_{j=1}^{M} \ln vol_{t,j}^{M} \tag{25}$$

$$Number\ of\ Transactions\,(TNV_t\,)\text{-}\,type = \sum_{j=1}^{M} \ln tf_{t,j}^{M} \tag{26}$$

$$Average\ Trade\ Size\,(TSV_t\,)\text{-}\,type = \sum_{j=1}^{M} \ln(vol_{t,j}^{M} / tf_{t,j}^{M}) \tag{27}$$

$$Bid\text{ - }Ask\ Frequceny\,(TNR_t\,)\text{-}\,type = \sum_{j=1}^{M} \ln bf_{t,j}^{M} \Big/ \sum_{j=1}^{M} \ln af_{t,j}^{M} \tag{28}$$

$$Bid\text{ - }Ask\ Volume\,(TVR_t\,)\text{-}\,type = \sum_{j=1}^{M} \ln bv_{t,j}^{M} \Big/ \sum_{j=1}^{M} \ln av_{t,j}^{M} \tag{29}$$

It should be noted that these types are defined as the summation of the corresponding M intra-daily logarithm of difference volume measures, with these intra-daily logarithm volumes representing the rate of information arrival into the market within the same time period, such as five-minute intervals; larger trading volume implies a more rapid rate of information arrival. Notice that TNR_t (TVR_t) is defined as bid frequency (volume) divided by ask frequency (volume) to measure excess supply or excess demand.

2.5 The forecasting Models

The Heterogeneous Auto-Regressive (HAR) model introduced by Corsi (2003) and Corsi et al. (2009) can effectively capture the long-term memory behavior of RV quite parsimoniously. Moreover, Corsi (2003) provided empirical evidence that the HAR model is able to reproduce the observed hyperbolic decay of the sample autocorrelations of RV. Hence, this paper employs the HAR model to forecast the RV.

The HAR model uses averaged future RV as the dependent variable and uses averages of past values of variance measures as the independent variables. This allows the models to take advantage of information from past price variation. We will define the multi-period normalized realized variation over H discrete periods as:

$$RV_{t,t+H} = H^{-1}(RV_{t+1} + RV_{t+2} + ... + RV_{t+H}), \quad RV_{t,t+1} \equiv RV_{t+1} \tag{30}$$

In this paper, the values 1, 5, 10, 15, and 20 are used for H, referring to one-day, weekly, bi-weekly, tri-weekly and monthly frequencies respectively. To keep the HAR model simple and intuitive, the HAR-RV model of Corsi (2003), including only the daily, weekly and monthly RV components, is then expressed as:

$$HAR-RV-X : RV_{t,t+H} = \beta_0 + \beta_d X_{t-1,t} + \beta_w X_{t-5,t} + \beta_m X_{t-20,t} + \varepsilon_{t,t+H} \tag{31}$$

where $\varepsilon_{t,t+H}$ is a standard IID noise and where X= RV, BPV, CTBPV, RPV, C, TC, with C and TC denoting the continuous and threshold continuous components of RV, respectively. This study follows the HAR-RV model introduced by Corsi (2003) and uses the regressors such as RV, BPV, CTBPV, RPV, C and TC for predicting RV.

In addition to the HAR-RV model, this paper also uses the following model suggested by ABD (2007) and Chung et al. (2008).

$$HAR-RV-CJ: RV_{t,t+H} = \beta_0 + \beta_{cd}C_{t-1,t} + \beta_{cw}C_{t-5,t} + \beta_{cm}C_{t-20,t} + \beta_{jd}J_{t-1,t} + \beta_{jw}J_{t-5,t} + \beta_{jm}J_{t-20,t} + \varepsilon_{t,t+H} \quad (32)$$

where C and J denote the continuous and jump components of RV, as separated by the jump test of BN-S (2006).

Corsi et al. (2009) proved that corrected threshold BPV (CTBPV) is an unbiased estimation of $\sigma^2_{t,t+1}$ in the presence of jumps, but the BPV estimator is a biased measure of $\sigma^2_{t,t+1}$ in days where jumps are present. For this reason, this paper modifies the HAR-RV-CJ model of Andersen et al. (2007) and Chung et al. (2008) using the continuous and jump component of RV as separated by the jump test of Corsi et al. (2009) using the CTBPV measure.

$$HAR-RV-TCJ: RV_{t,t+H} = \beta_0 + \beta_{cd}TC_{t-1,t} + \beta_{cw}TC_{t-5,t} + \beta_{cm}TC_{t-20,t} + \beta_{jd}TJ_{t-1,t} + \beta_{jw}TJ_{t-5,t} + \beta_{jm}TJ_{t-20,t} + \varepsilon_{t,t+H} \quad (33)$$

where TC and TJ are the continuous and jump components. This paper computes these regressors analogously to the realized volatility measures given by Equation (30), including $BPV_{t-5,t}$, $BPV_{t-20,t}$, $CTBPV_{t-5,t}$, $CTBPV_{t-20,t}$, $RPV_{t-5,t}$, $RPV_{t\text{-}20,t}$, $C_{t\text{-}5,t}$, $C_{t\text{-}20,t}$, $J_{t\text{-}5,t}$, $J_{t\text{-}20,t}$, $TC_{t\text{-}5,t}$, $TC_{t\text{-}20,t}$, $TJ_{t\text{-}5,t}$, *and* $TJ_{t\text{-}20,t}$.

Another form of the HAR regression that is used in this paper incorporates leverage effects into the HAR-RV model, and then proposes the following specification to obtain the LHAR-RV model:

$$LHAR-RV-X: RV_{t,t+H} = \beta_0 + \beta_d X_{t-1,t} + \beta_w X_{t-5,t} + \beta_m X_{t-20,t} + \beta_{rd} r^{(-)}_{t-1,t} + \beta_{rw} r^{(-)}_{t-5,t} + \beta_{rm} r^{(-)}_{t-20,t} + \varepsilon_{t,t+H} \quad (34)$$

where X = RV, BPV, CTBPV, RPV, C, TC and where $r^{(-)}$ denotes the ‘leverage effect’ components, given by Equation (24). Note that in order to keep the model as parsimonious as possible, only the negative returns as suggested by Corsi and Reno (2009) are included.

In what follows, the HAR-RV-TCJ model is extended to directly incorporate the leverage effects to obtain the LHAR-RV-TCJ model of Corsi and Reno (2009):

$$\begin{aligned} LHAR-RV-TCJ : RV_{t,t+H} &= \beta_0 + \beta_{cd} TC_{t-1,t} + \beta_{cw} TC_{t-5,t} + \beta_{cm} TC_{t-20,t} \\ &+ \beta_{jd} TJ_{t-1,t} + \beta_{jw} TJ_{t-5,t} + \beta_{jm} TJ_{t-20,t} \\ &+ \beta_{rd} r^{(-)}_{t-1,t} + \beta_{rw} r^{(-)}_{t-5,t} + \beta_{rm} r^{(-)}_{t-20,t} + \varepsilon_{t,t+H} \end{aligned} \tag{35}$$

where the definitions of TC and TJ are the same as those in Equation (33).

As suggested by Tseng (2009), this paper also incorporates lagged trading volume into the LHAR-RV and LHAR-RV-TCJ model; hence, the proposed models read:

$$\begin{aligned} & LHAR-RV-X-cum-Vol: \\ & RV_{t,t+H} = \beta_0 + \beta_d X_{t-1,t} + \beta_w X_{t-5,t} + \beta_m X_{t-20,t} \\ & \qquad + \beta_{rd} r^{(-)}_{t-1,t} + \beta_{rw} r^{(-)}_{t-5,t} + \beta_{rm} r^{(-)}_{t-20,t} \\ & \qquad + \beta_{vd} Vol_{t-1,t} + \beta_{vw} Vol_{t-5,t} + \beta_{vm} Vol_{t-20,t} + \varepsilon_{t,t+H} \end{aligned} \tag{36}$$

$$\begin{aligned} & LHAR-RV-TCJ-cum-Vol: \\ & RV_{t,t+H} = \beta_0 + \beta_{cd} TC_{t-1,t} + \beta_{cw} TC_{t-5,t} + \beta_{cm} TC_{t-20,t} \\ & \qquad + \beta_{jd} TJ_{t-1,t} + \beta_{jw} TJ_{t-5,t} + \beta_{jm} TJ_{t-20,t} \\ & \qquad + \beta_{rd} r^{(-)}_{t-1,t} + \beta_{rw} r^{(-)}_{t-5,t} + \beta_{rm} r^{(-)}_{t-20,t} \\ & \qquad + \beta_{vd} Vol_{t-1,t} + \beta_{vw} Vol_{t-5,t} + \beta_{vm} Vol_{t-20,t} + \varepsilon_{t,t+H} \end{aligned} \tag{37}$$

where X = RV, BPV, CTBPV, RPV, C, TC and where *Vol* defines the summation of the corresponding M intra-daily logarithm trading volume given by Equation (25). Also, the measures of lagged volume, including $Vol_{t-5,t}$ and $Vol_{t-20,t}$, are calculated using the same formula presented in Equation (30).

Following ABD (2007) and Forsberg and Ghysels (2007), this paper also models RV using square root and log-transform methods. The square root forms of the above equations are as follows:

$$HAR-RV^{1/2}-X : RV^{1/2}_{t,t+H} = \beta_0 + \beta_d X_{t-1,t} + \beta_w X_{t-5,t} + \beta_m X_{t-20,t} + \varepsilon_{t,t+H} \tag{38}$$

where X = $RV^{1/2}$, $BPV^{1/2}$, $CTBPV^{1/2}$, RPV, $C^{1/2}$, $TC^{1/2}$.

$$\begin{aligned} HAR-RV^{1/2}-CJ : RV^{1/2}{}_{t,t+H} &= \beta_0 + \beta_{cd} C^{1/2}_{t-1,t} + \beta_{cw} C^{1/2}_{t-5,t} + \beta_{cm} C^{1/2}_{t-20,t} \\ &+ \beta_{jd} J^{1/2}_{t-1,t} + \beta_{jw} J^{1/2}_{t-1,t} + \beta_{jm} J^{1/2}_{t-1,t} + \varepsilon_{t,t+H} \end{aligned} \tag{39}$$

and the same transformations will be estimated for models (33) to (37).

The logarithmic forms of the above equations are as follows:

$$HAR-\ln RV-X : \ln RV_{t,t+H} = \beta_0 + \beta_d \ln X_{t-1,t} + \beta_w \ln X_{t-5,t} + \beta_m \ln X_{t-20,t} + \varepsilon_{t,t+H} \tag{40}$$

where X = RV, BPV, CTBPV, RPV, C, TC.

$$\begin{aligned} &HAR-\ln RV-CJ : \\ &lnRV_{t,t+H} = \beta_0 + \beta_{cd} \ln C_{t-1,t} + \beta_{cw} \ln C_{t-5,t} + \beta_{cm} \ln C_{t-20,t} \\ &\qquad + \beta_{jd} \ln(J_{t-1,t}+1) + \beta_{jw} \ln(J_{t-5,t}+1) + \beta_{jm} \ln(J_{t-20,t}+1) + \varepsilon_{t,t+H} \end{aligned} \tag{41}$$

and the same transformations will be estimated for models (33) to (37).

2.6 Measure of Performance

In their study, Andersen et al. (2007) compared the results of the different models using only adjusted R^2. The report demonstrated hat the adjusted R^2 was highest when modeling the log transform of the realized volatility. However, Forsberg and Ghysels (2007) and Chung *et al*. (2008) continued to place more focus on ‘mean square error’

(MSE) than adjusted R^2. There are two reasons cited for this. The first reason was based on the argument of Forsberg and Ghysels (2007) who suggested that, when transformed variables (such as log or square root) are used as dependent variables, the adjusted R^2 from the regressions with different dependent variables are not comparable.

The second reason was related to the recent work of Hansen and Lunde (2006) and Patton (2006). They have shown that the MSE loss function is robust with regards to the volatility proxy used. Therefore, this study will use the adjusted R^2 and 'root mean square error' (RMSE) evaluation measures in the in-sample forecasts, but only RMSE in the out-of-sample forecasts.

This paper takes the inverse transformation of $RV_{t,t+H}^{1/2}$ and $\ln RV_{t,t+H}$, and then computes RMSEs. To be specific, let $RV_{t,t+H}$ be the true in-sample value of RV for H days, and let $\hat{R}V_{t,t+H}$ be the in-sample prediction value of the dependent variable. Additionally, the following RMSE evaluation measure is employed for the different transformed measures:

$$\text{origin-type:} \quad RMSE = \sqrt{N^{-1}\sum_{t=1}^{N}(RV_{t,t+H} - \hat{R}V_{t,t+H})^2} \tag{42}$$

$$\text{square roots-type:} \quad RMSE = \sqrt{N^{-1}\sum_{t=1}^{N}(RV_{t,t+H} - (\hat{R}V_{t,t+H}^{1/2})^2)^2} \tag{43}$$

$$\text{log-type:} \quad RMSE = \sqrt{N^{-1}\sum_{t=1}^{N}(RV_{t,t+H} - \exp(\ln \hat{R}V_{t,t+H}))^2} \tag{44}$$

where N is the number of in-sample forecasts. Hence, after recovering the dependent variable to its original form, this paper can compare the in-sample and out-of sample predictability for different models with different regressors.

3 Empirical Analysis

3.1 Data Source and Descriptions

The data set analyzed in this paper is the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) time series from the Taiwan Economic Journal (TEJ) database. Also, all high-frequency transactions from 2 January 2003 to 30 June 2008 are disposed of, a total of 1361 trading days. In order to mitigate the impact of microstructure effects on the estimates, as in ABD (2007), this study sets a sampling frequency as $\Delta = 5$ minutes, corresponding to 54 returns per day. Table 1 reports the descriptive statistics of the realized volatility levels. These realized volatility levels include realized variance (RV_t), realized bi-power variance (BPV_t), corrected threshold bi-power variance ($CTBPV_t$), realized power variance (RPV_t), the threshold continuous element (TC_t), the threshold jump element (TJ_t), and the jump element (J_t), as well as their square root transforms and log-transforms. The columns from two to seven of Table 1 report the sample mean, standard deviation, skewness, kurtosis, and the minimum and maximum of the different variables. The last column (LB_{10}) reports the Ljung-Box test statistic for up to tenth order serial correlation and the critical value of LB_{10} is 18.307.

First, the LB statistic demonstrates that RPV_t has the highest serial correlation and $CTBPV_t$ has the second highest correlation for the original variables and their transformations. For the RPV_t the Ljung-Box statistic is 4821.6 compared to 2563.3 for the $CTBPV_t$. For the square root transform the LB statistic is 5377.9 for the $RPV_t^{1/2}$ against

4006.4 for the $CTBPV_t^{1/2}$. For the log-transform this study finds the highest LB-statistic again for the $lnRPV_t$ which is 5650.2, which implies that RPV_t may provide a better predictable power on future volatility. It is worth noting that TJ_t and J_t have the lowest serial correlation for the level and their transformations, while the TJ_t is much higher than J_t. This finding shows that the jump element measured by means of threshold bipower variation will be better forecasted.

Second, RV_t has the highest standard deviation and more volatile than others. Finally, all the square-transformed and square-root-transformed variables are severely skewed to the right. In contrast, the skewness values for the log transform are quit close to symmetric value of 0 except for the TJ_t and J_t. The estimates of the sample kurtosis are well above the normal value of 3 for all the transformations, indicating that the distributions of these variables are highly leptokurtic. This is especially true for the RV_t of Taiwan Stock Exchange (TSE) returns as argued by Andersen et al. (2001) (henceforth ABDE) and ABDL (2001, 2003).

Figure 1 displays the time series plots of logarithm realized variance and logarithms of the threshold continuous and jump elements. Also displayed is the jump element computed using the bi-power variation of *RV*. Clearly, the jump is quit large in TJ_t and J_t plot since that the average logarithms of the threshold jump and jump elements for the TAIEX are 0.494 and 0.442 respectively, whilst Forsberg and Ghysels (2007) found that the average logarithm of jump for the S&P500 was only 0.023. Figure 2 reports the percentage contribution of threshold jumps estimated by Equation (20) to total quadratic variation computed on a 1-month moving average window for the full sample. Jumps contribute between 10% and 40% of total variation with an overall sample mean of about 25.7%. This finding is not in line with the results in Andersen et al. (2007), Huang and Tauchen (2005) and Corsi and Reno (2009). Their empirical work documents the mean of jumps contribution as only about 6% or 7% of total variance.

Chung et al. (2008) provided two possible explanations for the occurrence of higher jump percentages for the TAIEX. Firstly, market participants in the Taiwan stock market are mostly individual investors, known as noise traders. Their investment decisions are often undertaken without sound and rational analysis, and can be greatly affected by macroeconomic news announcements. Secondly, the Taiwan stock market has insufficient market depth and as a result, major investors can artificially generate volatility through excessive trading.

Figure 3 shows the sample autocorrelation function (SACF) of RV_t, BPV_t, $CTBPV_t$, RPV_t, TC_t, TJ_t, and J_t. It is clear that RPV_t is the most persistent, followed by BPV_t, $CTBPV_t$, and TC_t. Similar to ABD (2007) and Forsberg and Ghysels (2007) this study finds little persistence in the threshold jump and jump components of the realized variance.

Descriptive Statistics on daily trading volume are presented in Table 2. Five measures of volume are examined, including the daily trading volume value (VOL_t), the daily number of transactions (TNV_t), the daily average trade size (TSV_t), the daily bid-ask frequency (TNR_t), and the daily bid-ask volume (TVR_t). These measures of volume are transformed by taking their logarithms as the detrended versions of the corresponding measures. The last column of Table 2 reports the Ljung-Box test statistic for up to tenth-order serial correlation (LB_{10}); the corresponding the critical value is 18.307. The LB statistics indicate strong autocorrelation in all the volume variables. VOL_t is the highest followed by TNV_t and TSV_t. TNR_t and TVR_t are less persistent. Additionally, the standard deviation of VOL_t is 23.038 which is the highest variable, meaning that the trading volume value is more volatile.

Regarding sample skewness, the value of VOL_t and TNV_t are -0.091 and 0.048 which is quite close to symmetric value of 0, while the TSV_t, TNR_t and TVR_t are clearly asymmetric. Turning to the estimates of the sample kurtoses, this study finds that only the value of VOL_t is close to the normal value 3, indicating that the distribution is approximately Gaussian. In contrast, the values of the TSV_t, TNR_t and TVR_t are well above the normal value of 3, indicating that the distributions are highly leptokurtic.

Table 3 presents the Pearson correlation matrix between the five measures of daily trading volumes. Notably, it can be seen here that the highest correlation is between trading volume value and the number of transactions, while the second highest correlation is between trading volume and average trade size. This finding may support the idea of using TNV_t and TSV_t as volume explanation variables to replace VOL_t.

Figure 4 presents the logarithm of trading volumes including VOL_t, TNV_t, TSV_t, TNR_t, and TVR_t. As is obvious from the figure, periods with volumes above the mean are often followed by periods of volumes below the mean. Further, the first three panels show an increasing trend toward the end of the sample period.

Summarizing, RPV_t and $CTBPV_t$ exhibit the highest serial correlation, which implies that RPV_t and $CTBPV_t$ will be better at predicting future volatility. However, the threshold jump and jump elements of RV_t are much less persistent, while the jump element is less persistent than the threshold jump element. Vol_t is most persistent and volatile, and TVR_t is much less persistent and volatile. This study finds that the there is a high correlation between trading volume value and number of transactions. Other proxies of volume, such as bid-ask frequency, has a very low correlation with trading volume value.

3.2 In-sample Results

The data set covers a long time span of nearly 6 years of high frequency data for the TAIEX. In this section, the main determinants of future realized volatility are analyzed, and the performance of various HAR models in predicting future realized volatility are examined.

3.2.1 Comparing Volatility Forecasts with Two Jump Detection Methods

The primary purpose of this sub-section is to analyze the impact of jumps on future volatility when threshold bipower variation is employed as a measure of jumps. This study shall measure the corrected threshold bi-power variation with a value $c_\theta = 3$, using the C-Tz statistics to detect jumps. The C-Tz statistics are not only computed with a confidence level α = 99.9% but the most interesting quantities will be computed and plotted for different values of α as well.

According to Corsi et al. (2009), the C-Tz statistics provide a more effective estimation of the jump component than the z statistics. Furthermore, they proved that the z statistics were biased and noisy when the jumps appeared in the form of two consecutive jumps in the intra-daily returns. This bias can be completely removed by the corrected threshold estimators.

To compare the detecting power of the C-Tz and z statistics, this study computes the number of days which contain jumps in the TAIEX sample. Jumps are detected by the condition C-Tz $> \Phi_{\alpha}$ and z $> \Phi_{\alpha}$ as a function of α and plots the results in Figure 5. Thus, by using the statistics based on corrected threshold bi-power variation, it can be seen that this study achieves a higher number of jumps than were achieved by BN-S (2004b, 2006), using the statistics based on bi-power variation.

To further evaluate the relative contribution of the newly proposed C-Tz statistics, this paper compares the estimation results with those of the standard HAR model (31), the HAR-RV-CJ model (32), and the HAR-RV-TCJ model (33), as is adopted by Corsi et al. (2009). This study will also consider the standard deviation and logarithm forms following ABD (2007) and Forsberg and Ghysels (2007). Empirical results are reported in Tables 4, 5 and 6, where all jumps have been estimated with the C-Tz statistics. Three prediction horizons are considered: one-day, one-week, and four-week, corresponding to $RV_{t,t+H}$, for H = 1 , 5 and 20, together with their statistical significance evaluated with the Newey-West robust t-statistic.

As is seen in Table 4, most of the estimates for β_d, β_w, and β_m of the three models are highly significant. The only exception is the coefficient of daily continuous component. This result may be due to the fact that the time series of the TAIEX realized volatility seems to exhibit a higher level of noise, due to a lower mean tick arrival frequency and higher market microstructure effects, such as price discreteness. These microstructure effects make the daily continuous component have an insignificant impact on future volatility. Meanwhile, the weekly and monthly continuous components, being averaged over long periods, arguably contain less noise and more information on the volatility process. Therefore, higher weights are received from these models. However, this confirms the existence of highly persistent volatility dependence for all the transformations. Importantly, the coefficient (daily, weekly and monthly) of jump component as measured using BN-S (2004b, 2006) and Corsi et al. (2009) is not significant in many cases. This result also shows that only the weekly jump coefficient is positive and significant at least 10 % level. This suggests that the weekly jump may play a role in future daily volatility forecast for the Taiwan Stock Exchange (TSE) market. It is noteworthy that the constant term in these regressions is always significant, suggesting that all the regression models are biased. Most importantly, the HAR-RV-TCJ model yields a higher R^2 and a lower RMSE, thus showing a better forecasting power.

To understand this point in depth, the sample is divided into days immediately following the occurrence of a jump, and days following days with no jumps, as introduced by Corsi et al. (2009). On these two samples this computes the adjusted R^2 and

RMSE statistics separately, denoting them by J-R^2, J-RMSE and C-R^2, C-RMSE, respectively. The results reported in Table 4 show that the HAR-RV-TCJ model greatly improves the forecasting power on realized volatility in days which do not follow a jump, and it is also slightly outperforming in days immediately following a jump. These results are not in line with the findings reported by Corsi et al. (2009), which suggests that using the newly proposed C-Tz statistics provide a superior forecasting, especially in days which follow the occurrence of a jump. These results demonstrate that not only are C-Tz statistics superior at measuring the jump component, but they also remove noise from the continuous component in the explanatory variables. Therefore, this study introduces a superior method for future realized volatility forecasting in the continuous component.

This study also examines the forecasting models for H = 5 and 20. Results are reported in Tables 5 and 6, respectively. Again, Table 5 shows that the estimates of β_d, β_w, and β_m, which quantify the impact of the continuous sample path variability on the total future variation, are all generally highly significant. The coefficient (daily, weekly and monthly) of the jump component is also not significant in many cases. The result shows that the daily and weekly jump coefficients are positive and significant up to at least a level of 10%. This suggests that the daily and weekly jumps may play a role in future weekly volatility forecasts. Furthermore, the *t*-statistics of the coefficient β_{jd} is larger for the HAR-RV-TCJ model than for the HAR-RV-CJ model. Again, the adjusted R^2 and the RMSE suggest that the HAR-RV-TCJ model has better forecasting ability than the HAR-RV-CJ model. Also, both days after a jump and days without one show that the

HAR-RV-TCJ model provides superior forecasts compared to the HAR-RV-CJ model measured in terms of adjusted R^2 and the RMSE. This is especially true for the days not following a jump. Table 6 presents almost the same results as reported in Table 5. However, it should be noted that the impact of weekly jump components on future monthly volatility is insignificant. The impact of the monthly jump component on future monthly volatility, however, is in fact significant for the HAR-RV-TCJ model. This suggests that the daily and monthly jumps may play a role in future monthly volatility forecasting.

Thus, these results reveal that since the newly proposed C-Tz statistics can effectively estimate the jump component and completely remove jump noise from the continuous component, it has superior detecting power over the z test and enhances the forecasting ability on future realized volatility.

3.2.2 The LHAR-RV-TCJ Model for Predicting Future Realized Volatility

In sub-section 3.2.1, this paper demonstrates that the HAR-RV-TCJ model is better at predicting future realized volatility. However, it was suggested in the recent studies by Figlewski and Wang (2001) and Bollerslev et al. (2006) that equities and stock indexes often exhibit the so called “leverage effect”, i.e. volatility rises when stocks prices go down, but decreases when stock prices go up. So it can be seen that the leverage effect exhibits a negative correlation between past returns and future volatility. Thus, this paper incorporates leverage effects into the HAR-RV-TCJ model to obtain the so-called

LHAR-RV-TCJ model newly proposed by Corsi and Reno (2009). The purpose of this sub-section, then, is to demonstrate that the newly proposed model indeed improves the performance of realized variance forecasting.

The in-sample regression results of the LHAR-RV-TCJ model for the TAIEX considering five prediction horizons. One-day and one to four-week periods are presented in Table 7, together with their statistical significance, evaluated with the Newey-West robust t-statistic. Table 7 shows that all the coefficients of the three continuous volatility components are positive and, in general, highly significant. Interestingly, the weekly continuous component affects future volatility more strongly than the daily and monthly continuous components. This suggests that the weekly continuous component may play a role in future daily volatility forecasting. The hierarchical asymmetric propagation of the volatility cascade is also confirmed by these results. The impact of weekly volatility decreases with the forecasting horizon, while the impact of monthly volatility increases. The coefficient which measures the impact of monthly volatility on future daily volatility is approximately three times than that of daily volatility in future monthly volatility. These findings are consistent with Corsi (2009) and Corsi and Reno (2009). As noted by Muller et al. (1997), Arneodo (1998) and Lynch and Zumbach (2003), who suggested that volatility over longer time intervals has greater influence than volatility over shorter time intervals, not the other than around.

For the jump components, most of jump coefficient estimates are insignificant, while the daily jump components remain highly significant and positive. This suggests that the daily jump may play a role in future daily volatility forecasting. It is worth mentioning that the impact of daily and weekly jump components decreases with the horizon, but the impact of monthly jump component increases with the horizon. This finding indicates that monthly jumps affect future volatilities more significantly over a long period of time. For the leverage components, the coefficients of the negative returns in the LHAR-RV-TCJ model are generally negative and highly significant, especially at the daily aggregation frequency, which unveils a so-called "leverage effect".

The square root transformed regression results are presented in Table 8. The daily, weekly and monthly coefficients of continuous component are almost significant at the 1 % level. This confirms the existence of highly persistent volatility dependence. A similar hierarchical structure is also confirmed in the square root model.

For the jump and leverage components, the results are the same as reported in Table 7. The daily jumps remain positive and highly significant. This result also reveals the strong significance of the negative returns at only the daily aggregation frequency.

The log-transformed regression results are presented in Table 9. Again, the estimates for β_{cd}, β_{cw}, and β_{cm} confirm the existence of highly persistent volatility dependence. Most of the jump components are highly significant and positive, indicating that when they are measured by means of threshold bipower variation, the jump terms will have a substantial impact in determining future volatility. This empirical result is not in line with ABD (2007), which suggest that the jump components are not statistically significant and slightly increase the R^2 of the regression.

This result also indicates that not only the daily negative returns, but also the negative returns of the past week and past month, affect the next day volatility. This finding further confirms the views of the Heterogeneous Market Hypothesis; especially that heterogeneity originates from the difference in the time horizons. Most importantly, in order to evaluate the relative contribution of the newly proposed model, this paper compares their in-sample predictions with those of the standard HAR model and the HAR-RV-TCJ model with jumps, but with no leverage effects, using corrected threshold bipower variation as in Corsi et al. (2009), presented in Table 10. Panel B and C denote the square root and log transformation of the variance. The results presented in Panel A, B and C also give the highest adjusted R^2 and lowest RMSE for the LHAR-RV-TCJ model at any forecasting horizon. These empirical results demonstrate that the LHAR-RV-TCJ model including the leverage effects significantly improves the forecasting performance of the TAIEX volatility at any forecasting horizon. This study also finds that the HAR-RV-TCJ model outperforms the HAR model at any forecasting horizon in all the transformations.

3.2.3 Is leverage effect induced by jumps?

The empirical results of Christie (1982), and French, Schwert, and Stambaugh (1987) all suggest that stock price changes and volatility are inversely related (the so-called leverage effect). According to the previous discussion in section 3.2.2, the empirical results also suggest that the leverage effect is a vital explanatory component on the future

TAIEX volatility. In this sub-section, we set out to examine whether the leverage effect is induced by jumps.

Corsi and Reno (2009) indicates that the leverage effect is only somewhat attributable to jumps, and that it appears instead as a feature mostly induced by continuous returns. This study closely follows Corsi and Reno (2009) separating the daily jump contribution to quadratic variation into positive and negative parts, which are defined as:

$$\begin{aligned} J_t^+ &= J_t \cdot I(r_t > 0) \\ J_t^- &= J_t \cdot I(r_t < 0) \end{aligned} \tag{45}$$

and this study replaces J_t in the LHAR-RV-TCJ model with J_t^+ and J_t^- to obtain the newly model called LHAR-RV-TCJ$^+$ proposed by Corsi and Reno (2009). This paper also estimates the HAR-RV-TCJ$^+$ model but with no leverage terms. Results are reported in Tables 11, 12 and 13, corresponding to prediction horizon H=1(one day), 5(one week) and 20(one month). In Tables 11,12 and 13, estimating the HAR-RV-TCJ$^+$ model, this finds that the impact of negative jumps as measured by the corresponding coefficient in the regression, is significantly larger than that of positive jumps, and this is true at prediction horizon H=5 and 20. Moreover, the coefficient of positive jumps is sometimes negative but the coefficient of negative jumps is always positive at prediction horizon H=1. This result enhances the asymmetry in positive and negative jumps. Additionally , when this study estimates the full LHAR-RV-TCJ$^+$ model with the leverage terms, the impact of negative and positive jumps is estimated to be roughly the same for all the transformations at prediction horizon H= 5 and 20. At the same time, the coefficients of negative jumps remain positive, while those of positive jumps remain negative, at prediction horizon H=1.

This result is evidence of the fact that the leverage effect is indeed attributable to jumps for a short-run prediction horizon. This empirical result is not consistent with Corsi and Reno (2009). However, leverage is not induced by jumps for the longer-run prediction horizon. Instead, it appears to be a feature primarily induced by continuous returns. This empirical result is in line with Corsi and Reno (2009).

3.2.4 The Reformed Model for Prediction Future Realized Volatility

The prior studies have clearly demonstrated that the LHAR-RV-TCJ model is better at predicting future realized volatility (Corsi and Reno, 2009). The impact of trading volume measured in value (dollar volume or notional value) on realized volatility, however, was not discussed in the LHAR-RV-TCJ model proposed by Corsi and Reno (2009). Anderson (1996) suggested that according to the MDH concept, trading volume could serve as a proxy measure of the amount of unobservable information flowing into the market. Therefore, this study incorporates lagged trading volume value into the LHAR-RV-TCJ model in an attempt to improve the performance of realized volatility forecasting. The reformed forecasting model is referred to as the LHAR-RV-TCJ-cum-Vol model.

The in-sample regression results of the LHAR-RV-TCJ-cum-Vol model for TAIEX are presented in Table 14, with the adjusted R^2 and the "root mean squared error" (RMSE) for the one-day and one- to four-week in-sample RV predictions. In the first case, the estimates of $\beta_{cd,}$ β_{cw} and β_{cm} are mostly positively significant for all the prediction horizons, indicating that the *RV* also exists in highly persistent volatility dependence. The coefficient of the jump component shows that only the daily jump coefficient is positive and significant up to at least 5% level, with the exception of the one-day prediction horizons. The coefficient of the negative returns also reveals that only the daily negative return is positively significant to a level of 1 %.

In the second case, the estimates of the three lagged trading volume components are mostly positive but, in general, not significant. This result suggests that trading volume as an explanation variable is not helpful in this regression.

The square root form regression results are presented in Table 15. These results are similar to the results reported in Table 14. This study finds that the continuous, jump and negative returns components are the primary determining factors for future realized volatility. Again, trading volume as the regressors has no significant explanatory power on future volatility.

The log-transformed regression results are presented in Table 16. As expected, all the coefficients of the three continuous components are positive and highly significant. A similar hierarchical structure is confirmed by these results. Indeed, the impact of daily and weekly volatility decreases with the prediction horizon of future volatility, while the impact of monthly volatility increases. The daily jump component remains highly significant and positive, with the exception of the three-week prediction horizons, indicating that daily jumps may play a role in future volatility forecasting. The estimation of model (37) also reveals the strong significance of the negative returns at the daily aggregation frequency. It is noteworthy that the impact of the daily negative returns decreases with the prediction horizon of future volatility. Additionally, the negative returns of the past week also have a significant impact on future volatility, when computed over a two week period. Most importantly, however, this study finds that the estimates of β_{vd} are positively significant to a level of at least the 10%, with the exception of the three-week prediction horizons. This indicates that lagged trading volume indeed has a significant forecasting power on future volatility. As pointed out by Karpoff (1987), Gallant et al. (1992) and Bessembinder and Seguin (1993), there is a positive relationship between volatility and trading volume. The majority of the coefficients of trading volume in the far past, such as β_{vw}, β_{vm}, however, are revealed to be insignificant.

This result demonstrates that the lagged weekly and monthly volumes have relatively weaker impact on future volatility than the lagged daily volume. As noted by Tseng (2009), the impact of the lagged volume on the future volatility decreases from high to low aggregation frequency.

In order to evaluate the relative contribution of the newly proposed model, this paper compares its in-sample prediction with that of the LHAR-RV-TCJ model; leverage effects are included but trading volume is not. The results of Table 17 unambiguously indicate that the inclusion of the lagged trading volume considerably improves the forecasting performance of the TAIEX volatility at any prediction horizon for all the transformations. In line with Tseng (2009), this study finds that the difference in the RMSE between the LHAR-RV-TCJ and LHAR-RV-TCJ-cum-Vol model increases with the prediction horizon of future volatility. This is because the standardized measure of the multi-period for the RV, leverage effects, and volume over longer horizons, contribute to easier RV predictions. The percentage of reduction in the RMSE of the LHAR-RV-TCJ-cum-Vol model, for instance, ranges from 0.36% (H=1) to 8.46% (H=20).

Thus, this study has successfully demonstrated that the LHAR-RV-TCJ-cum-Vol model can significantly improve the performance of future realized volatility forecasting after adding trading volumes as the regressors. Coefficients of the trading volume components, however, are insignificant for any prediction horizon. This implies that the trading volume components have no explanatory capabilities.

All of this data clearly indicates that the inclusion of lagged trading volume (the proxy measure of the amount of unobservable information flowing into stock market) in the reformed model does indeed affect future realized volatility forecasting. As noted by Bowe et al. (2007) and Manganelli (2005), the trading volume contains relevant market information about the financial asset's true price movement.

3.2.5 In-sample Results for Different Measures of Trading Volume

Trading volume can be determined by number of trades (i.e. trading frequency) and average trade size. From the viewpoint of the market, bid-ask frequency (volume) can represent supply and demand. As a result, the discrepancy between bid-ask frequency (volume) carries information on excess supply or excess demand in the stock market. This sub-section will empirically analyze which of these will contribute the most to the future realized volatility forecasting for the Taiwan capital market. Using the LHAR-RV-TCJ-cum-Vol model, trading volume, trading frequency, average trade size, bid-ask volume, and bid-ask frequency will all be compared. The results of the in-sample predictions of $RV_{t,t+H}$ for H = 1, 5, 10, 15, and 20, using the five measures of trading volume are presented in Table 18. The left-hand column under each prediction horizon report the adjusted R^2 and the right-hand columns report the RMSE. This study also considers the square root and logarithm models. Five measures of trading volume are examined: trading volume value (VOL), number of trades (TNV), average trade size (TSV), bid-ask frequency (TNR), and bid-ask volume (TVR), corresponding to Equations (25) to (29).

We will first consider the results obtained by implementing the square volatility measure, which are presented in Panel A of Table 18. These results indicate that the highest adjusted R^2 value is achieved with trading frequency as the explanatory variable; the results range between 0.209 for the one-day horizon and 0.514 for the four-week horizon among trading volume value, trading frequency, and average trade size. The results for RMSE also reveal that trading frequency dominates all other measures of volume, with the RMSE of trading frequency ranging between 0.808 and 2.165 across five prediction horizons. This finding is consistent with Gopinath and Krishnamurti (2001) and Chiang et al. (2006).

In the second case, the highest adjusted R^2 and the lowest RMSE are both achieved with bid-ask volume as the explanatory variable for two- to five-week prediction horizons between bid-ask volume and bid-ask frequency. The highest adjusted R^2 and lowest RMSE, meanwhile, are achieved with bid-ask frequency for one-day and one-week prediction horizons.

The results of the $RV_{t,t+H}$ predictions using the square root volatility measure with different measures of volume regressors are presented in Panel B of Table 18. Again, trading frequency achieves the highest adjusted R^2 among trading volume value, trading frequency, and average trade size. This is with the adjusted R^2 of trading frequency ranging between 0.393 and 0.652, while the RMSE of trading frequency ranges between 0.211 and 0.442 across the five prediction horizons. With the exception of the one-day prediction horizons, bid-ask volume ratio achieves the highest adjusted R^2, and the lowest RMSE, between bid-ask volume ratio and bid-ask frequency ratio across the five prediction horizons.

The results of the $RV_{t,t+H}$ predictions using the logarithmic volatility measure with different measures of volume regressors are presented in Panel C of Table 18. First of all, trading frequency achieves the highest adjusted R^2 and the lowest RMSE between trading volume value, trading frequency, and average trade size. This indicates that the number of transactions (trading frequency) may actually carry more information than the average trade size does. This result is consistent with Gopinath and Krishnamurti (2001). This empirical result suggests that, in the Taiwan stock exchange (TSE) market, the number of transactions is more important in explaining the future realized volatility than the average trade size. Admati and Pfleiderer (1988), and Foster and Vishwanathan (1990) explained that informed traders may strategically break a large trade into many trades of smaller size, so the number of transactions could contain more information.

Secondly, bid-ask volume achieves the highest adjusted R^2 between bid-ask volume and bid-ask frequency, with the adjusted R^2 of trading frequency ranging between 0.468 and 0.682. The results for RMSE also reveal that bid-ask frequency is dominated by bid-ask volume. This paper using the LHAR-RV-TCJ-cum-Vol model, finds that the empirical results are also in line with results of Chen (2005), which demonstrated that bid-ask volume was better at predicting future realized volatility.

3.2.6 In-sample Results for Different Volatility Predictors

Recent literature suggests that realized power variation (RPV) is the most effective predictor of future volatility. Such publications include the work of Ghysels et al. (2006), ABD (2007), Forsberg and Ghysels (2007), Fradkin (2008) and Chung et al. (2008), to name a few. This study employs eight measures of realized variance: realized volatility (RV), realized bi-power variation (BPV), corrected threshold bi-power variation (CTBPV), realized power variation (RPV), continuous (C), jump and continuous (CJ), threshold continuous (TC), and threshold continuous and jump (TCJ).

RPV outperforms other volatility measures for the forecasting of future volatility, for several reasons. First, BN-S (2004) showed that realized power variation is immune to jumps. Jumps are generally large outliers that may have a strong effect on model estimates and forecasts. Second, Taylor (1986) and Ding et al. (1993) found that the absolute value of returns displays stronger persistence than squared returns. In other words, RPV may provide a better signal than RV when predicting volatility. Third, Forsberg and Ghysels (2007), Ghysels et al. (2006) and Ghysels and Sinko (2006) demonstrated that absolute returns enhance volatility forecasts. Lastly, Forsberg and Ghysels (2007) argue that the gains are due to the higher predictability, smaller sampling error, and immunity to jumps which means jumps don't have any affect on the model.

To the best of our knowledge, the issue of different market conditions, in the context of volatility forecasting using different regressors and different model structures, has not been previously addressed. Hence, in this sub-section, this study sets out to examine the in-sample fit for realized variance with different regressors, using the standard HAR model, the LHAR-RV model, and the LHAR-RV-cum-Vol model. The next sub-section will then focus on the performance of these models in forecasting future volatility during 'up-market days' (U) versus 'down-market days' (D).

The results of the in-sample predictions of $RV_{t,t+H}$ for the different regressors are presented in Table 19, with the left-hand columns under each prediction horizon reporting the adjusted R^2, and the right-hand columns reporting the RMSE. The results using the standard HAR models are reported in Panel A of Table 19. These results indicate that RPV, as a regressor, achieves the highest adjusted R^2 and lowest RMSE, followed by TCJ and

TC. It is interesting to note that the predictive ability on future volatility between RPV and TCJ is nearly the same. This is also true for BPV and CTBPV. For the HAR-RV-TC and HAR-RV-TCJ regressions, the difference in adjusted R^2 between TC and TCJ regressors is quite small, increasing the fit by at most 0.03. This indicates that there is not much to gain by separating RV into its continuous and jump parts and modeling them separately. The difference in RMSE between TC and TCJ regressors also yields similar results. Further, the adjusted R^2 of TC and TCJ regressors is higher than that of C and CJ regressors. This suggests the use of the tests and measures proposed by Corsi et al. (2009), which provide a better identification and more precise measurement of jumps. Interestingly, the RMSE decreases for all regressors as the prediction horizon increases, an indication that RV computed over longer horizons are easier to predict since they are smoother series, as discussed in Forsberg and Ghysels (2007).

The results of the in-sample predictions of $RV_{t,t+H}$ using the LHAR-RV model with different regressors, are presented in Panel B of Table 19. Again, RPV achieves the highest adjusted R^2 ranging between 0.194 and 0.433 in the LHAR-RV model, followed by TCJ and TC. The results for RMSE also reveal that RPV dominates almost all other regressors. The differences in the adjusted R^2 and RMSE between BPV and CTBPV are quit small, indicating that using CTBPV as the regressor, as proposed by Corsi et al. (2009), yields no better predictive power on future volatility than BPV. For the LHAR-RV-TC and LHAR-RV-TCJ regressions, using TC and TCJ as regressors produces essentially the same results; indicating that only the continuous component has predictive power. In other words, the jump component is simply ‘noise, due to the fact that the jump component is less persistent. This study also does not find a large improvement in explanatory power from dividing the continuous and jump components, as suggested by ABD (2007). From the previous discussion, the possible explanation for this result is that the new jump test proposed by Corsi et al. (2009) improves significantly the accuracy of volatility forecasts especially in periods not following a jump.

The results of the $RV_{t,t+H}$ predictions using the LHAR-RV-cum-Vol models with different regressors are presented in Panel C of Table 19; in general, these results are found to be similar to the results reported in Panel A and B of Table 19. Panel C, however, yielded a different result, finding that with TCJ as the regressor, the adjusted R^2 is the highest and RMSE is the lowest. Obviously, this means that TCJ has a more accurate predictive ability than RPV. In other words, RPV is not a top forecaster anymore in the models. This result confirms that when corrected threshold bipower variation is employed as a measure of jumps, the threshold continuous and jump components can improve the forecasting of future volatility.

Summarizing, the best regressors are RPV and TCJ. This study does not find that explanatory power is significantly improved by dividing the continuous and jump components using corrected threshold bipower variation. Nevertheless, threshold-based tests indeed perform much better than bipower variation-based tests. These results are consistent with the findings of Corsi et al. (2009). It has also been demonstrated that RPV, BPV and CTBPV all outperform the RV. This confirms Forsberg and Ghysels’ (2007) finding that the realized absolute value is a better forecaster of future realized variance than realized variance itself.

The square root transform regression results are presented in Table 20. The results using the HAR-$RV^{1/2}$ models with different regressors are presented in Panel A of Table 20. $TCJ^{1/2}$ almost achieves the highest adjusted R^2 within the HAR-$RV^{1/2}$ model, with the adjusted R^2 of $TCJ^{1/2}$ ranging from 0.363 for the on-day horizon, to 0.589 for the three-week horizon. Turning to the RMSE of the in-sample predictions, it is clear that $TCJ^{1/2}$ also dominates in RMSE terms. Comparing across prediction horizons, the RMSE is always lowest when $TCJ^{1/2}$ is used as the regressor, with the RMSE of $TCJ^{1/2}$ ranging between 0.941 and 2.229 across the five prediction horizons. For the HAR-$RV^{1/2}$-RPV regression, the RMSE ranges between 0.946 and 2.233 across the five prediction horizons. Notice again that the performance between RPV and $TCJ^{1/2}$ is almost the same. For the HAR-$RV^{1/2}$-$TC^{1/2}$ and HAR-$RV^{1/2}$-$TCJ^{1/2}$ regressions, the results of the difference in adjusted R^2 between $TC^{1/2}$ and $TCJ^{1/2}$ as regressors are quite small, with increasing the fit by at most 0.030. Again, this indicates that the jump component of RV is of little help in predicting future volatility. The difference in RMSE between the TC and TCJ regressors also yields similar results. It is obvious that the C and CJ regressors do not outperform the TC and TCJ regressors as independent variables in these regressions. Again, this study finds that $CTBPV^{1/2}$ does not seem to be of any help in improving forecasting. These results hold up for Panel B and C of Table 20, using the LHAR-$RV^{1/2}$ and LHAR-$RV^{1/2}$-cum-Vol models, respectively.

The log-transformed regression results are presented in Table 21. From the results of the HAR-ln RV, LHAR-ln RV, LHAR-ln RV-cum-Vol models reported in Panels A, B and C of Table 21, this study finds that ln RPV achieves nearly the highest adjusted R^2 and the lowest RMSE. This result is followed by ln TCJ, both for the HAR-ln RV and LHAR-ln RV models, while lnBPV achieves the highest adjusted R^2 and the lowest RMSE within the LHAR-ln RV-cum-Vol model. It is noteworthy that the predictive powers on future volatility provided by ln RPV and ln TCJ are identical. Again, Table 21 shows that the difference in adjusted R^2 and RMSE between the ln TC and ln TCJ regressors within the different models is quite small. Corsi and Reno (2009) explained that this is because the difference between the two forecasts is mainly in the days which come after a jump, which are quite few. Also, the ln TCJ regressor outperforms the ln CJ regressor in these models, although the ln C regressor outperforms the ln TC regressor. This result confirms that the threshold continuous and jump components seem to capture the main determinants of volatility dynamics.

It should be noted that when comparing the square and square root forms, the adjusted R^2 is higher with log-transformed regressors as it reduces the variance of the data. In other words, logarithmic models outperform the square root and square models, as noted by ABD (2007) and Forsberg and Ghysels (2007). In addition, when the RMSE of the HAR-RV-TCJ is compared with the HAR-$RV^{1/2}$-$TCJ^{1/2}$ and HAR-lnRV-lnTCJ models, this study finds that for the last two models, the square root and log transform do appear to be detrimental. Considered from the RMSE perspective, HAR-RV-TCJ model is found to provide the best predictions. The same results are found when this paper models the realized volatility in the LHAR-$RV^{1/2}$ and LHAR-$RV^{1/2}$-cum-Vol models with $TCJ^{1/2}$ as the regressor. Interestingly, this seems to be the case for the RPV-based regressions as well. This finding is in line with Forsberg and Ghysels (2007) and Chung et al. (2009).

As noted above in the discussion of Table 20, TCJ indeed has more accurate predictive ability on future volatility than all other regressors, especially the RPV for square root transformation. This finding contradicts the results of Forsberg and Ghysels (2007). Importantly, there is significant gain to using TCJ and TC instead of CJ and C as the explanatory variables. These surprising results are clear evidence that using corrected threshold bipower variation not only better measures the jump components, it also removes noise from the continuous component in the explanatory variables better than the bipower variation of BN-S (2006).With this, the significant impact of the continuous and jump components on future volatility as noted by Corsi et al. (2009), is easily seen.

Finally, in order to evaluate the relative contribution of the inclusion of both the leverage effects and the jumps, this paper compares its in-sample prediction with those of the HAR-RV, LHAR-RV, and LHAR-RV-cum-Vol regressions. From an adjusted R^2 or RMSE perspective, this paper finds that with the same regressor, the best performance is obtained by the LHAR-RV-cum-Vol model. This implies that not only does the leverage effect accurately reflect the asymmetric responses of realized volatility, but also that lagged trading volume does indeed contain extra information that affects future realized volatility. This study also demonstrates that the LHAR-RV-TCJ-cum-Vol model provides the most accurate predictions, because the TCJ component successfully captures the dynamics in future volatility. Thus, the choice of regressor is clearly more important than either the model or the weighting scheme selected for use as noted by Ghysels et al. (2006) and Forsberg and Ghysels (2007).

To sum up, the results provided by the HAR-RV, LHAR-RV and LHAR-RV-cum-Vol models suggest that using realized power variance, threshold continuous and jump components as the regressors provides most accurate predictions, in terms of capturing the fluctuations in future volatility for these models. The results of these comparisons show that threshold jumps contribute only marginally to the performance of these models. Additionally, the corrected threshold bipower variation (CTBPV) proposed by Corsi et al. (2009) is a good estimation of the jump component of realized volatility, but it does not improve the forecasts.

3.2.7 In-sample Results of an Up-market and a Down-market

This sub-section is inspired by Fuertes et al. (2008) which compares the value-added intraday information for future volatility forecasting during an 'up-market' (U) versus a 'down-market' day (D). The definition of an up/down market day is based on the moving average of daily returns over the most recent 20-day window, defined as:

$$I_t^{(20)} = \begin{cases} 1 & \text{if} \quad \frac{1}{20}\sum_{i=1}^{20} r_{t-i+1} > 0 \quad \text{(Up - market day)} \\ 0 & \text{else} \qquad\qquad\qquad\qquad \text{(Down - market day)} \end{cases} \tag{46}$$

There are two questions to be asked: (a) Dose the performance of future volatility forecasting with different regressors differ over market conditions? (b) Do the benefits from exploiting intraday data differ over market conditions?

The results of the in-sample predictions of $RV_{t,t+H}$ for the various different regressors are presented in Table 22, which reports the adjusted R^2 and RMSE, respectively, for up- and down- market days as defined by Equation (46). For the sake of brevity, this study limits itself to the HAR-RV model, since this model provides the simplest, most parsimonious method of capturing the memory persistence of realized variance. The square transform regression results with different regressors are presented in Panel A of Table 22. In the first case, Panel A shows that TCJ, as a regressor, achieves the highest adjusted R^2 and the lowest RMSE for all prediction horizons in both regimes as pointed out by the previous discussion. Regardless of TC and TCJ, this study finds that the adjusted R^2 and RMSE of RPV, as a regressor, dominates all other regressors for all prediction horizons in down-market days. This result does not hold in up-market days, however, implying that RPV has more forecasting ability in down-market days than in up-market days. This interesting result implies that RPV regressors are invariant to jumps in down-market days, but not invariant to jumps in up-market days.

Additionally, for the HAR-RV-TC and HAR-RV-TCJ regressions, the inclusion of the jump component on down-market days increases the adjusted R^2, ranging from 21.58% for the one-day prediction horizon, to 0.98% for the three-week prediction horizon. Conversely, it decreases to -0.47% for the one-day prediction horizon and 0.86% for the three-week prediction horizon over up-market days. Similarly, the percentage reduction in the inclusion of jump component forecast errors for up-market days is 0.05% for the one-day prediction horizon and 0.68% for the three-week prediction horizon, compared to an increase to 8.15% and 1.28% over down-market days, respectively. This finding indicates that the jump element of RV is of little help in predicting future volatility in up-market days. In the second case, the forecast losses tend to be smaller for down-market days. This suggests that realized volatility at day t is relatively more difficult to forecast when t-1 is an up-market day.

The square-root transform regression results are presented in Panel B of Table 22. A glance at these values shows the superiority of the HAR-RV-RPV and HAR-RV-TCJ models in a down-market regime, and the superiority of the HAR-RV-TCJ model in an up-market regime. Regardless of TC and TCJ, this study also shows that the performance of RPV is greatly superior to other regressors in down-market days, in terms of the adjusted R^2 and RMSE. For the HAR-$RV^{1/2}$-$TC^{1/2}$ and HAR-$RV^{1/2}$-$TCJ^{1/2}$ regressions, the difference in adjusted R^2 between $TC^{1/2}$ and $TCJ^{1/2}$ regressors is larger in up-market days, but smaller in down-market days. This result is dissimilar to that of the square form. The explanation for this finding might be that the predictive ability in down-market days is originally high, so that adding the jump element only marginally contributes to the performance of the model. The result of the difference in RMSE between $C^{1/2}$ and $CJ^{1/2}$ regressors is ambiguous. Furthermore, the RMSE from down-market days is relatively smaller than that from up-market days. This finding is consistent with Fuertes et al. (2008), suggesting that the future volatility is easier to forecast when the market is underperforming (i.e. in down-market days).

The log-transformed regression results are presented in Panel C. This study notes that ln TCJ achieves almost the highest adjusted R^2 in both regimes; ln RV is also the best forecaster in up-market days. However, the results for RMSE reveal that ln RV achieves a low RMSE in down-market days, indicating that there is a smaller deviation between the actual and predicted values; meanwhile, ln TCJ dominates all other regressors in up-market days. Regardless of ln TCJ, ln TC and ln RV, ln RPV performs relatively well in down-market days. For the HAR-lnRV-lnTC and HAR-lnRV-lnTCJ regressions, the results of the difference in adjusted R^2 and RMSE between lnTC and ln TCJ regressors are similar to the results reported in Panel A. This study finds that there is a relatively significant gain derived from implementing lnTCJ as the explanatory variable in down-market days compared to up-market days. From the RMSE perspective, this paper can compare the HAR-RV-TCJ with the HAR-$RV^{1/2}$-$TCJ^{1/2}$ and HAR-ln RV-ln TCJ models over different market conditions. For the last two models, the square and log transform do appear to be detrimental and indeed the HAR-RV-TCJ model is found to provide the best predictions.

Overall, it is not difficult to see that the most tangible benefits from exploiting intraday data in order to predict a future volatility occur during a bearish market. Fuertes et al. (2008) explained that if market is bearish, volatility is higher during down days than up days. This effect is exacerbated because, as Admati and Pfleiderer (1988) demonstrate, trades from both informed and discretionary liquidity traders come in clusters, with both groups preferring to trade during 'thick' markets[1] in order to minimize transaction costs. This clustering of trades, when trading activity is already high, triggers the release of even more information, which aids forecasting future volatility in down-market days. It is also possible high trading activity may mitigate the microstructure noise (e.g. infrequent-trading effects); this would explain why the HAR-RV model tends to provide better forecasts during down-market days. Hence, the use of intraday data is relatively more beneficial during down-market days.

Take summarize section 3.2, this paper shows that jumps can be effectively detected using the newly proposed C-Tz statistics and that the HAR-RV-TCJ model provides a superior forecasting, especially in days which do not follow a jump. The inclusion of the leverage effects in the HAR-RV-TCJ model noticeably improves the forecasts and the model confirms the views of the Heterogeneous Market Hypothesis. Moreover, the leverage effect is indeed attributable to jumps for short-run prediction horizons, while the leverage effect is not induced by jumps for long-run prediction horizons. Trading volume contains stock-market-relevant information and as a result the modified model, the so-called LHAR-RV-TCJ-Cum-Vol model, improves the performance of future volatility forecasting. This study also reveals that TNV provides more information on explaining volatility and that TVR has a significantly greater impact on volatility than TNR does. The results of comparing the forecast performance of different volatility predictors

---

[1] In Lippman and McCall (1986), a 'thick' market indicates that more transactions of a homogeneous good (an equity) take place in a unit of time, so-called a 'liquid' market.

suggest that RPV and TCJ are the top forecasters. This study also notes that the TJ regressor contributes only marginally to the forecast performance. From market conditions analysis, TCJ is the top forecaster in both regimes. Additionally, RPV performs relatively well in down-market days as compared to up-market days. Importantly, the forecast losses tend to be smaller on down-market days, indicating that the volatility is relatively is difficult to forecast within up-market days.

3.3 Out-of-sample Results

In this sub-section, out-of-sample predictions of the realized volatility of TAIEX are reported. The data is split into two parts to provide an in-sample section for estimating the models, and an out-of-sample section for measuring the predictions. The in-sample period covers 2 January, 2003 to 29 December, 2006, for a total of 994 days. The out-of-sample period is 2 January, 2007 to 30 June, 2008, for a total of 367 days.

In terms of predictive ability, the in-sample fit measures clearly demonstrate greater accuracy for the LHAR-RV-cum-Vol model compared to the HAR-RV and LHAR-RV models; the measures also indicate that the TCJ and RPV elements are the most effective forecaster for predicting future volatility. Since the primary purpose of this study is to seek out improvements in the forecasting of future volatility, it is important to determine whether the superior performance of the LHAR-RV-cum-Vol model is confirmed in the out-of-sample forecasts. Table 23 reports the out-of-sample results for the HAR-RV, LHAR-RV, and LHAR-RV-cum-Vol models, as well as their square root transforms and log-transforms. Panel A of Table 23 presents results pertaining to the prediction of $RV_{t,t+H}$. The out-of-sample results confirm the in-sample results; with the same regressor, and for all prediction horizons, all of the RMSE results are significantly lower in the LHAR-RV-cum-Vol regressions than those in the HAR-RV and LHAR-RV regressions. Compared to the other regressors within these models, RPV achieves the lowest RMSE for all the prediction horizons, making it by far the most preferable. This empirical result is consistent with the results achieved by both Forsberg and Ghysels (2007) and Chung et al. (2008), which indicate that the predictable features of absolute returns (realized power variation) are shown to improve forecasting the usual squared return-based measures of volatility. Interestingly, the TCJ regressor does not outperform the TC regressor.

Panel B of Table 23 reports the results for the square root transform. Again, RPV-based measures continue to provide the best out-of-sample predictions. Most importantly, it shows that $TCJ^{1/2}$ models slightly improve the forecasting power with respect to $TC^{1/2}$ models. As noted by ABD (2007), jumps have a null impact in determining future volatility when the jump component is estimated through the concept of corrected threshold bipower variation. Furthermore, $CTBPV^{1/2}$ does not provide more accurate out-of-sample predictive ability than $BPV^{1/2}$.

Panel C of Table 23 reports the results for the log transform. The results are displayed on Panel A and B. Surprisingly, it seems that ln RV is the top forecaster for all the prediction horizons. In light of the findings of Corsi et al. (2009), it is not surprising to see that ln TCJ produces a better predictive power than lnTC. Indeed, the RMSE is always lower for the LHAR-lnRV-cum-Vol model than for the HAR-lnRV and LHAR-lnRV models. These findings suggest that the LHAR-RV-cum-Vol model does improve the forecasting of future volatility for all the transformations with the results holding for both the in-sample and out-of-sample forecasts.

4 Concluding Remarks

This paper provides five contributions to the literature. The first contribution is to show that C-Tz statistics, which are based on corrected threshold bipower variation, are more efficient in detecting jumps than z statistics, which are based on bipower variation. The HAR-RV-CJ model is also evaluated, using the jumps detected by the z statistics, as in BN-S (2004a, 2006), and compared with the HAR-RV-TCJ model estimated with the jumps detected by the C-Tz statistics. The results indicate that the predictability of the HAR-RV-TCJ model is more accurate than HAR-RV-CJ model, especially in days which do not follow the occurrence of a jump. Meanwhile, the continuous volatility components present the hierarchical asymmetric propagation of the volatility cascade.

The second contribution is the incorporation of the leverage effects into the HAR-RV-TCJ model enhances the forecasting of future volatility. Results reveal not only daily but also weekly and monthly negative past returns are highly significant. This empirical finding confirms the view of the Heterogeneous Market Hypothesis. In addition, this study demonstrates that the leverage effect is indeed attributable to jumps for one-day prediction horizons. However, the leverage is not linked to jumps for the one-week or four-week prediction horizons; it appears instead to be a feature primarily induced by continuous returns.

The third contribution of this paper is to present a new improved model for volatility forecasting. This model extends the LHAR-RV-TCJ model by incorporating lagged trading volume, which is a proxy for the rate of information arrival into the market.

Results find that the only daily trading volume and daily jump components appear to play a role in future volatility forecast and the LHAR-RV-TCJ-cum-Vol model shows remarkably good forecasting performance. This result seems to support the Sequential Information Arrival Hypothesis and Mixture of Distributions Hypothesis. In particular, the number of transactions as a proxy for information flows provides the best predictive ability on the volatility. The informational content of the bid-ask volume appear to be higher than that of the bid-ask frequency in the TSE market.

The fourth contribution of this paper finds that the absolute returns-based volatility measure is better at forecasting than the squared returns-based volatility measure. Besides, RPV and TCJ under different transformations are the preferred regressors for future volatility predictions. The magnitude and occurrence of jumps in price does not have a significant effects on future volatility even if jumps are effectively detected using the newly proposed C-Tz statistics. This suggests that it is not necessary to separate the continuous and jump components of volatility for the purpose of forecasting. However, given the inadequacy of bipower variation in measuring volatility in the presence of jumps, this paper employs the tests and measures introduced by Corsi et al. (2009) to separate the RV into the continuous and jump components; this indicates that the TC and TCJ regressors indeed outperform the C and CJ regressors as independent variables. Moreover, CTBPV does not provide a better forecast of future volatility than BPV. These results hold good for both the in-sample and out-of-sample data.

Last, but definitely not least, using intraday data to model and forecast future volatility in 'up-market' and 'down-market' days, results suggest that TCJ is the most preferred regressor to predict future volatility in both regimes. It also appears that the jump component contains more information about future price movements when the market is down. Moreover, the RPV regressor performs relatively well in down-market days, indicating that it is invariant to jumps in down-market days, but it is variant to jumps in up-market days. In other words, there is no change in the RPV repressor with regard to its ability to predict future volatility in down-market days. Interestingly, the RMSE tends to be smaller for down-market days. That is, when the market is down the amount of market information increases, the predictive ability of future volatility also increases.

As the future research directions, we may include the realized volatility of international stock markets in the HAR-RV models. Meanwhile, for checking the adequacy of the fitted model, we should examine the sample ACF and the time plot of the residual series. Visual examination of such a residual plot often is useful in detecting problems with the estimated model. Besides, we believe choosing an optimal sampling frequency could improve the forecasting performance of realized volatility. These issues warrant further study in the future.

References

Admati, A.R., and Pfleiderer (1988), "A theory of intraday patterns: Volume and price variability," *Review of Financial Studies*, 1, 3-40.

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

Andersen, T.G. and T. Bollerslev (1998), "Answering the Skeptics: Yes, Standard volatility models do provide accurate forecasts," *International Economic Review*, 39, 885-905.

Andersen T.G., T. Bollerslev, F.X. Diebold (2007), "Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility," *Review of Economics and Statistics*, 89, 701–720.

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

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys (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 N. Meddahi (2005), "Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities," *Econometrica*, 73(1), 279-296.

Arneodo, A., J. Muzy, and D. Sornette (1998), "Causal cascade in stock market from the infrared to the ultraviolet," *European Physical Journal*, B2, 277-282.

Barndorff-Nielsen, O.E. and N. Shephard (2002a), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society*, 64, 253-280.

Barndorff-Nielsen, O.E. and N. Shephard (2002b), "Estimating Quadratic Variation Using Realized Variance," *Journal of Applied Econometrics*, 17, 457-478.

Barndorff-Nielsen, O.E., and N. Shephard (2004a), "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1-48.

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

Barndorff-Nielsen O.E., N. Shephard (2005), "Variation, jumps, market frictions and high frequency data in financial econometrics," Working paper, Nuffield College, University of Oxford.

Barndorff-Nielsen, O.E., and N. Shephard (2006), "Econometrics of testing for jumps in financial economics using bi-power variation," *Journal of Financial Econometrics*, 4, 1-30.

Bekaert, G. and Wu, G. (2000), "Asymmetric Volatility and Risk in Equity Markets," *The Review of Financial Studies*, 13, 1-42.

Bessembinder, H., and P. Seguin (1993), "Price volatility, trading volume, and market depth: Evidence from futures markets," *Journal of Financial and Quantitative Analysis*, 28, 21-39.

Black, F. (1976), "Studies of stock price volatility changes," Proceeding of the American Statistical Association, Business and Economic Statistics Section, 177-181.

Bollerslev, T., J. Litvinova and G. Tauchen (2006), "Leverage and volatility feedback effects in high-frequency data," *Journal of Financial Econometrics*, 4(3), 353-384.

Bowe, M., Hyde, S J., Lavern McFarlane (2007), "Duration, trading volume and the price impact of trades in an emerging futures market," Working paper.

Chiang, Yao-Min, Vivien W. Tai, and Robin K. Chou (2006), "Market condition, number of transactions, and price volatility: Evidence from an electronic, order driven, call market," *Managerial Finance*, 32(11), 903-914.

Christie, A.C. (1982), "The stochastic behavior of common stock variances-value, leverage and interest rate effects," *Journal of Financial Economics*, 3, 145-166.

Chen, Chun-Hung (2005), "Stock index volatility, number of transaction, and bid-ask volume: An evidence from Taiwan stock exchange," (in Chinese) Unpublished master's thesis, Department of Finance Chaoyang University of Technology.

Clark, P. (1973), "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica*, 41, 135-56.

Copeland, T. (1976), "A model of asset trading under the assumption of sequential information arrival," *Journal of Finance*, 31, 1149-68.

Corsi, F. (2003), "A Simple Long Memory Model of Realized Volatility," Manuscript, University of Southern Switzerland.

Corsi, F. (2009), "A simple approximate long-memory model of realized-volatility," *Journal of Financial Econometrics*, 7, 174-196.

Corsi, F., D., Pirinos and R., Ren. (2009), "Threshold bipower variation and the impact of jumps on volatility forecasting," Working paper, Dipartimento di Economia Politica Università di Siena.

Corsi, F. and Reno, R. (2009), "Volatility determinants: heterogeneity, leverage, and jumps," Working paper.

Ding Z, Granger CWJ, Engle RF. (1993), "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1, 83–106.

Easley, D., and M. O'Hara (1990), "The process of price adjustment in securities markets," Working paper, Cornell University.

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

Epps, T.W. and M.L. Epps (1976), "The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distribution hypothesis," *Econometrica*, 44, 305-321.

Figlewski, S., and X.Wang (2001), "Is the 'leverage effect' a leverage effect?," Working Paper, Department of Finance, New York University.

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

Foster, F.D., and S. Vishwanathan (1990), "A theory of the intraday variations in volume, variance and trading costs in securities markets," *Review of Financial Studies*, 3 , 593-624.

French, K.R., Schwert, G.W., Stambaugh, R.F. (1987), "Expected stock return and volatility," *Journal of Financial Economics*, 19, 3-29.

Fuertes, A-M and M. Izzeldin and E. Kalotychou (2008), "On forecasting daily stock volatility: The role of intraday information and market conditions," Working paper. Department of Economics, Lancaster University.

Gallant, A.R., P.E. Rossi, and G. Tauchen (1992), "Stock prices and volume," *Review of Financial Studies*, 50, 199-242.

Ghysels E, Santa-Clara P, Valkanov R. (2006), "Predicting volatility: how to get most out of returns data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.

Ghysels E, Sinko A. (2006), "Comment on Hansen and Lunde JBES paper," *Journal of Business and Economic Statistics*, 24(2), 192–194.

Gopinath, S. and C. Krishnamurti (2001), "Number of transactions and volatility: An empirical study using high-frequency data from Nasdaq stocks," *The Journal of Financial Research*, 24(2), 205-18.

Hansen PR and Lunde A. (2006), "Consistent ranking of volatility models," *Journal of Econometrics*, 131, 97–121.

Harris, L. (1987), "Transaction data tests of the mixture of distributions hypothesis," *Journal of Financial and Quantitative Analysis*, 22, 127-141.

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

Huang, X. and G. Tauchen (2005), "The relative contribution of jumps to total price variance," *Journal of Financial Econometrics*, 3(4), 456-499.

Huimin Chung, Chin-Sheng Huang, Tseng-Chan Tseng (2008), "Modeling and forecasting of realized volatility based on high-frequency data: Evidence from Taiwan," *International Research Journal of Finance and Economics*, ISSN 1450-2887 Issue 22.

Jennings, R. H. and Christopher B. (1983), "Information dissemination and portfolio choice," *Journal of Financial and Quantitative Analysis*, 18, 1-19.

Jones, C. M., G. Kual, and M. L. Lipson (1994), "Transactions, volume and volatility," *Review of Financial Studies*, 7, 631-51.

Karpoff, J.M. (1987), "The relationship between price changes and trading volume: A survey," *Journal of Financial and Quantitative Analysis*, 22, 109–126.

Lee, S. and P. Mykland (2007), "Jumps in financial markets: A new nonparametric test and jump clustering," *Review of Financial Studies*. Forthcoming.

Lamoureux, C.G. and Lastrapes, W.D. (1990), "Heteroskedasticity in stock return data: Volume versus GARCH effects," *Journal of Finance*, 45, 221-229.

Lynch, P., and G. Zumbach (2003), "Market heterogeneities and the causal structure of volatility," *Quantitative Finance*, 3, 320-331.

Mancini, C. and R. Reno (2006), "Threshold estimation of jump-diffusion models and interest rate modeling," Manuscript, University of Florence and University of Siena.

Manganelli, S. (2005), "Duration, volume, and the price impact of trades," *Journal of Financial Markets*, 8, 377-399.

Meddahi, N. (2002), "A theoretical comparison between integrated and realized Volatility," *Journal of Applied Econometrics*, 17, 479-508.

Morse, D. (1981), "Price and trading volume reaction surrounding earnings announcements: A closer examination," *Journal of Accounting Research*, 19, 374-383.

Miller, U., M. Dacorogna, R. Dav, 0. Pictet, R. Olsen, and J. Ward (1993), "Fractals and intrinsic time - A challenge to econometricians," 39th International AEA Conference on Real Time Econometrics, 14-15 October 1993, Luxembourg.

Muller, U., M. Dacorogna, R. Dav, R. Olsen, 0. Pictet, and J. von Weizsacker (1997), "Volatilities of different time resolutions - Analyzing the dynamics of market components," *Journal of Empirical Finance*, 4, 213-239.

Patton A. (2006), "Volatility forecast comparison using imperfect volatility proxies," Working paper 175, Quantitative Finance Research Centre, University of Technology Sydney.

Tauchen G.E. and M. Pitts (1983), "The price variability-volume relationship on solution markets," *Econometrica*, 51, 485-505.

Taylor S. (1986), “Modeling financial time series,” New York: John Wiley.

Tseng, Tseng-Chan (2009), “Predicting realized volatility based on high frequency data in stocks markets,” Unpublished doctoral dissertation, University of Yunlin.

Wu, G. (2001), “The determinants of asymmetric volatility,” *Review of Financial Studies*, 14, 837-859.

Zakoian, J.M. (1994), “Threshold Heteroskedastic Models,” *Journal of Economic Dynamics and Control*, 18, 931-955.

Table 1 Descriptive Statistics of Realized Volatility Levels in the TAIEX

Variables[a]	Mean	Std. Dev.	Skew	Kurtosis	Min.	Max.	$LB_{(10)}$[d]
RV_t	1.506	2.451	9.504	137.429	0.080	46.846	574.630
BPV_t	0.896	0.870	3.182	18.167	0.063	8.199	2517.100
$CTBPV_t$	0.781	0.757	3.224	19.540	0.019	7.614	2563.300
RPV_t	0.853	0.354	1.296	5.706	0.256	2.826	4821.600
TC_t	0.862	0.874	3.231	18.575	0.019	8.902	2273.600
TJ_t[b]	0.643	2.234	11.926	197.696	0.000	46.826	83.266
J_t[c]	0.470	2.060	13.762	253.528	0.000	46.735	52.715
$RV_t^{1/2}$	1.084	0.574	2.901	19.976	0.284	6.844	2034.1
$BPV_t^{1/2}$	0.872	0.368	1.422	6.091	0.252	2.863	3935.6
$CTBPV_t^{1/2}$	0.813	0.346	1.369	5.999	0.138	2.759	4006.4
$RPV_t^{1/2}$	0.905	0.181	0.666	3.515	0.506	1.681	5377.9
$TC_t^{1/2}$	0.850	0.374	1.446	6.152	0.138	2.983	3590.8
$TJ_t^{1/2}$[b]	0.434	0.674	2.697	16.490	0.000	6.843	87.155
$J_t^{1/2}$[c]	0.306	0.613	3.421	23.117	0.000	6.836	57.847
$\ln RV_t$	-0.047	*0.886*	*0.419*	*3.397*	*-2.515*	*3.846*	*3045.5*
$\ln BPV_t$	*-0.432*	*0.786*	*0.220*	*2.793*	*-2.750*	*2.104*	*4707.6*
$\ln CTBPV_t$	*-0.575*	*0.804*	*0.091*	*2.978*	*-3.950*	*2.030*	*4751.1*
$\ln RPV_t$	*-0.236*	*0.393*	*0.145*	*2.661*	*-1.361*	*1.039*	*5650.2*
$\ln TC_t$	*-0.497*	*0.825*	*0.110*	*3.054*	*-3.950*	*2.186*	*4352.2*
$\ln(TJ_t+1)$[b]	*0.295*	*0.494*	*2.437*	*10.984*	*0.000*	*3.867*	*128.66*
$\ln(J_t+1)$[c]	0.209	0.442	3.080	15.781	0.000	3.865	89.298

Notes:

a. RV_t denotes realized variance; BPV_t denotes bi-power variance; $CTBPV_t$ denotes corrected threshold bi-power variance; RPV_t denotes realized power variance; C_t and J_t are the respective continuous and jump elements of RV_t as separated by the $CTBPV_t$ jump test of Corsi et al. (2009).

b. A significant daily jump test are computed using a critical value of $\alpha = 99.9\%$ while the C-Tz statistics are computed using $c_\theta = 3$.

c. A significant level of $\alpha = 99.9\%$ was used in the bi-power jump test.

d. The critical value of the test statistic for $LB_{(10)}$ was 18.307.

Table 2 Descriptive Statistics of Daily Trading Volume for the TAIEX

variables[a]	Mean	Std. Dev.	Skew	Kurtosis	Min.	Max.	$LB_{(10)}$
$lnVOL_t$	394.986	23.038	-0.091	2.814	320.204	463.193	7190.0
$lnTNV_t$	504.578	19.223	0.048	2.662	433.893	562.7526	7154.7
$lnTSV_t$	-109.592	6.107	-0.676	3.853	-130.042	-92.593	7102.6
$lnTNR_t$	1.004	0.013	0.211	3.605	0.938	1.054	245.3
$lnTVR_t$	0.999	0.007	-0.204	3.773	0.965	1.030	19.916

Notes:

a. VOL_t denotes daily trading volume value (unit: NTD million); TNV_t denotes daily number of transactions (trading frequency); TSV_t denotes daily average trade size; TNR_t denotes daily bid-ask frequency; TVR_t denotes daily bid-ask volume.

b. All measures of volume are transformed by taking their logarithm.

c. The critical value of the test statistic for $LB_{(10)}$ was 18.307.

Table 3 Pearson Correlation Matrix Between Measures of Daily Trading Volume

	VOL_t	TNV_t	TSV_t	TNR_t	TVR_t
VOL_t	1.000	0.974	0.705	-0.098	0.033
TNV_t		1.000	0.527	-0.089	0.026
TSV_t			1.000	-0.091	0.042
TNR_t				1.000	0.329
TVR_t					1.000000

Note:

a. This table reports correlation estimates of daily trading volume value (VOL_t), daily number of transactions (TNV_t), daily average trade size (TSV_t), daily bid-ask frequency (TNR_t), and daily bid-ask volume (TVR_t).

Table 4 Coefficient Estimates and Significance of Jump Terms in HAR-RV Regressions (H=1)

HAR-	RV_{t+1}			$RV_{t+1}^{1/2}$			$\ln RV_{t+1}$		
	RV	CJ	TCJ	RV	CJ	TCJ	RV	CJ	TCJ
β_0	0.429***	0.247***	0.141*	0.158***	0.103***	0.076**	-0.008	0.097***	0.149***
	(5.089)	(3.027)	(1.769)	(3.875)	(2.620)	(2.037)	(0.435)	(2.657)	(3.193)
β_d	0.159	-0.078	0.100	0.068	0.029	0.139	0.028	0.072	0.126**
	(1.440)	(-0.522)	(0.552)	(1.014)	(0.349)	(1.513)	(0.711)	(1.391)	(2.154)
β_w	0.330**	0.859***	0.990***	0.530***	0.660***	0.685***	0.520***	0.459***	0.428***
	(2.303)	(2.977)	(2.726)	(5.636)	(4.950)	(4.469)	(7.723)	(5.818)	(5.138)
β_m	0.219***	0.296*	0.293	0.252***	0.252***	0.194*	0.350***	0.299***	0.228***
	(2.963)	(1.668)	(1.405)	(4.238)	(2.754)	(1.850)	(5.716)	(4.294)	(3.114)
β_{jd}		0.205	0.168		0.031	0.022		-0.005	-0.006
		(1.579)	(1.365)		(0.665)	(0.551)		(-0.086)	(-0.134)
β_{jw}		0.196**	0.150*		0.265***	0.234***		0.512***	0.488***
		(2.002)	(1.647)		(4.039)	(4.456)		(4.188)	(4.499)
β_{jm}		-0.108	-0.055		0.045	0.068**		0.083	0.160
		(-0.668)	(-0.548)		(0.413)	(0.854)		(0.459)	(1.091)
R^2	0.168	0.188	0.194	0.345	0.352	0.363	0.432	0.439	0.445
RMSE	2.228	2.198	2.190	2.235	2.234	2.229	2.271	2.277	2.275
J-R^2	0.184	0.216	0.244	0.223	0.397	0.595	0.201	0.217	0.226
J-RMSE	1.699	1.664	1.634	1.747	1.723	1.710	1.863	1.826	1.831
C-R^2	0.051	0.090	0.162	0.140	0.329	0.578	0.236	0.243	0.243
C-RMSE	1.824	1.785	1.712	1.860	1.786	1.693	1.843	1.805	1.784

Notes:

a. OLS estimate for daily (H=1) HAR-RV, HAR-RV-CJ, HAR-RV-TCJ volatility forecast regressions.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 5 Coefficient Estimates and Significance of Jump Terms in HAR-RV Regressions (H=5)

HAR-	RV_{t+1}			$RV_{t+1}^{1/2}$			ln RV_{t+1}		
	RV	CJ	TCJ	RV	CJ	TCJ	RV	CJ	TCJ
β_0	0.563***	0.350***	0.217**	0.211***	0.149***	0.122***	-0.012	0.105**	0.157**
	(5.229)	(3.682)	(2.237)	(4.383)	(3.274)	(2.812)	(-0.544)	(2.355)	(2.807)
β_d	0.097***	0.163**	0.246***	0.112***	0.142***	0.216***	0.101***	0.119***	0.151***
	(3.741)	(2.042)	(2.606)	(4.608)	(3.230)	(4.648)	(4.569)	(4.055)	(4.860)
β_w	0.240**	0.217	0.706***	0.349***	0.411***	0.464***	0.316***	0.297***	0.285***
	(2.244)	(1.109)	(2.662)	(3.934)	(3.516)	(3.509)	(4.506)	(3.873)	(3.708)
β_m	0.278***	0.678***	0.475**	0.338***	0.359***	0.315***	0.449***	0.392***	0.324***
	(3.456)	(3.506)	(2.413)	(4.907)	(3.789)	(2.847)	(6.804)	(5.415)	(4.237)
β_{jd}		0.098***	0.097***		0.053**	0.064***		0.083**	0.105***
		(3.137)	(3.847)		(2.560)	(3.785)		(2.570)	(3.735)
β_{jw}		0.093	0.049		0.172***	0.110*		0.326***	0.245**
		(1.455)	(0.878)		(2.641)	(1.816)		(2.750)	(2.176)
β_{jm}		-0.081	-0.064		0.047	0.082		0.094	0.220
		(-0.568)	(-0.743)		(0.474)	(1.097)		(0.508)	(1.486)
R^2	0.263	0.320	0.351	0.528	0.554	0.573	0.640	0.653	0.656
RMSE	1.356	1.301	1.271	1.329	1.308	1.293	1.367	1.360	1.360
J-R^2	0.214	0.223	0.236	0.229	0.274	0.288	0.489	0.505	0.507
J-RMSE	0.900	0.894	0.886	1.019	1.005	1.004	0.983	0.976	0.981
C-R^2	0.148	0.246	0.299	0.213	0.323	0.393	0.491	0.501	0.501
C-RMSE	0.830	0.781	0.753	0.936	0.910	0.890	0.820	0.810	0.806

Notes:

a. OLS estimate for weekly (H=5) HAR-RV, HAR-RV-CJ, HAR-RV-TCJ volatility forecast regressions.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 6 Coefficient Estimates and Significance of Jump Terms in HAR-RV Regressions (H=20)

HAR-	RV_{t+1}			$RV_{t+1}^{1/2}$			$\ln RV_{t+1}$		
	RV	CJ	TCJ	RV	CJ	TCJ	RV	CJ	TCJ
β_0	0.724***	0.581***	0.435***	0.323***	0.267***	0.244***	-0.019	0.095*	0.125**
	(7.108)	(6.308)	(4.480)	(6.457)	(5.607)	(5.362)	(-0.741)	(1.942)	(2.058)
β_d	0.044**	0.107**	0.161***	0.063***	0.096***	0.146***	0.058***	0.077***	0.097***
	(2.133)	(2.377)	(3.015)	(2.880)	(2.826)	(3.793)	(3.350)	(3.342)	(4.027)
β_w	0.081	0.179	0.567***	0.183**	0.296***	0.395***	0.233***	0.271***	0.293***
	(0.987)	(1.219)	(3.367)	(2.165)	(2.961)	(3.754)	(3.458)	(3.685)	(4.016)
β_m	0.380***	0.546***	0.447**	0.448***	0.407***	0.318***	0.493***	0.385***	0.291***
	(5.081)	(3.021)	(2.185)	(5.661)	(3.758)	(2.696)	(6.854)	(4.626)	(3.355)
β_{jd}		0.044***	0.045***		0.034***	0.039***		0.055**	0.065***
		(3.539)	(4.582)		(2.765)	(3.645)		(2.445)	(3.190)
β_{jw}		-0.065	-0.099**		0.021	-0.035		0.073	0.014
		(-1.209)	(-2.345)		(0.333)	(-0.598)		(0.608)	(0.129)
β_{jm}		0.123	0.124		0.172	0.235***		0.293	0.454***
		(0.838)	(1.184)		(1.603)	(2.695)		(1.431)	(2.624)
R^2	0.268	0.333	0.389	0.513	0.549	0.571	0.649	0.654	0.657
RMSE	0.995	0.948	0.908	0.984	0.959	0.941	1.024	1.021	1.018
J-R^2	0.236	0.274	0.303	0.291	0.318	0.321	0.585	0.595	0.599
J-RMSE	0.585	0.569	0.558	0.739	0.735	0.733	0.665	0.666	0.667
C-R^2	0.210	0.306	0.359	0.348	0.447	0.476	0.593	0.593	0.595
C-RMSE	0.538	0.503	0.484	0.688	0.674	0.666	0.530	0.524	0.522

Notes:

a. OLS estimate for monthly (H=20) HAR-RV, HAR-RV-CJ, HAR-RV-TCJ volatility forecast regressions.

b. The significant daily jumps are computed using a critical value of $\alpha = 99.9\%$.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 7 In-sample Results modeling $RV_{t,t+H}$ of TAIEX using the LHAR-RV-TCJ Model

	Horizons				
Variables	1day	1week	2weeks	3weeks	4weeks
β_0	0.077	0.144	0.233**	0.309***	0.384***
	(0.857)	(1.401)	(2.347)	(3.280)	(4.072)
β_{cd}	-0.052	0.126	0.045	0.104	0.111**
	(-0.315)	(1.317)	(0.622)	(1.639)	(2.023)
β_{cw}	1.041**	0.723**	0.870***	0.762***	0.666***
	(2.253)	(2.389)	(3.336)	(3.405)	(3.569)
β_{cm}	0.307	0.595***	0.513**	0.485**	0.474**
	(1.357)	(3.070)	(2.300)	(2.186)	(2.243)
β_{jd}	0.122	0.071***	0.039**	0.031***	0.028***
	(1.015)	(2.660)	(2.527)	(2.849)	(2.913)
β_{jw}	0.159*	0.054	-0.012	-0.048	-0.079*
	(1.833)	(0.848)	(-0.220)	(-0.996)	(-1.823)
β_{jm}	-0.026	-0.038	-0.013	0.046	0.135
	(-0.276)	(-0.460)	(-0.134)	(0.471)	(1.328)
β_{rd}	-0.342***	-0.161**	-0.165***	-0.174***	-0.134***
	(-2.791)	(-2.375)	(-3.253)	(-3.925)	(-3.587)
β_{rw}	-0.140	-0.410	-0.223	-0.067	-0.002
	(-0.515)	(-1.241)	(-0.956)	(-0.320)	(-0.014)
β_{rm}	0.338	1.083	1.027	0.982*	0.911*
	(0.468)	(1.317)	(1.515)	(1.784)	(1.938)
R^2	0.203	0.366	0.408	0.416	0.401
RMSE	*2.175*	*1.254*	*1.040*	*0.945*	*0.898*

Notes:

a. OLS estimate for LHAR-TCJ regressions, estimated on 1day and 1- to 4 -week of the realized volatility.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 8 In-sample Results modeling $RV_{t,t+H}^{1/2}$ of TAIEX using the LHAR-RV-TCJ Model

	Horizons				
Variables	1day	1week	2weeks	3weeks	4weeks
β_0	0.100**	0.113**	0.149***	0.181***	0.215***
	(2.236)	(2.169)	(2.907)	(3.712)	(4.475)
β_{cd}	-0.026	0.106**	0.071*	0.104***	0.112***
	(-0.270)	(2.320)	(1.762)	(2.743)	(3.085)
β_{cw}	0.683***	0.446***	0.544***	0.501***	0.462***
	(3.708)	(3.352)	(3.925)	(3.964)	(4.130)
β_{cm}	0.269**	0.402***	0.354***	0.333***	0.320***
	(2.491)	(3.768)	(2.953)	(2.719)	(2.700)
β_{jd}	-0.022	0.036**	0.025**	0.027**	0.028***
	(-0.686)	(2.529)	(2.029)	(2.538)	(2.799)
β_{jw}	0.241***	0.129*	0.070	0.025	-0.007
	(4.310)	(1.945)	(1.021)	(0.393)	(-0.129)
β_{jm}	0.093	0.097	0.130*	0.178**	0.234***
	(1.280)	(1.356)	(1.669)	(2.279)	(2.716)
β_{rd}	-0.100***	-0.051***	-0.046***	-0.043***	-0.032***
	(-2.795)	(-3.454)	(-4.352)	(-4.541)	(-3.730)
β_{rw}	-0.114**	-0.122**	-0.050	-0.010	0.012
	(-2.340)	(2.083)	(-1.069)	(-0.220)	(0.292)
β_{rm}	0.076	0.246	0.239*	0.237**	0.213*
	(0.581)	(1.638)	(1.779)	(1.989)	(1.929)
R^2	0.388	0.592	0.605	0.599	0.579
RMSE	2.196	1.275	1.070	0.977	0.932

Notes:

a. OLS estimate for LHAR-TCJ regressions, estimated on 1day and 1- to 4 -week of the realized volatility.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 9 In-sample Results modeling $\ln RV_{t,t+H}$ of TAIEX using the LHAR-RV-TCJ Model

	Horizons				
Variables	1day	1week	2weeks	3weeks	4weeks
β_0	0.044	0.096*	0.134**	0.139**	0.127**
	(0.978)	(1.787)	(2.277)	(2.333)	(2.069)
β_{cd}	0.035	0.085***	0.062***	0.068***	0.077***
	(0.594)	(2.970)	(2.713)	(3.063)	(3.454)
β_{cw}	0.407***	0.273***	0.316***	0.329***	0.324***
	(4.670)	(3.681)	(4.013)	(4.283)	(4.379)
β_{cm}	0.285***	0.378***	0.356***	0.322***	0.293***
	(4.018)	(5.179)	(4.475)	(3.818)	(3.393)
β_{jd}	-0.076*	0.056**	0.040*	0.043**	0.045**
	(-1.648)	(2.226)	(1.789)	(2.111)	(2.359)
β_{jw}	0.456***	0.251**	0.188	0.120	0.063
	(4.185)	(2.103)	(1.526)	(1.014)	(0.554)
β_{jm}	0.234*	0.282**	0.311**	0.380***	0.464***
	(1.781)	(2.015)	(2.116)	(2.480)	(2.688)
β_{rd}	-0.116***	-0.070***	-0.061***	-0.056***	-0.042***
	(-3.431)	(-3.820)	(-4.495)	(-4.551)	(-3.567)
β_{rw}	-0.219***	-0.206***	-0.097	-0.038	0.001
	(-4.263)	(-3.168)	(-1.633)	(-0.614)	(0.026)
β_{rm}	0.097	0.317*	0.317*	0.323**	0.286*
	(0.692)	(1.787)	(1.836)	(1.993)	(1.811)
R^2	0.466	0.674	0.687	0.681	0.662
RMSE	2.240	1.341	1.147	1.058	1.013

Notes:

a. OLS estimate for LHAR-TCJ regressions, estimated on 1day and 1- to 4 -week of the realized volatility.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 10 Comparing Forecasting Models: HAR-RV, HAR-RV-TCJ and LHAR-RV-TCJ

	Horizons									
	1day		1week		2weeks		3weeks		4weeks	
Models	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Panel A:										
HAR-RV	0.168	2.228	0.263	1.356	0.261	1.165	0.263	1.064	0.268	0.995
HAR-TCJ	0.194	2.190	0.351	1.271	0.393	1.055	0.400	0.958	0.389	0.908
LHAR-TCJ	**0.203**	**2.175**	**0.366**	**1.254**	**0.408**	**1.040**	**0.416**	**0.945**	**0.401**	**0.898**
Panel B:										
HAR-$RV^{1/2}$	0.345	2.235	0.528	1.329	0.532	1.136	0.525	1.042	0.513	0.984
HAR-$TCJ^{1/2}$	0.363	2.229	0.573	1.293	0.592	1.084	0.589	0.990	0.571	0.941
LHAR-$TCJ^{1/2}$	**0.388**	**2.196**	**0.592**	**1.275**	**0.605**	**1.070**	**0.599**	**0.977**	**0.579**	**0.932**
Panel C:										
HAR-lnRV	0.432	2.271	0.640	1.367	0.662	1.169	0.662	1.077	0.649	1.024
HAR-lnTCJ	0.445	2.275	0.656	1.360	0.677	1.159	0.674	1.067	0.657	1.018
LHAR-lnTCJ	**0.466**	**2.240**	**0.674**	**1.324**	**0.687**	**1.147**	**0.681**	**1.058**	**0.662**	**1.013**

Notes:

a. The table presents the adjusted R^2 and root mean square error (RMSE) for 1-day and 1- to 4-week in-sample predictions for TAIEX.

b. The dependent variable for all models and for all horizons is the standardized realized variance: $RV_{t;t+H}/H$.

c. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 11 In-sample daily regressions for the LHAR-RV-CJ$^+$ and HAR-RV-CJ$^+$ model

	RV_{t+1}		$RV_{t+1}^{1/2}$		$\ln RV_{t+1}$	
	HAR-CJ$^+$	LHAR-CJ$^+$	HAR-CJ$^+$	LHAR-CJ$^+$	HAR-CJ$^+$	LHAR-CJ$^+$
β_0	0.158*	0.116	0.091**	0.100**	0.146***	0.055
	(1.957)	(1.283)	(2.460)	(2.242)	(3.126)	(1.200)
β_d	0.036	-0.036	0.105	-0.012	0.107*	0.039
	(0.190)	(-0.215)	(1.103)	(-0.131)	(1.859)	(0.694)
β_w	1.083***	1.095**	0.718***	0.689***	0.449***	0.411***
	(2.811)	(2.277)	(4.538)	(3.679)	(5.250)	(4.646)
β_m	0.278	0.309	0.176*	0.260**	0.218***	0.281***
	(1.317)	(1.358)	(1.660)	(2.363)	(2.993)	(3.963)
$\beta_{jd}^{(+)}$	-0.066	-0.055	-0.066**	-0.053*	-0.152***	-0.117**
	(-1.227)	(-1.093)	(-2.093)	(-1.651)	(-2.747)	(-2.103)
$\beta_{jd}^{(-)}$	0.269**	0.231	0.112*	0.026	0.155**	-0.010
	(2.231)	(1.618)	(1.898)	(0.450)	(2.105)	(-0.143)
β_{jw}	0.145	0.151	0.232***	0.239	0.485***	0.456***
	(1.361)	(1.455)	(4.360)	(4.305)	(4.612)	(4.223)
β_{jm}	-0.011	-0.001	0.090	0.099	0.189	0.239*
	(-0.116)	(-0.017)	(1.194)	(1.384)	(1.318)	(1.808)
$\beta_{rd}^{(-)}$		-0.163		-0.076***		-0.094***
		(-1.278)		(-2.793)		(-2.999)
$\beta_{rw}^{(-)}$		-0.129		-0.111**		-0.217***
		(-0.464)		(-2.220)		(-4.188)
$\beta_{rm}^{(-)}$		0.363		0.086		0.107
		(0.493)		(0.635)		(0.750)
R^2	0.215	0.216	0.378	0.389	0.455	0.467
RMSE	2.160	2.157	2.204	2.191	2.260	2.238

Notes:

a. OLS estimate for the LHAR-RV-CJ$^+$ and HAR-RV-CJ$^+$ model in which this separates daily jumps into positive and negative value.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 12 In-sample weekly regressions for the LHAR-RV-CJ+ and HAR-RV-CJ+ model

	RV_{t+1}		$RV_{t+1}^{1/2}$		ln RV_{t+1}	
	HAR-CJ+	LHAR-CJ+	HAR-CJ+	LHAR-CJ+	HAR-CJ+	LHAR-CJ+
β_0	0.221**	0.149	0.128***	0.113**	0.155***	0.095*
	(2.272)	(1.440)	(2.940)	(2.169)	(2.792)	(1.760)
β_d	0.228**	0.128	0.203***	0.107**	0.143***	0.084***
	(2.421)	(1.339)	(4.405)	(2.363)	(4.719)	(2.966)
β_w	0.731***	0.731**	0.478***	0.466***	0.295***	0.273***
	(2.748)	(2.407)	(3.610)	(3.350)	(3.821)	(3.673)
β_m	0.471**	0.595***	0.308***	0.401***	0.319***	0.379***
	(2.400)	(3.071)	(2.778)	(3.753)	(4.175)	(5.182)
$\beta_{jd}^{(+)}$	0.034	0.044	0.028**	0.033**	0.040	0.060*
	(1.191)	(1.377)	(1.866)	(2.057)	(1.406)	(1.948)
$\beta_{jd}^{(-)}$	0.125***	0.087***	0.100***	0.040*	0.177***	0.050
	(6.851)	(2.808)	(4.864)	(1.718)	(4.513)	(1.221)
β_{jw}	0.048	0.053	0.109*	0.128*	0.244**	0.251**
	(0.838)	(0.811)	(1.808)	(1.939)	(2.194)	(2.101)
β_{jm}	-0.053	-0.035	0.091	0.098	0.233	0.282**
	(-0.611)	(-0.415)	(1.231)	(1.361)	(1.586)	(2.013)
$\beta_{rd}^{(-)}$		-0.134		-0.049***		-0.072***
		(-1.607)		(-2.765)		(-3.328)
$\beta_{rw}^{(-)}$		-0.408		-0.121**		-0.206***
		(-1.229)		(-2.082)		(-3.175)
$\beta_{rm}^{(-)}$		1.087		0.247		0.316*
		(1.321)		(1.637)		(1.776)
R^2	0.354	0.367	0.577	0.592	0.659	0.674
RMSE	1.267	1.253	1.288	1.274	1.356	1.341

Notes:

a. OLS estimate for the LHAR-RV-CJ+ and HAR-RV-CJ+ model in which this separates daily jumps into positive and negative values.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 13 In-sample monthly regressions for the LHAR-RV-CJ^+ and HAR-RV-CJ^+ model

	RV_{t+1}		$RV_{t+1}^{1/2}$		ln RV_{t+1}	
	HAR-CJ^+	LHAR-CJ^+	HAR-CJ^+	LHAR-CJ^+	HAR-CJ^+	LHAR-CJ^+
β_0	0.437***	0.386***	0.245***	0.215***	0.125**	0.126**
	(4.499)	(4.106)	(5.400)	(4.474)	(2.053)	(2.034)
β_d	0.153***	0.111**	0.143***	0.111***	0.095***	0.076***
	(2.896)	(2.026)	(3.744)	(2.996)	(3.974)	(3.405)
β_w	0.577***	0.668***	0.398***	0.462***	0.295***	0.323***
	(3.412)	(3.573)	(3.763)	(4.127)	(4.022)	(4.371)
β_m	0.445**	0.474**	0.316***	0.321***	0.290***	0.293***
	(2.180)	(2.243)	(2.682)	(2.710)	(3.337)	(3.397)
$\beta_{jd}^{(+)}$	0.018	0.020	0.032**	0.031**	0.050**	0.049**
	(0.995)	(0.998)	(2.474)	(2.334)	(2.139)	(1.989)
$\beta_{jd}^{(-)}$	0.057***	0.033**	0.047***	0.024	0.082***	0.038
	(4.468)	(2.334)	(3.238)	(1.403)	(2.751)	(1.137)
β_{jw}	-0.099**	-0.080*	-0.035	-0.007	0.014	0.063
	(-2.340)	(-1.817)	(-0.599)	(-0.125)	(0.126)	(0.553)
β_{jm}	0.129	0.137	0.237***	0.233***	0.457***	0.463***
	(1.240)	(1.342)	(2.718)	(2.716)	(2.640)	(2.688)
$\beta_{rd}^{(-)}$		-0.125***		-0.034***		-0.045***
		(-3.106)		(-3.157)		(-2.952)
$\beta_{rw}^{(-)}$		-0.002		0.012		0.001
		(-0.012)		(0.285)		(0.022)
$\beta_{rm}^{(-)}$		0.912*		0.212*		0.285*
		(1.942)		(1.923)		(1.805)
R^2	0.389	0.400	0.571	0.578	0.657	0.662
RMSE	0.907	0.898	0.940	0.932	1.018	1.013

Notes:

a. OLS estimate for the LHAR-RV-CJ^+ and HAR-RV-CJ^+ model in which this separates daily jumps into positive and negative values.

b. The significant daily jumps are computed using a critical value of α = 99.9%.

c. The t-statistics based on Newey-West correction are given in parentheses.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 14 In-sample Results modeling $RV_{t,t+H}$ of TAIEX using the LHAR-TCJ-cum-Vol Model

	Horizons				
Variables	1day	1week	2weeks	3weeks	4weeks
β_0	-1.347**	-1.822***	-2.223***	-2.485***	-2.529***
	(-2.483)	(-2.619)	(-3.139)	(-3.643)	(-3.887)
β_{cd}	-0.064	0.077	0.024	0.056	0.058
	(-0.416)	(0.867)	(0.405)	(0.994)	(1.256)
β_{cw}	0.937**	0.637**	0.626***	0.436**	0.336**
	(1.982)	(2.061)	(2.596)	(2.264)	(1.987)
β_{cm}	0.435	0.748***	0.806***	0.898***	0.896***
	(1.642)	(3.817)	(3.159)	(3.824)	(4.051)
β_{jd}	0.121	0.067**	0.038**	0.030***	0.027***
	(0.986)	(2.316)	(2.456)	(2.653)	(2.712)
β_{jw}	0.156*	0.048	-0.017	-0.051	-0.083*
	(1.780)	(0.769)	(-0.327)	(-1.172)	(-1.922)
β_{jm}	-0.124	-0.172*	-0.195*	-0.176	-0.095
	(-1.349)	(-1.828)	(-1.682)	(-1.429)	(-0.731)
β_{rd}	-0.329***	-0.158**	-0.130***	-0.130***	-0.090***
	(-2.689)	(-2.445)	(-2.625)	(-3.231)	(-2.687)
β_{rw}	-0.122	-0.391	-0.241	-0.159	-0.095
	(-0.342)	(-1.458)	(-1.219)	(-0.874)	(-0.565)
β_{rm}	0.162	0.891	0.570	0.292	0.209
	(0.201)	(1.040)	(0.827)	(0.531)	(0.439)
β_{vd}	0.002	0.008***	0.002	0.004	0.005**
	(0.231)	(2.663)	(0.817)	(1.508)	(2.035)
β_{vw}	-0.005	-0.014	-9.38E-4	0.004	0.003
	(-0.449)	(-1.265)	(-0.101)	(0.604)	(0.516)
β_{vm}	0.012	0.019	0.015	0.010	0.011
	(0.917)	(1.452)	(1.294)	(1.105)	(1.462)
RMSE	2.167	1.225	0.994	0.879	0.822

Notes:

a. OLS estimate for LHAR-RV-CJ-cum-Vol regressions.

b. The t-statistics based on Newey-West correction are given in parentheses.

c. The significant daily jumps are computed using a critical value of α = 99.9%.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 15 In-sample Results modeling $RV_{t,t+H}^{1/2}$ of TAIEX using the LHAR-TCJ-cum-Vol Model

	Horizons				
Variables	1day	1week	2weeks	3weeks	4weeks
β_0	-0.223*	-0.307*	-0.401**	-0.464***	-0.481***
	(-1.769)	(-1.946)	(-2.350)	(-2.704)	(-2.823)
β_{cd}	-0.080	0.068	0.054	0.070**	0.068**
	(-0.946)	(1.534)	(1.457)	(2.006)	(2.083)
β_{cw}	0.655***	0.420***	0.390***	0.292**	0.245**
	(3.342)	(2.918)	(2.891)	(2.429)	(2.235)
β_{cm}	0.360***	0.496***	0.540***	0.597***	0.602***
	(2.685)	(4.360)	(4.123)	(4.669)	(4.963)
β_{jd}	-0.032	0.028*	0.023*	0.024**	0.023**
	(-0.838)	(1.909)	(1.874)	(2.331)	(2.432)
β_{jw}	0.232***	0.115*	0.049	-1.81E-4	-0.034
	(4.015)	(1.795)	(0.753)	(-0.003)	(-0.617)
β_{jm}	0.021	0.009	0.002	0.020	0.063
	(0.286)	(0.121)	(0.031)	(0.220)	(0.645)
β_{rd}	-0.102***	-0.052***	-0.038***	-0.033***	-0.023***
	(-2.940)	(-3.515)	(-3.711)	(-3.632)	(-2.777)
β_{rw}	-0.123*	-0.120**	-0.059	-0.035	-0.015
	(-1.845)	(-2.292)	(-1.338)	(-0.809)	(-0.372)
β_{rm}	0.021	0.194	0.119	0.062	0.026
	(0.149)	(1.248)	(0.856)	(0.506)	(0.223)
β_{vd}	0.002	0.001**	5.5E-4	9.4E-4	0.001*
	(1.087)	(2.024)	(0.694)	(1.186)	(1.854)
β_{vw}	-0.002	-0.002	4.4E-4	0.001	0.001
	(-0.911)	(-1.245)	(0.220)	(0.965)	(0.860)
β_{vm}	0.002	0.003	0.003	0.001	0.002
	(0.887)	(1.585)	(1.259)	(0.962)	(1.183)
RMSE	2.191	1.254	1.037	0.931	0.878

Notes:

a. OLS estimate for LHAR-RV-CJ-cum-Vol regressions.

b. The t-statistics based on Newey-West correction are given in parentheses.

c. The significant daily jumps are computed using a critical value of $\alpha = 99.9\%$.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 16 In-sample Results modeling $\ln RV_{t,t+H}$ of TAIEX using the LHAR-TCJ-cum-Vol Model

	Horizons				
Variables	1day	1week	2weeks	3weeks	4weeks
β_0	-0.558***	-0.632***	-0.767***	-0.870***	-0.952***
	(-3.222)	(-3.106)	(-3.469)	(-3.840)	(-4.085)
β_{cd}	-0.017	0.050*	0.039*	0.036*	0.0.38*
	(-0.289)	(1.719)	(1.711)	(1.697)	(1.774)
β_{cw}	0.368***	0.206***	0.180**	0.168**	0.162**
	(3.775)	(2.718)	(2.311)	(2.225)	(2.182)
β_{cm}	0.391***	0.497***	0.539***	0.542***	0.522***
	(4.738)	(6.425)	(6.521)	(6.272)	(6.015)
β_{jd}	-0.101**	0.041*	0.035	0.036*	0.035*
	(-2.066)	(1.686)	(1.641)	(1.926)	(1.945)
β_{jw}	0.418***	0.203*	0.118	0.039	-0.020
	(3.766)	(1.753)	(1.000)	(0.348)	(-0.190)
β_{jm}	0.066	0.083	0.051	0.083	0.148
	(0.513)	(0.595)	(0.332)	(0.484)	(0.773)
β_{rd}	-0.118***	-0.066***	-0.047***	-0.040***	-0.028**
	(-3.361)	(-3.562)	(-3.522)	(-3.411)	(-2.440)
β_{rw}	-0.243***	-0.217***	-0.122**	-0.075	-0.037
	(-3.923)	(-3.495)	(-2.190)	(-1.349)	(-0.698)
β_{rm}	-0.054	0.151	0.054	0.005	-0.044
	(-0.367)	(0.857)	(0.309)	(0.034)	(-0.267)
β_{vd}	0.004**	0.003**	0.001	0.001*	0.002**
	(2.036)	(2.258)	(1.261)	(1.796)	(2.389)
β_{vw}	-0.002	-0.001	0.003	0.004	0.003
	(-0.678)	(-0.458)	(1.114)	(1.483)	(1.264)
β_{vm}	0.002	0.003	0.002	0.001	0.001
	(0.782)	(1.211)	(0.726)	(0.519)	(0.723)
RMSE	2.231	1.319	1.114	1.013	0.964

Notes:

a. OLS estimate for LHAR-RV-CJ-cum-Vol regressions.

b. The t-statistics based on Newey-West correction are given in parentheses.

c. The significant daily jumps are computed using a critical value of $\alpha = 99.9\%$.

d. * Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level.

Table 17 Comparison for the In-sample Performance of LHAR-RV-TCJ and HAR-RV-TCJ-cum-Vol Models

	Horizons									
	1day		1week		2weeks		3weeks		4weeks	
Models	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Panel A: LHAR-RV										
TCJ	0.203	2.175	0.366	1.254	0.408	1.040	0.416	0.945	0.401	0.898
TCJ-cum-Vol	**0.207**	**2.167**	**0.394**	**1.225**	**0.459**	**0.994**	**0.493**	**0.879**	**0.496**	**0.822**
Panel B: LHAR-$RV^{1/2}$										
$TCJ^{1/2}$	0.388	2.196	0.592	1.275	0.605	1.070	0.599	0.977	0.579	0.932
$TCJ^{1/2}$-cum-Vol	**0.393**	**2.191**	**0.610**	**1.254**	**0.637**	**1.037**	**0.647**	**0.931**	**0.639**	**0.878**
Panel C: LHAR-lnRV										
lnTCJ	0.466	2.240	0.674	1.324	0.687	1.147	0.681	1.058	0.662	1.013
lnTCJ-cum-Vol	**0.475**	**2.231**	**0.694**	**1.319**	**0.720**	**1.114**	**0.727**	**1.013**	**0.717**	**0.964**

Notes:

a. The table presents the adjusted R^2 and root mean square error (RMSE) for 1-day and 1- to 4-week in-sample predictions for TAIEX.

b. The dependent variable for all models and for all horizons is the standardized realized variance: $RV_{t;t+H}/H$.

c. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 18 In-sample Results for Five Measures of Trading Volume using the HAR-RV-TCJ-cum-Vol Model

	Horizons									
	1 day		1 week		2 weeks		3 weeks		4 weeks	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Panel A : LHAR-RV-TCJ-cum-Vol										
VOL	0.207	2.167	0.394	1.225	0.459	0.994	0.493	0.879	0.496	0.822
TNV	**0.209**	**2.165**	**0.398**	**1.221**	**0.468**	**0.985**	**0.506**	**0.868**	**0.514**	**0.808**
TSV	0.203	2.172	0.373	1.246	0.417	1.031	0.430	0.933	0.416	0.885
TNR	**0.208**	**2.165**	**0.380**	**1.239**	0.426	1.024	0.432	0.930	0.418	0.884
TVR	0.205	2.171	0.376	1.243	**0.428**	**1.021**	**0.446**	**0.919**	**0.438**	**0.868**
Panel B: LHAR-$RV^{1/2}$-$TCJ^{1/2}$-cum-Vol										
VOL	**0.393**	0.443	0.610	0.264	0.637	0.234	0.647	0.220	0.639	0.215
TNV	**0.393**	**0.442**	**0.613**	**0.263**	**0.644**	**0.232**	**0.657**	**0.217**	**0.652**	**0.211**
TSV	0.389	0.444	0.597	0.268	0.610	0.243	0.607	0.232	0.589	0.229
TNR	**0.390**	**0.444**	0.598	0.268	0.613	0.242	0.606	0.232	0.586	0.230
TVR	0.389	**0.444**	**0.600**	**0.267**	**0.618**	**0.240**	**0.618**	**0.229**	**0.602**	**0.226**
Panel C : LHAR-lnRV-ln TCJ-cum-Vol										
VOL	0.475	0.634	0.694	0.378	0.720	0.339	0.727	0.324	0.717	0.323
TNV	**0.476**	**0.633**	**0.698**	**0.376**	**0.727**	**0.335**	**0.737**	**0.319**	**0.729**	**0.316**
TSV	0.467	0.639	0.677	0.388	0.691	0.357	0.687	0.347	0.670	0.349
TNR	0.466	0.640	0.676	0.389	0.689	0.358	0.683	0.349	0.664	0.352
TVR	**0.468**	**0.638**	**0.682**	**0.386**	**0.699**	**0.352**	**0.699**	**0.341**	**0.682**	**0.342**

Notes:

a. Entries to the table represent Adjusted R^2 and RMSE for one day, one week through four weeks in-sample predictions of the RV in the TAIEX with five measures of trading volume.

b. These different regressors are employed in use of the LHAR-RV-cum-Vol model.

c. The dependent variable for all horizons is the standardized realized.

d. *VOL* denotes trading volume value; *TNV* denotes number of transactions (trading frequency); *TSV* denotes average trade size; *TNR* denotes bid-ask frequency; *TVR* denotes bid-ask volume.

e. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 19 In-sample Forecasts Evaluation modeling $RV_{t,t+H}$ using Different HAR-RV Models

	Horizons									
	1 day		1 week		2 weeks		3 weeks		4 weeks	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
	Panel A : HAR-RV-X Model									
RV	0.168	2.228	0.263	1.356	0.261	1.165	0.263	1.064	0.268	0.995
BPV	0.161	2.237	0.327	1.295	0.375	1.071	0.387	0.970	0.378	0.917
CTBPV	0.163	2.235	0.325	1.297	0.378	1.069	0.388	0.970	0.379	0.916
RPV	0.175	2.218	**0.359**	**1.265**	**0.411**	**1.040**	**0.423**	**0.942**	**0.410**	**0.893**
C	0.145	2..259	0.290	1.330	0.331	1.109	0.341	1.006	0.337	0.947
CJ	**0.199**	**2.183**	0.313	1.308	0.342	1.098	0.350	0.998	0.343	0.941
TC	0.161	2.237	0.330	1.292	0.385	1.063	0.395	0.964	0.382	0.914
TCJ	0.194	2.190	0.351	1.271	0.393	1.055	0.400	0.958	0.389	0.908
	Panel B: LHAR-RV-X Model									
RV	0.186	2.206	0.294	1.326	0.290	1.141	0.289	1.044	0.284	0.983
BPV	0.185	2.202	0.351	1.271	0.394	1.054	0.405	0.955	0.391	0.906
CTBPV	0.183	2.205	0.347	1.274	0.397	1.052	0.406	0.954	0.393	0.905
RPV	0.194	2.190	**0.373**	**1.249**	**0.423**	**1.029**	**0.433**	**0.932**	**0.417**	**0.887**
C	0.176	2.215	0.325	1.295	0.359	1.084	0.365	0.986	0.352	0.935
CJ	**0.205**	**2.173**	0.332	1.287	0.361	1.081	0.366	0.984	0.353	0.933
TC	0.184	2.204	0.356	1.266	0.406	1.043	0.415	0.947	0.397	0.902
TCJ	0.203	2.175	0.366	1.254	0.408	1.040	0.416	0.945	0.401	0.898
	Panel C : LHAR-RV-X-cum-Vol Model									
RV	0.185	2.199	0.312	1.307	0.331	1.107	0.353	0.994	0.366	0.923
BPV	0.191	2.191	0.377	1.243	0.439	1.013	0.473	0.898	0.480	0.836
CTBPV	0.189	2.194	0.375	1.246	0.442	1.011	0.474	0.896	0.483	0.834
RPV	0.199	2.181	**0.396**	**1.224**	**0.459**	0.995	0.487	0.886	0.492	0.827
C	0.186	2.198	0.363	1.257	0.421	1.029	0.457	0.911	0.469	0.845
CJ	**0.216**	**2.155**	0.377	1.242	0.440	1.011	0.482	0.889	0.490	0.827
TC	0.191	2.192	0.385	1.236	0.452	1.001	0.484	0.888	0.488	0.830
TCJ	0.207	2.167	0.394	1.225	**0.459**	**0.994**	**0.493**	**0.879**	**0.496**	**0.822**

Notes:

a. Entries to the table represent Adjusted R^2 and RMSE for one-day, and one- to four-week in-sample predictions of the RV of TAIEX using the HAR-RV, LHAR-RV, and LHAR-RV-cum-Vol models.

b. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 20 In-sample Forecasts Evaluation modeling $RV_{t,t+H}^{1/2}$ using Different HAR-RV Models

	Horizons									
	1 day		1 week		2 weeks		3 weeks		4 weeks	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
	Panel A : HAR-$RV^{1/2}$-X Model									
$RV^{1/2}$	0.345	2.235	0.528	1.329	0.532	1.136	0.525	1.042	0.513	0.984
$BPV^{1/2}$	0.346	2.254	0.559	1.319	0.583	1.099	0.576	1.002	0.560	0.951
$CTBPV^{1/2}$	0.347	2.255	0.546	1.325	0.577	1.102	0.575	1.005	0.557	0.954
RPV	**0.365**	2.232	0.572	1.296	**0.593**	1.085	0.587	0.994	0.567	0.946
$C^{1/2}$	0.327	2.270	0.522	1.344	0.549	1.125	0.549	1.026	0.532	0.972
$CJ^{1/2}$	0.349	**2.229**	0.539	1.324	0.558	1.115	0.555	1.019	0.539	0.966
$TC^{1/2}$	0.341	2.259	0.543	1.325	0.574	1.100	0.572	1.003	0.552	0.955
$TCJ^{1/2}$	0.363	**2.229**	**0.573**	**1.293**	0.592	**1.084**	**0.589**	**0.990**	**0.571**	**0.941**
	Panel B: LHAR-$RV^{1/2}$-X Model									
$RV^{1/2}$	0.373	2.206	0.556	1.308	0.551	1.120	0.540	1.029	0.522	0.970
$BPV^{1/2}$	0.376	2.216	0.576	1.297	0.594	1.084	0.589	0.989	0.567	0.943
$CTBPV^{1/2}$	0.373	2.223	0.570	1.304	0.590	1.087	0.585	0.993	0.564	0.946
RPV	0.386	**2.196**	0.590	1.280	0.604	1.072	0.595	0.982	0.573	0.939
$C^{1/2}$	0.365	2.224	0.556	1.314	0.570	1.105	0.563	1.010	0.541	0.963
$CJ^{1/2}$	0.373	2.205	0.564	1.303	0.573	1.101	0.566	1.028	0.547	0.958
$TC^{1/2}$	0.370	2.223	0.569	1.301	0.589	1.084	0.584	0.989	0.560	0.946
$TCJ^{1/2}$	**0.388**	**2.196**	**0.592**	**1.275**	**0.605**	**1.070**	**0.599**	**0.977**	**0.579**	**0.932**
	Panel C : LHAR-$RV^{1/2}$-X-cum-Vol Model									
$RV^{1/2}$	0.375	2.204	0.567	1.295	0.572	1.095	0.579	0.994	0.571	0.934
$BPV^{1/2}$	0.387	2.204	0.603	1.268	0.634	1.044	**0.647**	0.935	0.640	0.881
$CTBPV^{1/2}$	0.384	2.211	0.599	1.274	0.634	1.045	**0.647**	0.936	**0.641**	0.880
RPV	0.392	2.195	0.608	1.258	0.633	1.042	0.639	0.941	0.630	0.888
$C^{1/2}$	0.380	2.208	0.591	1.278	0.622	1.055	0.636	0.945	0.631	0.889
$CJ^{1/2}$	0.384	2.196	0.595	1.272	0.624	1.052	0.639	0.941	0.633	0.885
$TC^{1/2}$	0.382	2.210	0.600	1.269	0.634	1.041	0.646	0.932	0.638	0.879
$TCJ^{1/2}$	**0.393**	**2.191**	**0.610**	**1.254**	**0.637**	**1.037**	**0.647**	**0.931**	0.639	**0.878**

Notes:

a. Entries to the table represent Adjusted R^2 and RMSE for one-day, and one- to four-week in-sample predictions of the RV of TAIEX using the HAR-RV, LHAR-RV, and LHAR-RV-cum-Vol models.

b. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 21 In-sample Forecasts Evaluation modeling $\ln RV_{t,t+H}$ using Different HAR-RV Models

	Horizons									
	1 day		1 week		2 weeks		3 weeks		4 weeks	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Panel A : HAR-ln RV-X Model										
ln RV	0.432	2.271	0.640	1.367	0.662	1.169	0.662	1.077	0.649	1.024
ln BPV	0.436	2.293	0.648	1.379	0.675	1.166	0.675	1.070	0.658	1.019
ln CTBPV	0.432	2.301	0.636	1.393	0.664	1.176	0.663	1.079	0.646	1.028
ln RPV	**0.449**	**2.274**	**0.660**	1.364	**0.683**	**1.158**	**0.680**	**1.066**	**0.664**	**1.017**
ln C	0.424	2.305	0.632	1.394	0.660	1.181	0.659	1.084	0.644	1.032
ln CJ	0.435	2.278	0.644	1.374	0.666	1.172	0.664	1.277	0.648	1.027
ln TC	0.426	2.306	0.629	1.396	0.658	1.179	0.657	1.082	0.639	1.032
ln TCJ	0.445	2.275	0.656	**1.360**	0.677	1.159	0.674	1.067	0.657	1.018
Panel B: LHAR-ln RV-X Model										
ln RV	0.459	2.228	0.667	1.342	0.677	1.154	0.673	1.064	0.657	1.017
ln BPV	0.460	2.254	0.669	1.355	0.686	1.153	0.682	**1.058**	0.662	1.015
ln CTBPV	0.455	2.266	0.658	1.368	0.674	1.162	0.669	1.070	0.649	1.024
ln RPV	**0.468**	**2.237**	**0.677**	1.345	**0.691**	**1.147**	**0.685**	**1.058**	**0.667**	**1.013**
ln C	0.452	2.259	0.659	1.366	0.674	1.164	0.668	1.073	0.648	1.026
ln CJ	0.456	2.245	0.665	1.354	0.677	1.160	0.672	1.070	0.653	1.027
ln TC	0.452	2.268	0.653	1.369	0.669	1.164	0.664	1.072	0.643	1.027
ln TCJ	0.466	2.240	0.674	**1.341**	0.687	**1.147**	0.681	**1.058**	0.662	**1.013**
Panel C : LHAR-ln RV-X-cum-Vol Model										
ln RV	0.463	2.232	0.675	1.332	0.695	1.136	0.700	1.038	0.692	0.985
ln BPV	0.473	2.239	**0.696**	1.325	**0.726**	**1.113**	**0.734**	**1.010**	**0.725**	**0.959**
ln CTBPV	0.470	2.248	0.690	1.334	0.720	1.118	0.729	1.013	0.719	0.962
ln RPV	**0.475**	**2.231**	0.694	1.324	0.720	1.118	0.725	1.019	0.716	0.969
ln C	0.469	2.242	0.691	1.330	0.721	1.118	0.730	1.014	0.721	0.963
ln CJ	0.469	2.237	0.692	1.324	0.722	1.115	0.731	1.011	0.722	0.961
ln TC	0.467	2.249	0.687	1.334	0.717	1.119	0.726	1.014	0.716	0.964
ln TCJ	**0.475**	**2.231**	0.694	**1.319**	0.720	1.114	0.727	1.013	0.717	0.964

Notes:

a. Entries to the table represent Adjusted R^2 and RMSE for one-day, and one- to four-week in-sample predictions o the RV of TAIEX using the HAR-RV, LHAR-RV, and LHAR-RV-cum-Vol models.

b. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 22 In-sample Forecast Evaluation and Market Conditions: Up- vs. Down- Market Days

Panel A : HAR-RV-X Model

		RV		BPV		CTBPV		RPV		TC		TCJ	
		up	down	up	down	up	down	up	down	up	down	up	down
1 day	R^2	0.169	0.478	0.205	0.422	0.205	0.425	0.179	0.430	0.212	0.417	**0.211**	**0.507**
	RMSE	1.823	1.265	1.783	1.331	1.783	1.328	1.811	1.322	1.775	1.337	**1.774**	**1.228**
1 week	R^2	0.426	0.535	0.483	0.591	0.485	0.590	0.441	0.605	0.490	0.591	**0.495**	**0.620**
	RMSE	0.993	0.933	0.942	0.875	0.940	0.876	0.980	0.859	0.935	0.875	**0.930**	**0.842**
2 weeks	R^2	0.466	0.564	0.547	0.675	0.557	0.677	0.524	0.682	0.563	0.679	**0.566**	**0.688**
	RMSE	0.869	0.813	0.800	0.701	0.791	0.699	0.821	0.694	0.786	0.697	**0.782**	**0.687**
3 weeks	R^2	0.504	0.578	0.571	0.709	0.573	0.713	0.564	0.714	0.581	0.711	**0.586**	**0.718**
	RMSE	0.795	0.755	0.739	0.626	0.737	0.622	0.745	0.621	0.730	0.624	**0.725**	**0.616**
4 weeks	R^2	0.504	0.608	0.555	0.739	0.555	0.744	0.567	0.741	0.558	0.740	**0.568**	**0.744**
	RMSE	0.771	0.696	0.730	0.568	0.730	0.563	0.720	0.566	0.727	0.568	**0.718**	**0.562**

Panel B: HAR-$RV^{1/2}$-X Model

		$RV^{1/2}$		$BPV^{1/2}$		$CTBPV^{1/2}$		RPV		$TC^{1/2}$		$TCJ^{1/2}$	
		up	down	up	down	up	down	up	down	up	down	up	down
1 day	R^2	0.715	0.798	0.722	0.795	0.721	0.796	0.724	**0.805**	0.720	0.793	**0.725**	**0.805**
	RMSE	2.628	2.489	2.628	2.518	2.631	2.518	2.623	2.502	2.628	2.522	**2.621**	**2.498**
1 week	R^2	0.873	0.892	0.877	0.903	0.873	0.902	0.879	**0.909**	0.872	0.901	**0.883**	**0.909**
	RMSE	1.830	1.758	1.830	1.753	1.833	1.755	1.821	1.743	1.832	1.755	**1.817**	**1.741**
2 weeks	R^2	0.890	0.901	0.897	0.920	0.893	0.920	0.897	**0.924**	0.892	0.920	**0.901**	0.923
	RMSE	1.650	1.612	1.642	1.591	1.643	1.591	1.637	**1.585**	1.642	1.591	**1.633**	1.586
3 weeks	R^2	0.895	0.906	0.898	0.928	0.894	0.929	0.897	**0.931**	0.893	0.928	**0.904**	0.930
	RMSE	1.557	1.547	1.552	1.521	1.554	1.520	1.549	1.518	1.553	1.521	**1.542**	**1.516**
4 weeks	R^2	0.885	0.915	0.887	0.937	0.883	0.938	0.886	**0.939**	0.882	0.936	**0.894**	0.938
	RMSE	1.507	1.503	1.505	1.477	1.508	1.476	1.504	**1.475**	1.508	1.503	**1.495**	**1.475**

Notes:

a. The up- and down-market classification are based on the moving average of daily return over the most recent 20-day window to forecast $RV_{t,t+H}$.

b. Entries to the table represent Adjusted R^2 and RMSE for one day, and one week through four weeks in-sample predictions of the RV in the TAIEX.

c. For all models, the dependent variable is the standardized realized variance i.e. $RV_{t,t+H}/H$ for all the horizons.

d. Bold values denote the highest adjusted R^2 and the lowest RMSE.

Table 22 In-sample Forecast Evaluation and Market Conditions: Up- vs. Down- Market Days (*cont.*)

		Panel A : HAR-ln RV-X Model											
		ln RV		ln BPV		ln CTBPV		ln RPV		ln TC		ln TCJ	
		up	down	up	down	up	down	up	down	up	down	up	down
1 day	R^2	0.384	0.471	0.355	0.384	0.333	0.321	0.351	0.400	0.342	0.348	**0.396**	**0.478**
	RMSE	2.427	**2.335**	2.438	2.396	2.445	2.419	2.438	2.384	2.443	2.411	**2.424**	2.349
1 week	R^2	0.627	**0.649**	0.556	0.556	0.508	0.469	0.548	0.568	0.526	0.510	**0.633**	0.664
	RMSE	1.545	**1.483**	1.569	1.535	1.581	1.564	1.570	1.530	1.578	1.552	**1.530**	1.492
2 weeks	R^2	**0.662**	0.662	0.576	0.597	0.522	0.514	0.567	0.600	0.541	0.558	0.661	**0.688**
	RMSE	1.333	**1.284**	1.354	1.322	1.367	1.348	1.356	1.321	1.363	1.335	**1.329**	1.288
3 weeks	R^2	**0.659**	0.670	0.565	0.618	0.511	0.537	0.556	0.617	0.529	0.582	0.655	**0.701**
	RMSE	1.224	**1.188**	1.248	1.216	1.261	1.240	1.250	1.217	1.257	1.228	**1.222**	1.189
4 weeks	R^2	**0.616**	0.692	0.527	0.642	0.478	0.559	0.520	0.641	0.493	0.604	0.609	**0.725**
	RMSE	1.168	**1.122**	1.189	1.145	1.201	1.168	1.191	1.148	1.198	1.157	**1.167**	1.122

Table 23 Out-of-Sample Forecasts of the TAIEX 2007-2008

Panel A:		RV	BPV	CTBPV	RPV	TC	TCJ
	Model 1	3.158	2.978	2.980	**2.951**	2.995	3.145
1 day	Model 2	3.129	2.978	2.987	**2.956**	2.995	3.134
	Model 3	3.123	2.952	2.961	**2.947**	2.968	3.115
	Model 1	1.708	1.620	1.633	**1.561**	1.638	1.639
1 week	Model 2	1.707	1.627	1.644	**1.588**	1.636	1.651
	Model 3	1.665	1.556	1.569	**1.525**	1.559	1.576
	Model 1	1.409	1.307	1.313	**1.278**	1.317	1.312
4 weeks	Model 2	1.397	1.291	1.295	**1.271**	1.298	1.296
	Model 3	1.312	1.147	1.131	1.142	**1.128**	1.149
Panel B :		$RV^{1/2}$	$BPV^{1/2}$	$CTBPV^{1/2}$	RPV	$TC^{1/2}$	$(TCJ)^{1/2}$
	Model 1	3.038	3.042	3.049	**3.004**	3.064	3.028
1 day	Model 2	3.009	3.011	3.023	**2.987**	3.033	3.003
	Model 3	2.997	2.977	2.995	**2.975**	3.001	2.981
	Model 1	1.688	1.725	1.748	**1.675**	1.760	1.713
1 week	Model 2	1.682	1.714	1.739	**1.680**	1.744	1.706
	Model 3	1.636	1.643	1.669	**1.632**	1.673	1.644
	Model 1	1.429	1.432	1.449	1.415	1.458	**1.409**
4 weeks	Model 2	1.417	1.422	1.439	1.406	1.447	**1.397**
	Model 3	1.326	1.269	1.283	1.284	1.293	**1.276**
Panel C :		lnRV	lnBPV	lnCTBPV	lnRPV	lnTC	ln (TCJ)
	Model 1	**3.067**	3.115	3.130	3.088	3.142	3.094
1 day	Model 2	**3.027**	3.064	3.077	3.047	3.087	3.044
	Model 3	**2.993**	3.011	3.029	3.021	3.034	3.000
	Model 1	**1..765**	1.840	1.872	1.812	1.886	1.832
1 week	Model 2	**1.744**	1.810	1.840	1.794	1.851	1.804
	Model 3	**1.689**	1.722	1.753	1.735	1.762	1.726
	Model 1	**1.491**	1.547	1.573	1.536	1.583	1.524
4 weeks	Model 2	**1.477**	1.539	1.567	1.530	1.576	1.513
	Model 3	**1.372**	1.387	1.414	1.410	1.425	1.388

Notes:

a. The table presents the out-of-sample results for TAIEX from 2 January 2007 to 30 June 2008. The dependent variable is RV. Data from 2 January 2003 to 29 December 2006 is used to estimate the parameters of the models

b. Entries to the table represent RMSE for the out-of-sample predictions, based upon one-say and one- to four-week out-of-sample RV prediction horizons.

c. Model 1 denotes the HAR-RV model; Model 2 denotes the LHAR-RV model; Model 3 denotes the LHAR-RV-cum-Vol model.

d. Bold values denote the highest adjusted R^2 and the lowest RMSE.

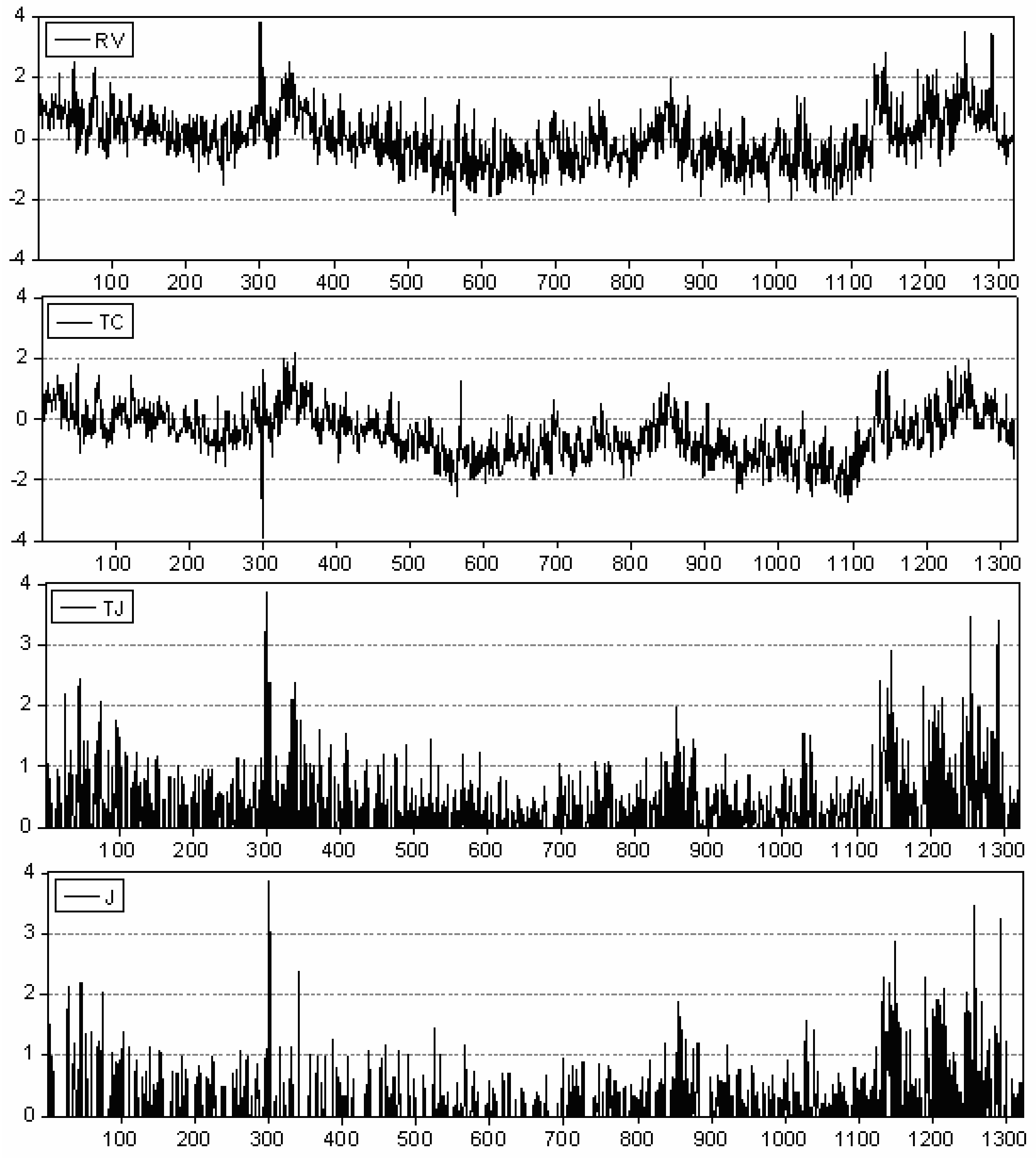

Figure 1 Time Series Plot for Log Realized Volatility of the TAIEX 2003-2008

Notes:

a. The top panel shows daily realized volatility in log form, or $\ln RVt$.

b. The second panel graphs the threshold continuous component defined in Equation (22), $TC_{t.}$

c. The third panel graphs the significant threshold jumps corresponding to α = 99.9% defined in Equation (20), TJ_t.

d. The bottom panel graphs the significant jumps defined in Equation (16), J_t.

e. The sample period covers from 2 January 2003 to 30 June 2008.

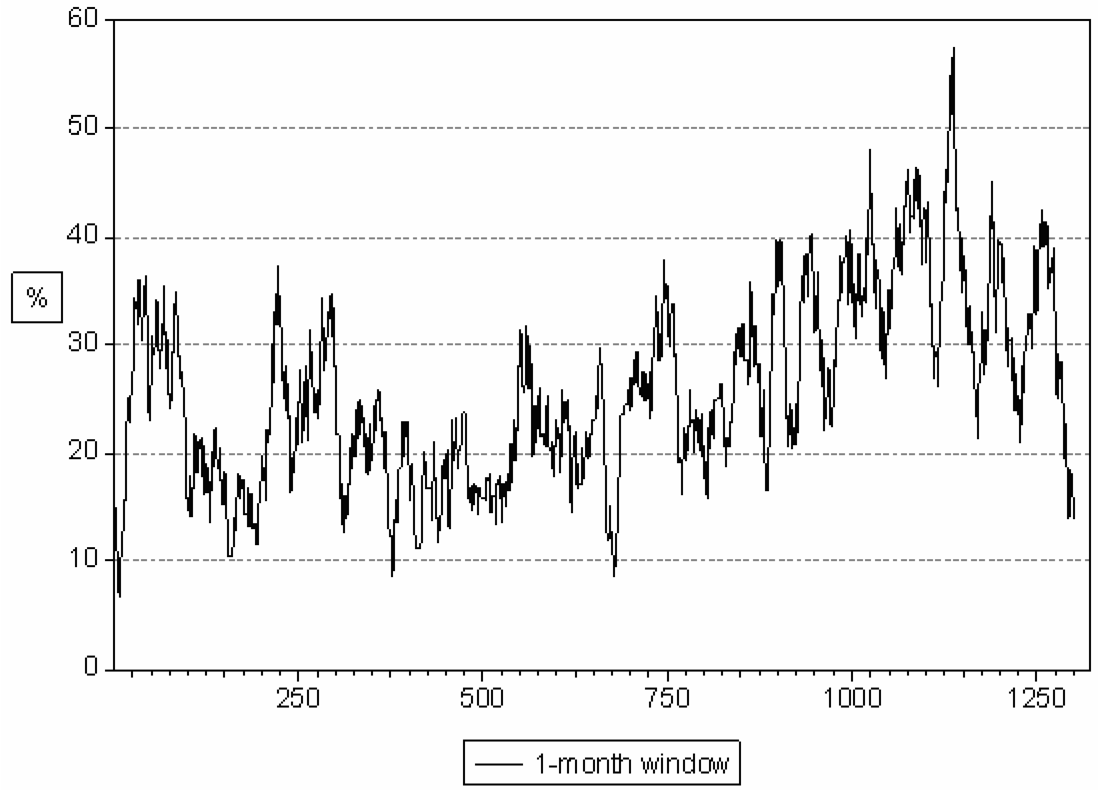

Figure 2 Jump Contribution to Total Variance

Notes:

a. Percentage contribution of daily jump estimated by Equation (20) to total quadratic variation measured over a moving average window of 1-month for the TAIEX.

b. The C-Tz statistic in Equation (20) is computed with confidence interval α= 99.9% .

c. The sample period covers from 2 January 2003 to 30 June 2008.

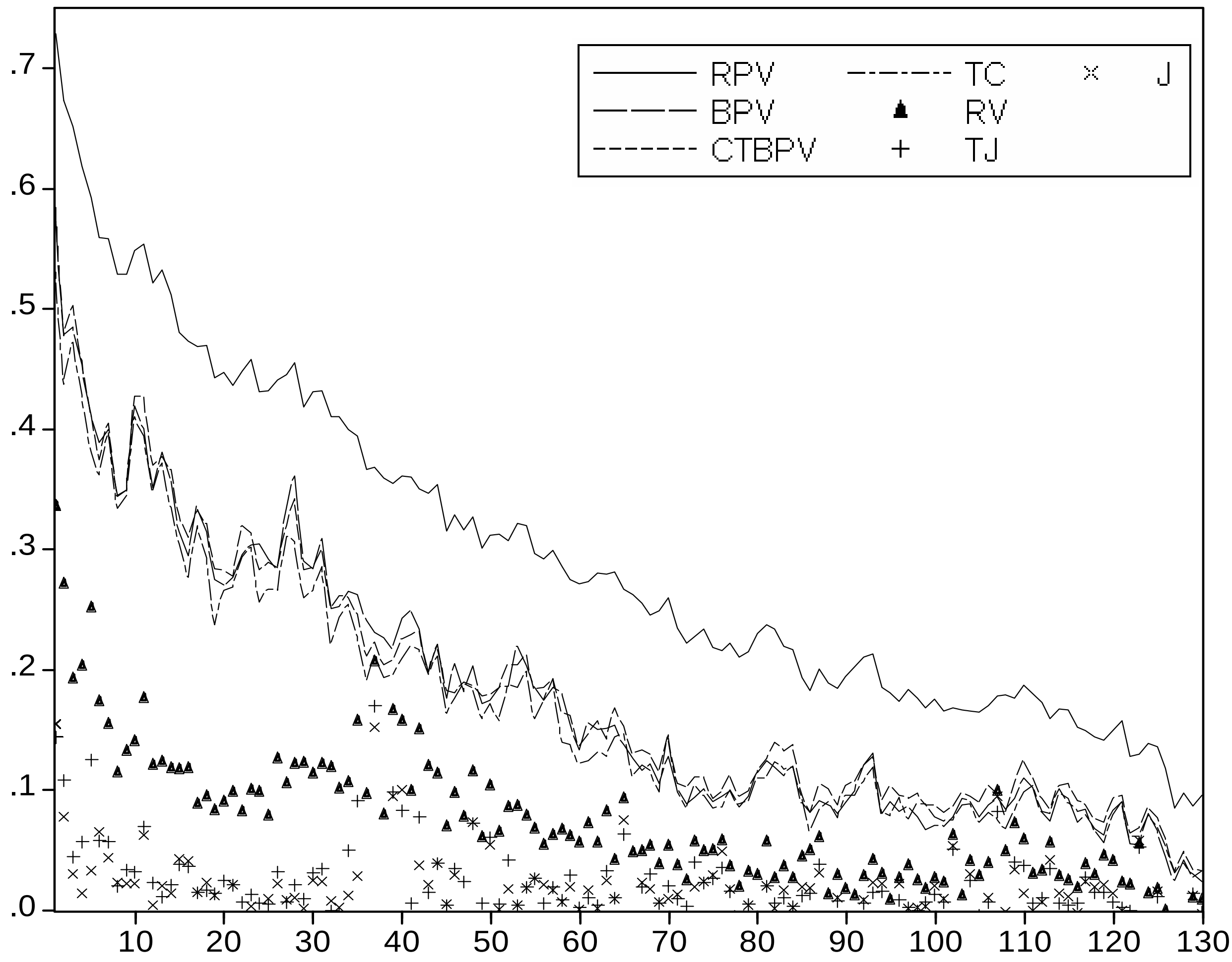

Figure 3 Sample Autocorrelation Function for the TAIEX Volatility Measures and Decompositions

Notes:

a. The figure shows the SACF for *RV*, *RPV*, *CTBPV*, *RPV*, *TC*, *TJ*, and Jumps for the period.

b. The significance level of the threshold bipower test and bipower test is 0.999.

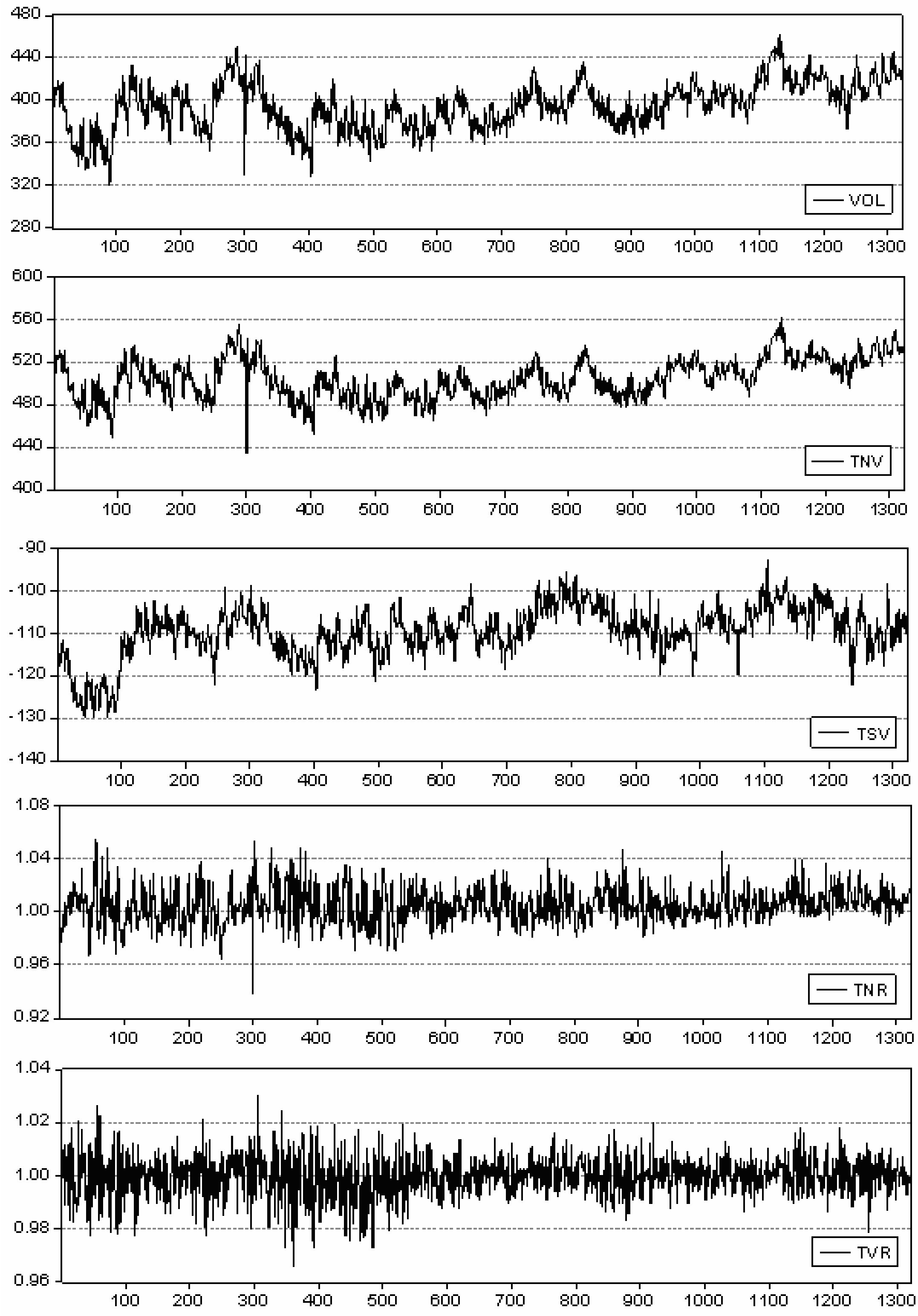

Figure 4 Daily Trading Volumes in the TAIEX

Notes: a. The top three panels graph trading volume value (Vol_t) , number of transactions (TNV_t), and average trade size (TSV_t), respectively.

b. The bottom two panels graph bid-ask frequency and bid-ask volume, respectively.

c. The sample period covers from 2 January 2003 to 30 June 2008.

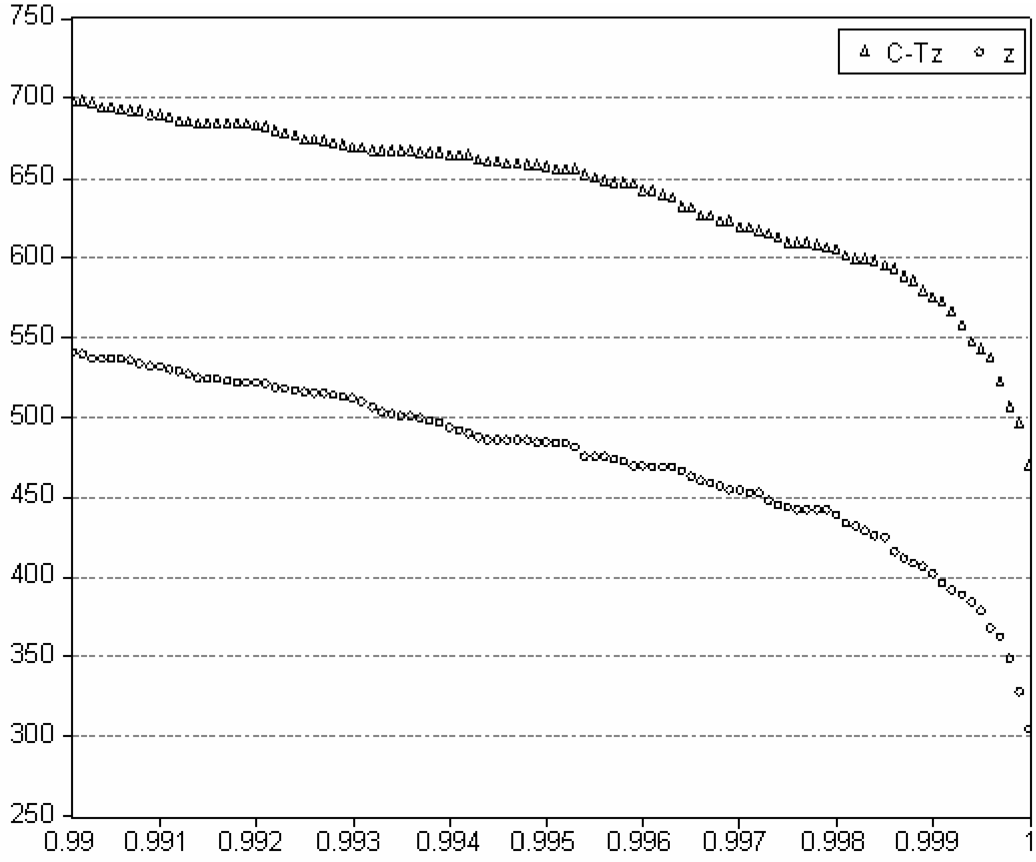

Figure 5　Number of Jump Days

Notes:

a. Number of days which contain jumps in the TAIEX sample obtained with the C-Tz statistics (21) and z statistics (17) , as a function of the confidence level α.

b. The sample period covers from 2 January 2003 to 30 June 2008 for a total of 1361 days.