IMES DISCUSSION PAPER SERIES

Pricing Nikkei 225 Options Using Realized Volatility

Masato Ubukata and Toshiaki Watanabe

Discussion Paper No. 2011-E-18

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

2-1-1 NIHONBASHI-HONGOKUCHO
CHUO-KU, TOKYO 103-8660
JAPAN

IMES Discussion Paper Series 2011-E-18
August 2011

Pricing Nikkei 225 Options Using Realized Volatility

Masato Ubukata * and Toshiaki Watanabe **

Abstract

This article analyzes whether daily realized volatility, which is the sum of squared intraday returns over a day, is useful for option pricing. Different realized volatilities are calculated with or without taking account of microstructure noise and with or without using overnight and lunch-time returns. ARFIMA, ARFIMAX, HAR, HARX models are employed to specify the dynamics of realized volatility. ARFIMA and HAR models can capture the long-memory property and ARFIMAX and HARX models can also capture the asymmetry in volatility depending on the sign of previous day's return. Option prices are derived under the assumption of risk-neutrality. For comparison, GARCH, EGARCH and FIEGARCH models are estimated using daily returns, where option prices are derived by assuming the risk-neutrality and by using the Duan (1995) method in which the assumption of risk-neutrality is relaxed. Main results using the Nikkei 225 stock index and its put options prices are: (1) ARFIMAX model with daily realized volatility performs best, (2) the Hansen and Lunde (2005a) adjustment without using overnight and lunch-time returns can improve the performance, (3) if the Hansen and Lunde (2005a), which also plays a role to remove the bias caused by the microstructure noise by setting the sample mean of realized volatility equal to the sample variance of daily returns, is used, the other methods for taking account of microstructure noise do not necessarily improve the performance and (4) the Duan (1995) method does not improve the performance compared with assuming the risk neutrality.

Keywords: microstructure noise; Nikkei 225 stock index; non-trading hours; option pricing; realized volatility

JEL classification: C22, C52, G13

***** Assistant Professor, Department of Economics, Kushiro Public University of Economics
(E-mail: ubukata@kushiro-pu.ac.jp)
****** Professor, Institute of Economic Research, Hitotsubashi University
(E-mail: watanabe@ier.hit-u.ac.jp)
Institute for Monetary and Economic Studies, Bank of Japan

Financial support from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government through Grant-in-Aid for Scientific Research (No.18203901; 21243018; 22243021; 23730301) and the Global COE program “Research Unit for Statistical and Empirical Analysis in Social Sciences” at Hitotsubashi University and from the Joint Usage and Research Center, Institute of Economic Research, Hitotsubashi University (IERPK1109) is gratefully acknowledged. The authors also thank seminar participants at the Bank of Japan, Tohoku University, Hiroshima University of Economics and Nihon University and the staff of the Institute for Monetary and Economic Studies (IMES), the Bank of Japan, for their useful comments. Views expressed in this paper are those of the authors and do not necessarily reflect the official views of the Bank of Japan.

1 Introduction

One of the most important variables in option pricing is the volatility of the underlying asset. While the well-known Black and Scholes (1973) model assumes that the volatility is constant, few would dispute the fact that the volatility changes over time. Many time series models are now available to describe the dynamics of volatility. One of the most widely used is the ARCH (autoregressive conditional heteroskedasticity) family including ARCH model by Engle (1982), GARCH (generalized ARCH) model by Bollerslev (1986) and their extensions.

The problem of using these models is that we must specify the model before estimating the volatility and the estimate of volatility depends on the specification of volatility dynamics. Recently, realized volatility has attracted the attentions of financial econometricians as an accurate estimator of volatility. Realized volatility is independent of the specification of volatility dynamics because it is simply the sum of squared intraday returns.

ARCH type models have already been applied to option pricing (Bollerslev and Mikkelsen, 1999; Duan, 1995). As far as we know, there are few which have applied realized volatility to option pricing compared with the applications to volatility forecasting (Koopman et al. 2005) and Value-at-Risk (Giot and Laurent, 2004; Clements et al., 2008). One exception is Bandi et al. (2008), which apply realized volatility to the pricing of S&P 500 index options. This article applies realized volatility to the pricing of Nikkei 225 stock index options traded at Osaka Securities Exchange and compares its performance with that of using the ARCH family.

There are two problems in calculating realized volatility. First, realized volatility is influenced by market microstructure noise such as bid-ask spread and non-synchronous trading (Campbell et al., 1997). There are some methods available for mitigating the effect of microstructure noise on realized volatility (Aït-Sahalia et al., 2005; Bandi and Russell, 2006, 2008, 2011; Barndorff-Nielsen et al., 2004, 2008; Hansen and Lunde, 2006; Kunitomo and Sato 2008; Oya 2011; Zhang, 2006; Zhang et al., 2005; Zhou 1996). It is worthwhile applying these methods and comparing the results. We use several different methods for mitigating the effect of microstructure noise on realized volatility. We analyze whether using these methods may improve the performance of option pricing. Second, the Tokyo stock exchange, where the 225 stocks that constitute the Nikkei 225 stock index are traded, opens only for 9:00–11:00 and 12:30–15:00. We cannot obtain high-frequency returns during the period when the market is closed. Adding the squares of overnight (15:00-9:00) and lunch-time (11:00-12:30) returns may make realized volatility noisy. Following Hansen and Lunde (2005a), we calculate realized volatility without overnight and lunch-time returns and multiply a constant such that the sample mean of daily realized volatility is equal to the sample variance of daily returns. We examine whether this method is effective in option pricing by comparing with simply adding the squares of overnight and lunch-time returns.

Many authors have documented that realized volatility follows a long-memory process (Andersen et al., 2001, 2003). We use the ARFIMA (autoregressive fractionally integrated moving average) model and HAR (heterogeneous interval autoregressive) model by Corsi (2009) to describe the dynamics of realized volatility. It is also well known in stock markets that today's volatility is negatively correlated with yesterday's return. We also extend ARFIMA and HAR models to take account of this asymmetry in volatility.

For ARCH type models, we use the simple GARCH model proposed by Bollerslev (1986), the EGARCH (exponential GARCH) model by Nelson (1991) that may capture the asymmetry in volatility and the FIEGARCH (fractionally integrated EGARCH) model by Bollerslev and Mikkelsen (1996) that may also allow for the long-memory property of volatility.

We calculate option prices under the assumption of risk neutrality. Duan (1995) has developed a more general method for pricing options in ARCH type models, which does not assume risk neutrality. We also calculate option prices both assuming the risk neutrality and by using the Duan (1995) method.

Main findings are: (1) ARFIMAX model with daily realized volatility performs best, (2) the Hansen and Lunde (2005a) adjustment without using overnight and lunch-time returns can improve the performance, (3) if the Hansen and Lunde (2005a), which also plays a role to remove the bias from the microstructure noise by setting the sample mean of realized volatility equal to the sample variance of daily returns, is used, the other methods for taking account of microstructure noise do not necessarily improve the performance and (4) the Duan (1995) method does not improve the performance compared with assuming the risk neutrality.

The article proceeds as follows. Section 2 explains several methods used in this article for calculating realized volatilities. Section 3 explains ARFIMA(X) and HAR(X) models to describe the dynamics of realized volatility and ARCH type models used in this article for comparison. Section 4 explains how to calculate option prices using the ARFIMA(X) and HAR(X) models with daily realized volatility and ARCH type models with daily returns. Section 5 explains the data and Section 6 compares the performance of option pricing. Section 7 concludes. The appendix provides a detailed description of realized volatilities employed in this article.

2 Realized Volatility

We start with a brief review of realized volatility using the following diffusion process.

$$dp(s) = \mu(s)ds + \sigma(s)dW(s), \tag{1}$$

where s is time, $p(s)$ is the log-price, $W(s)$ is a standard Brownian motion, and $\mu(s)$ and $\sigma(s)$ are the drift and the volatility respectively, which may be time-varying but are assumed to be independent of $dW(s)$. In this article, we call $\sigma(s)$ or $\sigma^2(s)$ volatility interchangeably although $\sigma(s)$ is usually called volatility in the finance literature. Then, the volatility for day t is defined as the integral of $\sigma^2(s)$ over the interval $(t-1, t)$ where $t-1$ and t represent the market closing time on day $t-1$ and t respectively, i.e.,

$$IV_t = \int_{t-1}^{t} \sigma^2(s)ds, \tag{2}$$

which is called integrated volatility. The integrated volatility is unobservable, but if we have the intraday return data $\left(r_{t-1+1/n}\ r_{t-1+2/n},\ \ldots,\ r_t\right)$, we can estimate it as the sum of their squares

$$RV_t = \sum_{i=1}^{n} r^2_{t-1+i/n}, \tag{3}$$

which is called realized volatility. If the prices do not include any noise, realized volatility RV_t will provide a consistent estimate of IV_t, i.e.,

$$\operatorname*{plim}_{n\to\infty} RV_t = IV_t. \tag{4}$$

There are two problems in calculating realized volatility. First, although the realized volatility is an accurate estimator of integrated volatility under the assumption of a continuous stochastic model, it fails when there is market microstructure noise as seen in real high-frequency data. The microstructure noise can be induced by various market frictions such as the discreteness of price changes, bid-ask bounces, and asymmetric information across traders, inter alia.[1] A growing literature attempts to study an integrated volatility estimation from microstructure noise-contaminated high-frequency data. In this article, we employ some influential integrated volatility estimators robust to the microstructure noise.

Second, the Tokyo Stock Exchange is open only for 9:00–11:00 (morning session) and 12:30–15:00 (afternoon session) except for the first and last trading days in every year, when it is open only for 9:00-11:00. It is impossible to obtain high-frequency returns for 15:00–9:00 (overnight) and 11:00–12:30 (lunch-time). Since realized volatility obtained using high-frequency returns over 4.5-hour trading period only captures the volatility during the part of the day that the market is open, we need to extend the realized volatility to a measure of volatility for the full day. If we simply add the squares of overnight and lunch-time returns, realized volatility may be subject to discretization error. Hansen and Lunde (2005a) propose to calculate realized volatility only when the market is open, which is denoted as $RV_t^{(o)}$, and multiply a constant c such that the sample mean of realized volatility is equal to the sample variance of daily returns, i.e.,

$$RV_t = cRV_t^{(o)}, \quad c = \frac{\sum_{t=1}^{T}(R_t - \overline{R})^2}{\sum_{t=1}^{T} RV_t^{(o)}}, \tag{5}$$

where $(R_1, \ldots, R_T)$ is the sample of daily returns and $\overline{R}$ is the sample mean[2].

In order to test the effects of taking into consideration the microstructure noise and the non-trading

[1] The literature on market microstructure provides important insights from early studies including Roll (1984), who derives a simple estimator of the bid-ask spread based on the negative autocovariance of returns. Harris (1990) examines the rounding effects emanating from the discreteness of transaction prices. In the recent literature on microstructure noise, Meddahi (2002) and Hansen and Lunde (2006) examine the variance of microstructure noise as well as the correlation between the microstructure noise and frictionless equilibrium price. Ubukata and Oya (2009) examine dependence of microstructure noise.

[2] See Martens (2002) and Hansen and Lunde (2005b) for the other methods.

hours on option pricing, we use as many as 30 daily realized volatilities listed in Table 1. Without microstructure noise, it would be desirable to use intraday returns sampled at the highest frequencies. Since the highest frequencies available for Nikkei 225 stock index is 1-minute, we first calculate realized volatility using 1-minute returns ($n = 270$). From the second to fifteenth methods in Table 1 are expected to correct the bias of the classical realized volatility and mitigate the variance increase of the estimator induced by the microstructure noise. A more detailed description of the methods is provided in the appendix. We apply the Hansen and Lunde (2005a) adjustment to the 15 kinds of realized volatilities, which are denoted as $RV(\text{1min})^{HL}$, $RV(\text{5min})^{HL}$, $RV(\text{15min})^{HL}$, $RV(BR)^{HL}$, $BK(BR)^{HL}$, $ZMA(ZMA)^{HL}$, $ZMA(BR)^{HL}$, $BC(ZMA, ZMA)^{HL}$, $BC(ZMA, BR)^{HL}$, $FBK(BNHLS)^{HL}$, $FCK(BNHLS)^{HL}$, $FMTH(BNHLS)^{HL}$, $FBK(BR)^{HL}$, $FCK(BR)^{HL}$, $FMTH(BR)^{HL}$. For comparison, we also calculate 15 kinds of daily realized volatilities constructed by adding the sqaures of overnight and lunch-time returns instead of the Hansen and Lunde (2005a) adjustment, which are denoted as $RV(\text{1min})^{SR}$, $RV(\text{5min})^{SR}$, $RV(\text{15min})^{SR}$, $RV(BR)^{SR}$, $BK(BR)^{SR}$, $ZMA(ZMA)^{SR}$, $ZMA(BR)^{SR}$, $BC(ZMA, ZMA)^{SR}$, $BC(ZMA, BR)^{SR}$, $FBK(BNHLS)^{SR}$, $FCK(BNHLS)^{SR}$, $FMTH(BNHLS)^{SR}$, $FBK(BR)^{SR}$, $FCK(BR)^{SR}$, $FMTH(BR)^{SR}$.

3 ARFIMA(X), HAR(X) and ARCH type Model

Many researchers have documented that realized volatility may follow a long-memory process. Let $\rho(h)$ denote the h-th order autocorrelation coefficient of variable X. Then, X follows a short-memory process if $\sum_{h=0}^{\infty} |\rho(h)| < \infty$ and a long-memory process if $\sum_{h=0}^{\infty} |\rho(h)| = \infty$. A stationary ARMA model is a short-memory process. As h increases, the autocorrelation coefficient $\rho(h)$ of the long-memory process decays more slowly than that of the short-memory process. More specifically, the former decays hyperbolically and the latter decays geometrically.

The most widely used for a long-memory process is ARFIMA(p, d, q) model[3]

$$\phi(L)(1-L)^d X_t = \theta(L) u_t, \quad u_t \sim \text{NID}(0, \sigma^2), \tag{6}$$

where L denotes the lag operator and $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$ are the p-th and q-th order lag polynomials assumed to have all roots outside the unit circle. The order of integration d is allowed to take non-integer values. If $d = 0$, ARFIMA model collapses to stationary ARMA model and if $d = 1$, it becomes non-stationary ARIMA model. If $0 < d < 0.5$, X_t follows a stationary long-memory process and if $0.5 \leq d < 1$, X_t follows a non-stationary long-memory process. $(1-L)^d$ may be written as follows.

$$(1-L)^d = 1 + \sum_{k=1}^{\infty} \frac{d(d-1)\cdots(d-k+1)}{k!}(-L)^k. \tag{7}$$

[3] See Beran (1994) for the details of long-memory and ARFIMA model.

We assume that u_t follows an independent normal distribution with zero mean and variance σ^2.

By setting $p = 0$ and $q = 1$, which are selected by SIC, and $X_t = \ln(RV_t) - \mu$ where μ is the unconditional mean of $\ln(RV_t)$, we consider the following model.

$$(1 - L)^d [\ln(RV_t) - \mu] = u_t + \theta u_{t-1}, \quad u_t \sim \text{NID}(0, \sigma^2). \tag{8}$$

We estimate parameters d, μ and θ jointly using the approximate maximum likelihood method (Beran, 1995), where it is assumed that $\ln(RV_t) = \mu$ $(t = 0, -1, \ldots)$. We can estimate σ^2 as the sample variance of residual.

We also employ HAR model by Corsi (2009) well-known as a simple approximate long-memory model of realized volatility. The model consists of three realized volatility components defined over different time periods as follows

$$\ln(RV_t) = \beta_0 + \beta_1 \ln(RV_{t-1}) + \beta_2 \ln(RV_{t-1}^w) + \beta_3 \ln(RV_{t-1}^m) + v_t, \quad v_t \sim \text{NID}(0, \sigma_v^2), \tag{9}$$

where $RV_{t-1}^w = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}$ and $RV_{t-1}^m = \frac{1}{22}\sum_{i=1}^{22} RV_{t-i}$ are the average of the past realized volatilities corresponding to time horizons of 5 trading days (one week) and 22 trading days (one month), respectively. We can estimate parameters β_0, β_1, β_2, β_3 and σ_v^2 by applying simple linear regression.

It is well-known that there is a negative correlation between today's return and tomorrow's volatility in stock markets. To take into account this phenomenon, we extend the above ARFIMA(0,d,1) model (8) to the following ARFIMA(0,d,1)-X model

$$(1 - L)^d \left[\ln(RV_t) - \mu_0 - \mu_1 |R_{t-1}| - \mu_2 D_{t-1}^- |R_{t-1}|\right] = u_t + \theta u_{t-1}, \quad u_t \sim \text{NID}(0, \sigma^2), \tag{10}$$

where D_{t-1}^- is a dummy variable that takes one if the return on day $t-1$ is negative and zero otherwise. We estimate parameters d, μ_0, μ_1, μ_2, θ and σ^2 using the same method as that for ARFIMA model. If the estimate of μ_2 has a statistically significant positive value, it is consistent with a well-known negative correlation between today's return and tomorrow's volatility in stock markets. The HAR model (9) can be naturally extended to HAR-X model taking account of the asymmetry in volatility as follows

$$\ln(RV_t) = \beta_0 + \beta_1 \ln(RV_{t-1}) + \beta_2 \ln(RV_{t-1}^w) + \beta_3 \ln(RV_{t-1}^m) + \beta_4 |R_{t-1}| + \beta_5 D_{t-1}^- |R_{t-1}| + v_t, \quad v_t \sim \text{NID}(0, \sigma_v^2). \tag{11}$$

We estimate parameters β_0, β_1, β_2, β_3, β_4, β_5 and σ_v^2 using the same method as that for the HAR model. The positive value of β_5 indicates the negative correlation between today's return and tomorrow's volatility.

Some researchers such as Barndorff-Nielsen et al. (2004), Barndorff-Nielsen and Shephard (2001, 2002b) and Nagakura and Watanabe (2010) have proposed a UC (unobserved components) model[4]. Assuming that the asset price follows a contimuous-time model called square-root stochastic variance model, they show that the realized volatility calculated using the discretely sampled data follows an ARMA(1,1) model. Since it is the realized volatility rather than its log that follows an ARMA(1,1) model and the distribution of the error term is unknown, the future volatility sampled for option pricing may possibly be negative if we assume that the distribution of error term is normal. Thus, we do not use this model in this article.

We also estimate ARCH type models using daily returns. We define daily return as

$$R_t = \ln(S_t) - \ln(S_{t-1}), \tag{12}$$

where S_t is the closing price on day t. We specify daily return as

$$R_t = \mathrm{E}(R_t|\boldsymbol{I}_{t-1}) + \epsilon_t,\ \epsilon_t = \sigma_t z_t,\ z_t \sim \text{NID}(0,1), \tag{13}$$

where $\mathrm{E}(R_t|\boldsymbol{I}_{t-1})$ is the expectation of R_t conditional on the information up to day $t-1$ and z_t is assumed to follow an independent standard normal distribution. Then, σ_t^2 is the variance of R_t conditional on the information up to day $t-1$. We will explain how to specify $\mathrm{E}(R_t|\boldsymbol{I}_{t-1})$ later.

For volatility specification, we use three different ARCH type models. First is the GARCH model proposed by Bollerslev (1986). Specifically, we use the GARCH$(1,1)$ model

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2,\ \omega > 0,\ \beta,\alpha \geq 0, \tag{14}$$

where ω, β and α are parameters, which are assumed to be non-negative to guarantee that volatility is always positive. This model can capture the volatility clustering. Volatility is stationary if $|\beta+\alpha| < 1$, and the speed for which the shock to volatility decays becomes slower as $\beta + \alpha$ approaches to one.

As has already been mentioned, another well-known phenomenon in stock markets is volatility asymmetry, which cannot be captured by the above GARCH model. To capture this phenomenon, we also use the EGARCH model proposed by Nelson (1991). Specifically, we use the EGARCH$(1,0)$ model

$$\ln(\sigma_t^2) = \omega + \phi\left[\ln(\sigma_{t-1}^2) - \omega\right] + \theta z_{t-1} + \gamma\left(|z_{t-1}| - \mathrm{E}\,|z_{t-1}|\right), \quad |\phi| < 1. \tag{15}$$

While the GARCH model specifies the process of σ_t^2, the EGARCH model specifies that of its logarithm. Thus, it does not require non-negativity constraints for parameters. If $\theta < 0$, it is consistent with the volatility asymmetry in stock markets. In this model, volatility is stationary if $|\phi| < 1$, and the speed for which the shock to volatility decays becomes slower as ϕ approaches to one. Since z_{t-1} is assumed to follow the standard normal distribution, $\mathrm{E}\,|z_{t-1}| = \sqrt{2/\pi}$.

Neither the GARCH nor EGARCH models allow volatility to have long-memory property. Hence,

[4]Nagakura and Watanabe (2010) consider microstructure noise while Barndorff-Nielsen et al. (2004) and Barndorff-Nielsen and Shephard (2001, 2002b) neglect it.

we also use the FIEGARCH model proposed by Bollerslev and Mikkelsen (1996). Since this model is an extension of the above EGARCH model to allow the long-memory of volatility, it can also capture the volatility asymmetry. We use the following FIEGARCH($1, d, 0$) model.

$$(1-\phi L)(1-L)^d\left[\ln(\sigma_t^2)-\omega\right]=\theta z_{t-1}+\gamma\left(|z_{t-1}|-\mathrm{E}\,|z_{t-1}|\right),\quad |\phi|<1. \tag{16}$$

Similarly to the EGARCH model, it is consistent with the volatility asymmetry in stock markets if $\theta < 0$. As for d, the same argument as that for ARFIMA model holds.

FIGARCH (Baillie et al., 1996) and FIAPGARCH (Tse, 1998) models can also take into account the possibility that the volatility follows a long-memory process. These models, however, have some drawbacks. First, the variance of return will be infinite even though $0 < d < 0.5$ (Schoffer, 2003). Second, the parameter constraints to guarantee that the volatility is always positive are complicated (Conrad and Haag, 2006). Thus, we do not use these models in this article. We estimate parameters in the GARCH, EGARCH and FIEGARCH models using the maximum likelihood method[5].

4 Option Pricing

We first calculate option prices under the assumption of risk neutrality. If the traders are risk neutral, the expected return may be represented by

$$\mathrm{E}(R_t|\boldsymbol{I}_{t-1})=r-d-\frac{1}{2}\sigma_t^2, \tag{17}$$

where r and d are continuously compounded risk-free rate and dividend rate.

The price of European option will be equal to the discounted present value of the expectation of option prices on the expiration date. For example, the price of European put option with the exercise price K and the maturity τ is given by

$$P_T=\exp(-r\tau)\mathrm{E}\left[\mathrm{Max}(K-\tilde{S}_{T+\tau},0)|\boldsymbol{I}_T\right], \tag{18}$$

where $\tilde{S}_{T+\tau}$ is the price of the underlying asset on the expiration date $T+\tau$.

We cannot evaluate this expectation analytically if the volatility of the underlying asset follows ARFIMA(X), HAR(X) or ARCH type models. We calculate option prices by simulating $\tilde{S}_{T+\tau}$ from ARFIMA(X), HAR(X) or ARCH type models. Suppose that $(S_{T+\tau}^{(1)},\ldots,S_{T+\tau}^{(m)})$ are simulated. Then, (18) may be calculated as follows.

$$P_T\approx\exp(-r\tau)\frac{1}{m}\sum_{i=1}^{m}\mathrm{Max}(K-S_{T+\tau}^{(i)},0). \tag{19}$$

We set $m = 10000$. For variance reduction, we used the control variate and the Empirical Martingale

[5]See Taylor (2001) for the estimation method for the FIEGARCH model.

Simulation proposed by Duan and Shimonato (1998) jointly.

Duan (1995) relaxed the assumption of risk neutrality to derive option prices when the price of underlying asset follows ARCH type models. We also use this method. Following Duan (1995), we set

$$\mathrm{E}(R_t|\boldsymbol{I}_{t-1}) = r - d - \frac{1}{2}\sigma_t^2 + \lambda\sigma_t, \tag{20}$$

where $\lambda\sigma_t$ captures the risk premium.

Unless the traders are risk neutral, we must convert the physical measure P into the risk neutral measure Q and evaluate the expectation in equation (18) under the risk neutral measure Q. Duan (1995) makes the following assumptions on Q, called local risk-neutral valuation relationship (LRNVR).

1. $R_t|I_{t-1}$ follows a normal distribution under the risk neutral measure Q.
2. $\mathrm{E}^Q[\exp(R_t)|I_{t-1}] = \exp(r-d)$.
3. $\mathrm{Var}^Q[R_t|I_{t-1}] = \mathrm{Var}^P[R_t|I_{t-1}]$ a.s.

Under assumptions 1 and 2, daily returns under the risk neutral measure Q must be represented by

$$R_t = r - d - \frac{1}{2}\sigma_t^2 + \xi_t, \quad \xi_t = \sigma_t w_t, \quad w_t \sim \mathrm{NID}(0,1). \tag{21}$$

Comparing equation (21) with equations (13) and (20) leads to

$$\epsilon_t = \xi_t - \lambda\sigma_t, \tag{22}$$

$$z_t = w_t - \lambda. \tag{23}$$

Since assumption 3 means that volatilities are the same between P and Q, all we have to do for volatility is to substitute equations (22) or (23) into ϵ_t in the GARCH volatility equation or z_t in the EGARCH and FIEGARCH volatility equations. For example, the GARCH$(1,1)$ volatility equation will be

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha(\xi_{t-1} - \lambda\sigma_{t-1})^2, \ \omega > 0, \ \beta, \alpha \geq 0. \tag{24}$$

Equations (21) and (24) constitute GARCH$(1,1)$ model under Q. Hence, we can evaluate the option prices as follows.

[1] Estimate the parameters λ, ω, β and α in GARCH$(1,1)$ model under P that consists of equations (13), (20) and (14).

[2] Simulate $\tilde{S}_{T+\tau}$ using GARCH$(1,1)$ model under Q that consists of equations (21) and (24) by setting the parameters λ, ω, β and α equal to their estimates in [1].

[3] Substitute $(S_{T+\tau}^{(1)}, \ldots, S_{T+\tau}^{(m)})$ simulated in [2] into equation (19) to obtain the option price.

Similarly, we can calculate the option price using the EGARCH and FIEGARCH models. The EGARCH $(1,0)$ and FIEGARCH$(1,d,0)$ volatility equations under Q will be

$$\ln(\sigma_t^2) = \omega + \phi\left[\ln(\sigma_{t-1}^2) - \omega\right] + \theta(v_{t-1} - \lambda) + \gamma\left(|v_{t-1} - \lambda| - \sqrt{2/\pi}\right), \tag{25}$$

$$(1-\phi L)(1-L)^d\left[\ln(\sigma_t^2) - \omega\right] = \theta(v_{t-1} - \lambda) + \gamma\left(|v_{t-1} - \lambda| - \sqrt{2/\pi}\right). \tag{26}$$

For comparison, we also calculate option prices using the Black-Scholes formula with volatility σ as the standard deviation of daily returns over the past 20 days.

5 Data

We analyze the Nikkei 225 stock index options traded at the Osaka Securities Exchange. The underlying asset is the Nikkei 225 stock index, which is the average of the prices of 225 representative stocks traded at the Tokyo Stock Exchange. The sample period is from May 29, 1996 to September 27, 2007. Following equation (12), we calculate the daily returns for the underlying asset as the log-difference of the closing prices of the Nikkei 225 index in consecutive days. Table 2 summarizes the descriptive statistics of the daily returns (%) for the full sample. The mean is not significantly different from zero. While the skewness is not significantly different from zero, the kurtosis is significantly above 3, indicating the well-known phenomenon that the distribution of the daily return is leptokurtic. LB(10) is the Ljung-Box statistic adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelations up to 10 lags. According to this statistic, the null hypothesis is not rejected at the 1% significance level although it is rejected at the 5% level. We do not consider autocorrelations in the daily return in the following analyses.

We calculate realized volatility using the Nikkei NEEDS-TICK data. This dataset includes the Nikkei 225 stock index for every minute from 9:01 to 11:00 in the morning session and from 12:31 to 15:00 in the afternoon session. Sometimes, the time stamps for the closing prices in the morning and afternoon sessions are slightly after 11:00 and 15:00 because the recorded time shows when the Nikkei 225 stock index is calculated. In such cases, we use all prices up to closing prices. Using these prices, the 30 daily different realized volatilities listed in Table 1 are calculated with or without using the adjustment coefficient c defined by equation (5).

Figure 1 plots some kinds of realized volatilities and Table 3 summarizes the descriptive statistics of the 30 daily different realized volatilities. From $RV(\text{1min})^{HL}$ to $FMTH(BR)^{HL}$ are adjusted such that the mean of realized volatility is equal to the sample variance of daily returns, but their means are different because the adjustment coefficient c is calculated day by day using the past 1200 realized volatilities and daily returns. From $RV(\text{1min})^{SR}$ to $FMTH(BR)^{SR}$ are not adjusted and their means are much lower than those of the others. Among the 15 realized volatilities with the Hansen and Lunde (2005a) adjustment, $RV(\text{1min})^{HL}$ has the smallest standard deviation. $RV(\text{15min})^{HL}$ has the largest standard deviation of them as induced by the range from the minimum at 0.0635 to the maximum at 35.9133. The standard deviation of $ZMA(ZMA)^{SR}$ is the smallest of all. These results are confirmed by Figure 1. Figure 1(a) shows that $RV(\text{15min})^{HL}$ is more volatile than $RV(\text{1min})^{HL}$ and $RV(BR)^{HL}$, and Figure 1(b) shows that $RV(\text{1min})^{SR}$ is smaller on average and less volatile than $RV(\text{1min})^{HL}$. The values of skewness and kurtosis indicate that the distributions of all realized volatilities are non-normal. LB(10) is so large that the null hypothesis of no autocorrelation is rejected. Table 3 (b) shows the descriptive statistics for log-realized volatilities. They are qualitatively the same as those of Table 3 (a) except skewness and kurtosis. While realized volatilities are positively skewed, log-realized volatilities are negatively skewed at the 5% significant level except $\ln(RV(\text{15min})^{HL})$, $\ln(ZMA(BR)^{HL})$ and $\ln(RV(\text{15min})^{SR})$. The kurtosis of log-realized volatilities is much smaller than those of realized volatilities. The kurtosis of $\ln(RV(\text{1min})^{HL})$, $\ln(RV(\text{1min})^{SR})$ and $\ln(ZMA(ZMA)^{SR})$ is not significantly above 3 at the 5% level. The distributions of log-realized volatilities are much closer to the normal distribution than those of realized volatilities. Thus, we use log-realized volatility as a dependent variable in the ARFIMA model (8), HAR model (9), ARFIMAX model (10) and HARX model (11).

To measure the performance of option pricing, we also use prices of the Nikkei 225 stock index options traded at the Osaka Securities Exchange. Nikkei 225 stock index options are European options and their maturities are the trading days previous to the second Friday every month. Considering theoretical option prices are with respect to a risk neutral measure, we assess the performance of option pricing using options which are most likely to be efficiently priced. For the Nikkei 225 stock index options, put options are traded more heavily than call options and the options with the maturity more than one month are not traded so much. Thus we concentrate on put options whose maturity is 30 days (29 days if the day when the maturity is 30 days is a weekend or holiday). On such days, we consider put options with different exercise prices whose bid and ask prices are both available at the same time between 14:00 and 15:00. For each option, we use the average of bid and ask prices at the same time closest to 15:00 as the market price at 15:00. The reason why we use the average of bid and ask prices instead of transaction prices is that transaction prices are subject to market microstructure noise due to bid-ask bounce (Campbell et al., 1997). We also exclude some kinds of put options which are not priced at the theoretical range from the lower bound at $P_T = \text{Max}(0, K\text{exp}(-r\tau) - S_T\text{exp}(-d\tau))$ to the upper bound at $P_T = K\text{exp}(-r\tau)$.

We estimate the ARFIMA(X) and HAR(X) models using 1200 daily realized volatilities up to the day before the options whose maturity is one month are traded, where the adjustment coefficient c defined by equation (5) is calculated using the same 1200 realized volatilities with 1200 daily returns. We also estimate ARCH type models using the same 1200 daily returns with risk-free rate and dividend. As mentioned, the daily returns are calculated as the log difference of closing prices. We use CD rate as a risk-free rate and fix the annual dividend rate as 0.5% following Nishina and Nabil (1997). The first date when options whose maturity is one month are traded is April 11, 2001. We first estimate the parameters in the ARFIMA(X), HAR(X) and ARCH type models using 1200 daily realized volatilities and returns up to April 10, 2001, where we calculate the adjustment coefficient c using the same 1200 daily realized volatilities and returns. Then, given the obtained parameter estimates, we calculate the put option prices on April 11, 2001 using CD rate and the Nikkei 225 index at 15:00 on that date. The next date when options whose maturity is one month are traded is May 9, 2001. We first estimate the parameters in the ARFIMA(X), HAR(X) and ARCH type models using 1200 daily realized volatilities and returns up to May 8, 2001, where we calculate the adjustment coefficient c using the same 1200 daily realized volatilities and returns. Then, given the obtained parameter estimates, we calculate the put option prices on May 9, 2001 using CD rate and the Nikkei 225 index at 15:00 on that date. We repeat this procedure up to September 2007.

Figure 2 plots the estimates of all parameters in all models for each of the above 78 iterations. Figure 2 (a) and (b) plot the estimates of parameters in the ARFIMA and ARFIMAX models using $RV(\text{1min})^{HL}$. The estimates of d in the ARFIMA and ARFIMAX models move around 0.5 and are above 0.5 in the latter half, indicating the long-memory and the possibility of non-stationarity of log-realized volatility. The estimates of μ_2 in the ARFIMAX model are positive for all periods, indicating the well-known phenomenon of a negative correlation between today's return and tomorrow's volatility. Figure 2 (c) and (d) plot the estimates of parameters in the HAR and HARX models using $RV(\text{1min})^{HL}$. The positive estimates of β_1, β_2, β_3 in the HAR and HARX models for all periods are consistent with the empirical results using S&P500 in Corsi (2009). The estimates of β_5 in the HARX model are positive, indicating the asymmetry in volatility. Figure 2 (e), (f) and (g) plot the estimates of parameters in ARCH type models using daily returns. The sum of the estimates of β and α in the GARCH model and the estimates of ϕ in the EGARCH model are close to 1 for all periods, indicating the well-phenomenon of volatility clustering. These models, however, do not allow for the long-memory of volatility. The estimates of d in the FIEGARCH model are more volatile than those of the ARFIMA(X) model. They move around 0.5 in the first half while they move up to 0.54 and down to 0 in the latter half. These results provide evidence that a structural change may occur during our sample period, but we leave it for future research. The estimates of θ in the EGARCH and FIEGARCH models are negative for all periods, indicating a negative correlation between today's return and tomorrow's volatility.

6 Results

To measure the performance of option pricing, we use four loss functions, MAE (Mean Absolute Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error) and RMSPE (Root Mean Square Percentage Error) defined as

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\tilde{P}_i - P_i\right|, \quad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{P}_i - P_i\right)^2},$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\tilde{P}_i - P_i}{P_i}\right|, \quad \text{RMSPE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\tilde{P}_i - P_i}{P_i}\right)^2}.$$

where N is the number of put options used for evaluating the performance, $\tilde{P}_i$ is the price of the ith put option calculated by each model and P_i is its market put price calculated as the average of bid and ask prices at the same time closest to 15:00. From the fact that the lowest market put price amounts to 1.5 yen which is calculated as the mid-point of the ask price at 2 yen and the bid price at 1 yen, any price $\tilde{P}_i$ less than the lowest price is approximated at 1.5 yen.

Following Bakshi et al. (1997), we classify put options into five categories such as DITM (deep-in-the-money), ITM (in-the-money), ATM (at-the-money), OTM (out-of-the-money) and DOTM (deep-out-of-the-money) using the moneyness which is the ratio of the underlying asset price over the exercise price. Table 4 shows this classification. We examine the performance in each category as well as in total.

Table 5 shows the values of loss functions for ARCH type models with daily returns, the ARFIMA (X) and HAR(X) models with $RV(\text{1min})^{HL}$ and the BS model. In total, the ARFIMAX model performs best for RMSPE and MAPE while the HARX model performs best for RMSE and MAE. The RMSE and MAE of the ARFIMAX model are, however, not so much different from those of the HARX model. In DOTM, ARFIMAX model performs best for RMSPE and MAPE while the FIEGARCH model performs best for the other loss functions. In OTM, the ARFIMAX model performs best for RMSE, RMSPE and MAPE while the ARFIMA model performs best for MAE. In ATM and ITM, either the ARFIMAX model or the HARX model performs best for all loss functions. In DITM, the GARCH model performs best for all loss functions. Although there are some exceptions depending on moneyness and loss function, we may conclude that the ARFIMAX model performs best.

Tables 6 and 7 show the values of loss functions for the ARFIMAX model with 30 different realized volatilities. Table 6 shows the result for the realized volatilities calculated simply by adding the squares of overnight and lunch-time returns instead of using the Hansen and Lunde (2005a) adjustment. In total and all moneyness, the loss functions of $RV(\text{1min})^{SR}$ have larger values than those of the other realized volatilities except $ZMA(ZMA)^{SR}$. This result is intuitive because $RV(\text{1min})^{SR}$ does not take account of microstructure noise at all. Table 7 shows the result for the realized volatilities calculated using the Hansen and Lunde (2005a) adjustment instead of adding the squares of overnight and lunch-time returns. In total and all moneyness, all loss functions in Table 7 are smaller than those in Table 6, indicating that the Hansen and Lunde (2005a) adjustment improves the performance of option pricing. It is also noteworthy that the performance of $RV(\text{1min})^{HL}$ is no longer bad. $RV(\text{1min})^{HL}$ performs best for RMSPE and MAPE in total, RMSE in OTM and RMSPE in DOTM. For the other moneyness and loss functions, $RV(\text{15min})^{HL}$, $ZMA(ZMA)^{HL}$, $BC(ZMA, ZMA)^{HL}$ and $FMTH(BR)^{HL}$, which take account of microstructure noise, perform best. This result means that the Hansen and Lunde (2005a) adjustment plays a role to remove not only the discretization noise included in the squares of the lunch-time and overnight returns but also the bias caused by the microstructure noise because the adjustment coefficient c is set such that the sample mean of realized volatility is equal to the sample variance of daily returns. We may conclude that if the the Hansen and Lunde (2005a) adjustment is used, the other methods for taking account of the microstructure noise do not necessarily improve the performance of option pricing. [6]

So far, we assumed risk neutrality. As explained in Section 4, Duan (1995) has proposed a method for GARCH option pricing relaxing this assumption. We also apply this method to the GARCH, EGARCH and FIEGARCH models. Table 8 shows the result. The values of loss functions using this method are not so much different from those assuming risk neutrality. This result means that the Duan (1995) method does not improve the performance of option pricing compared with assuming risk neutrality.

7 Conclusions

This article compares the performance of option pricing among the ARFIMA(X) and HAR(X) models with daily realized volatility and the ARCH models with daily returns. The main results are: (1) the ARFIMAX model with daily realized volatility performs best, (2) the Hansen and Lunde (2005a) adjustment without using overnight and lunch-time returns can improve performance, (3) if the Hansen and Lunde (2005a) adjustment, which also plays a role to remove the bias from the microstructure noise by setting the sample mean of realized volatility equal to the sample variance of daily returns, is used, the other methods for taking account of microstructure noise do not necessarily improve performance and (4) the Duan (1995) method does not improve performance compared with assuming risk neutrality.

Several extensions are possible. First, we did not consider jumps in returns. Barndorff-Nielsen and Shephard (2002a, 2004) have proposed a method for calculating realized volatility taking account of jumps. Andersen et al. (2007) show that the performance of forecasting future volatility is improved by removing significant jumps from realized volatility and adding significant jumps to the HAR model as an explanatory variable. It is interesting whether the performance of option pricing will also be improved by doing so. Second, Hansen et al. (2010) and Takahashi et al. (2009) have proposed to model daily returns and realized volatility jointly[7]. It is also interesting to apply their methods to option pricing.

[6] Bandi et al. (2008) compare the option pricing performance of the realized volatilities of the S&P 500 index. Their method is, however, different from ours as follows. (1) They compare the profits from the straddle trading strategy obtained by substituting the volarility forecasts from the ARFIMA model for realized volatility into the Black-Scholes option pricing formula. (2) They only analyze the performance of $RV(\text{5min})^{HL}$ to $FMTH(BR)^{HL}$, which are calculated using the Hansen and Lunde (2005a) adjustment, while we also analyze the performance of $RV(\text{1min})^{HL}$ and $RV(\text{1min})^{SR}$ to $FMTH(BR)^{SR}$, which are calculated by adding the lunch-time and overnight returns without using the Hansen and Lunde (2005a) adjustment. (3) They do not analyze ARCH-type models.

[7] Hansen et al. (2010) and Takahashi et al. (2009) extend ARCH type models and the stochastic volatility model respectively.

Appendix Integrated volatility estimators with microstructure noise

Here, we give a detailed review of various realized volatilities using the high-frequency returns employed in our analysis. Assume the i-th intraday return $r_{t-1+i/n}$ for day t contaminates with microstructure noise as follows

$$\begin{aligned} r_{t-1+i/n} &= p(t-1+i/n) - p(t-1+(i-1)/n) + \eta(t-1+i/n) - \eta(t-1+(i-1)/n) \\ &= p(t-1+i/n) - p(t-1+(i-1)/n) + e_{t-1+i/n}, \end{aligned} \tag{A.1}$$

where $e_{t-1+i/n} := \eta(t-1+i/n) - \eta(t-1+(i-1)/n)$ and η represents microstructure noise.

- Realized volatility with 1-, 5- and 15-minute returns, $RV(\text{1min})$, $RV(\text{5min})$ and $RV(\text{15min})$.

Without microstructure noise, it would be desirable to use intraday returns sampled at the highest frequencies. Since the highest frequencies available for the Nikkei 225 stock index is 1-minute, we first calculate realized volatility using 1-minute returns ($n = 270$), which is denoted as $RV(\text{1min})$. However, it may fail to satisfy the consistency condition when there is market microstructure noise as usually documented in real high-frequency data. Another classical approach is to use realized volatility constructed from intraday returns sampled at moderate frequencies rather than at the highest frequencies. This approach can partially offset the bias of the microstructure effect. In practice, researchers are necessarily forced to select a moderate sampling frequency. For example, it may be regarded as around those frequencies for which realized volatility signature plots under alternative sampling frequencies are leveled off. Evidence from previous studies suggests that it is optimal to use 5 to 30-minute return data. Hence, we employ $RV(\text{5min})$ and $RV(\text{15min})$ which are equal to the sum of squared 5- and 15-minute returns ($n = 54$ and 18), respectively.

- Optimally-sampled realized volatility, $RV(BR)$.

The selection of a moderate sampling frequency is important to get an accurate estimate of the integrated volatility because the noise-induced bias at high sampling frequencies can be traded off with the variance reduction obtained by high-frequency sampling. To take this trade off between the bias and variance into account, Bandi and Russell (2008) provide a theoretical justification for the choice of optimal sampling frequency in terms of the mean squared error (MSE) criterion. They derive the following approximated optimal number of observations n^* based on the minimization of MSE in a finite sample

$$n^* \approx \left[\frac{IQ}{\{\mathrm{E}(e^2)\}^2}\right]^{\frac{1}{3}}, \tag{A.2}$$

where IQ represents an integrated quarticity of the equilibrium price process ($IQ = \int_{t-1}^{t} \sigma^4(s)ds$). It is estimated by $\hat{IQ} = \frac{n}{3}\sum_{i=1}^{n} r_{t-1+i/n}^4$ (realized quarticity) with low frequency returns such as 15-

minute returns. Following the consistent estimator of noise moment as shown by Bandi and Russell (2008), $\mathrm{E}(e^2)$ can be estimated by $\hat{\mathrm{E}}(e^2) = \frac{1}{n}\sum_{i=1}^{n} r_{t-1+i/n}^2$ at the highest frequencies. Thus, the optimally-sampled realized volatility, $RV(BR)$, is equal to the realized volatility with the optimal number of observations calculated as $\hat{n}^* = \left[\hat{IQ}/(\hat{\mathrm{E}}\left(e^2\right))^2\right]^{1/3}$.

- The Bartlett-type kernel estimator in Barndorff-Nielsen et al. (2004) with a finite sample optimal number of autocovariances proposed by Bandi and Russell (2011), $BK(BR)$.

$RV(\text{1min})$, $RV(\text{5min})$, $RV(\text{15min})$ and $RV(BR)$ have the obvious drawback that they do not incorporate all data and whereby information is lost. The methods introduced here take advantage of the rich sources in all high-frequency data. The problem of estimating the integrated volatility under microstructure noise is similar to the autocorrelation corrections that are used in the long-run variance estimation in stationary time-series (Newey and West, 1987; Andrews, 1991). So it is natural to consider kernel-based estimators of integrated volatility under microstructure noise. The literature includes the earlier study by Zhou (1996) who proposes a particular kernel estimator which incorporates the first-order autocovariance. Barndorff-Nielsen et al. (2004) derive kernel-based estimators that are far more precise than that of Zhou (1996). They examine the Bartlett-type kernel estimator defined as

$$BK = \left(\frac{n-1}{n}\frac{H-1}{H}\right)\gamma_0 + 2\sum_{h=1}^{H}\left(\frac{H-h}{H}\right)\gamma_h, \tag{A.3}$$

where $\gamma_h = \sum_{i=1}^{n-h} r_{t-1+i/n} r_{t-1+(i+h)/n}$ is the h-th autocovariance of intraday returns and γ_0 is equal to realized volatility using returns sampled at the highest frequencies. This estimator weights the realized volatility and the H-th return autocovariances by Bartlett weights. The optimal number of autocovariances is given by the minimization of MSE of the estimator in finite sample (see equation 7 to 10 in Bandi and Russell, 2011 for exact MSE minimization expressions). There is a convenient rule-of-thumb for choosing H in practice as proposed in Bandi and Russell (2011). The expression is obtained as

$$H^* \approx \left(\frac{3IV^2}{2n^2IQ}\right)^{\frac{1}{3}} n, \tag{A.4}$$

where IV denotes integrated volatility. IV and IQ are estimated using realized volatility and realized quarticity with lower frequency returns such as 15-minute returns. Hence, BK with a finite sample optimal number of autocovariances H^* leads to $BK(BR)$.

- The two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang et al. (2005), $ZMA(ZMA)$.

Zhang et al. (2005) propose a two-scale or subsampling estimator in the spirit of the estimation of the long-rum variance studied by Carlstein (1986). Denote the original grid of observation times as

$\Psi = \{t-1, t-1+1/n, t-1+2/n \ldots, t\}$. Consider Ψ is partitioned into $\tilde{K}$ nonoverlapping subgrids, $\Psi_{\tilde{K}}^{(j)}$, $j = 1, \ldots, \tilde{K}$, for example, the first sub-grid starts at $t-1$ and takes every $\tilde{K}-$th arrival time ($\Psi_{\tilde{K}}^{(1)} = \{t-1, t-1+\tilde{K}/n, t-1+2\tilde{K}/n \ldots\}$), and the second sub-grid starts at $t-1+1/n$ and takes every $\tilde{K}-$th arrival time ($\Psi_{\tilde{K}}^{(2)} = \{t-1+1/n, t-1+(1+\tilde{K})/n, t-1+(1+2\tilde{K})/n \ldots\}$). Then, the realized volatility for the subgrid $\Psi_{\tilde{K}}^{(j)}$ is defined as

$$RV_{\tilde{K}}^{(j)} = \sum_{i=1}^{n_j} r_{t-1+(j-1+i\tilde{K})/n}^2, \tag{A.5}$$

where $r_{t-1+(j-1+i\tilde{K})/n}$ is subsampling return between transaction prices at times $t-1+(j-1+i\tilde{K})/n$ and $t-1+(j-1+(i-1)\tilde{K})/n$. The two-scale estimator in Zhang et al. (2005) is given by

$$ZMA = (1/\tilde{K}) \sum_{j=1}^{\tilde{K}} RV_{\tilde{K}}^{(j)} - (\bar{n}/n) RV, \tag{A.6}$$

where $\bar{n} = (n - \tilde{K} + 1)/\tilde{K}$ and RV is the realized volatility for the full grid Ψ. The second term corrects the bias in the first term. The asymptotic optimal number of subsamples $\tilde{K}^*(ZMA)$ derived by minimizing the estimator's asymptotic variance is given by

$$\tilde{K}^*(ZMA) = \left[\frac{3\left\{\mathrm{E}(e^2)\right\}^2}{IQ} \right]^{1/3} n^{2/3}. \tag{A.7}$$

IQ and $\mathrm{E}(e^2)$ are estimated by realized quarticity with 15-minute returns and $\hat{\mathrm{E}}(e^2) = \frac{1}{n}\sum_{i=1}^{n} r_{t-1+i/n}^2$ at the highest frequencies, respectively. Thus, $ZMA(ZMA)$ is equal to ZMA with $\tilde{K}^*(ZMA)$.

- The two-scale estimator in Zhang et al. (2005) with a finite sample optimal number of subsamples proposed by Bandi and Russell (2011), $ZMA(BR)$.

Barndorff-Nielsen et al. (2004) show that ZMA in (A.6) can be written as follows

$$ZMA = \left(1 - \frac{n-H+1}{nH}\right)\gamma_0 + 2\sum_{h=1}^{H}\left(\frac{H-h}{H}\right)\gamma_h - \frac{1}{H}\theta_H, \tag{A.8}$$

where $\theta_1 = 0$, and $\theta_H = \theta_{H-1} + (r_{t-1+1/n} + \cdots + r_{t-1+(H-1)/n})^2 + (r_{t-1+(n-H+2)/n} + \cdots + r_t)^2$ for $H \geq 2$. The third term guarantees consistency of ZMA and differentiates ZMA from the inconsistent BK. This equation implies the two-scale estimator in Zhang et al. (2005) is almost identical to the modified Bartlett kernel estimator. Bandi and Russell (2011) additionally show that the finite sample MSEs of BK and ZMA are very similar in practice. Hence, the ZMA with $\tilde{K} = H^*$ in (A.4) corresponds to $ZMA(BR)$.

- The bias-corrected two-scale estimator in Zhang et al. (2005) with an asymptotically optimal number of subsamples proposed by Zhang et al. (2005), $BC(ZMA, ZMA)$.

The two-scale estimator ZMA has a finite sample bias as shown in Zhang et al. (2005) who provide the approximate correction for this bias. On the other hand, Bandi and Russell (2011) report the exact bias-correction form. Following a suggestion by Bandi and Russell (2011), the bias-corrected estimator is defined as

$$\begin{aligned} BC(ZMA) &= c(\tilde{K}, n)ZMA, \\ c(\tilde{K}, n) &= \left(\frac{\tilde{K}n - 1 + 2\tilde{K} - \tilde{K}^2 - n}{\tilde{K}n} \right)^{-1}. \end{aligned} \tag{A.9}$$

Since $BC(ZMA)$ is asymptotically equivalent to ZMA, the asymptotically optimal number of subsamples is given by $\tilde{K}^*(ZMA)$. Thus, $BC(ZMA)$ with $\tilde{K}^*(ZMA)$ can be described by $BC(ZMA,$ $ZMA)$.

- The bias-corrected two-scale estimator in Zhang et al. (2005) with a finite sample optimal number of subsamples proposed by Bandi and Russell (2011), $BC(ZMA, BR)$.

Since $BC(ZMA)$ is unbiased in a finite sample, the optimal number of subsamples is provided by minimizing the finite sample variance of $BC(ZMA)$. Bandi and Russell (2008, 2011) show that the optimal number of subsamples is defined as

$$\tilde{K}^*(BR) = \underset{0<\tilde{K}/n\leq 1/2}{\arg\min} \left[\text{Var}\left(BC(ZMA) \right) \right] = \underset{0<\tilde{K}/n\leq 1/2}{\arg\min} \left[\left\{ c(\tilde{K}, n) \right\}^2 \ \text{Var}(ZMA) \right], \tag{A.10}$$

where, if $\tilde{K}/n \leq 1/2$,

$$\begin{aligned} \text{Var}(ZMA) =& (-4\sigma_\eta^4 - 8IV\sigma_\eta^2)\frac{1}{n} + \left(-4\sigma_\eta^4 - 8\sigma_\eta^2 IV + \frac{13}{3}IQ + \frac{79}{3}IV^2 \right) \frac{1}{n^2} + (2IQ + 8IV^2)\frac{1}{n^3} \\ & - \frac{1}{3}(IQ + IV^2)\frac{\tilde{K}^2}{n^2} + \left(-\frac{IV^2}{3n} - \frac{4IV^2}{n^2} + \frac{4}{3}IQ \right) \frac{\tilde{K}}{n} \\ & + \left[-\frac{4}{n^4}(IQ + IV^2) + \left(\frac{8\sigma_\eta^4 + 16\sigma_\eta^2 IV - 8IQ - \frac{56}{3}IV^2}{n^3} \right) \right. \\ & \left. + \left(\frac{24\sigma_\eta^2 IV - \frac{10}{3}IQ + 8\sigma_\eta^4}{n^2} \right) + \left(\frac{-8\sigma_\eta^4 + 8\sigma_\eta^2 IV}{n} \right) \right] \frac{n}{\tilde{K}} \\ & + \left[\frac{2}{n^5}IQ + \left(\frac{-4\sigma_\eta^4 - 8\sigma_\eta^2 IV + 4IQ - 8IV^2}{n^4} \right) \right. \\ & \left. + \left(\frac{-4\sigma_\eta^4 - 16\sigma_\eta^2 IV + 2IQ}{n^3} \right) + \left(\frac{8\sigma_\eta^4 - 8\sigma_\eta^2 IV}{n^2} \right) + \frac{8}{n}\sigma_\eta^4 \right] \frac{n^2}{\tilde{K}^2}, \end{aligned} \tag{A.11}$$

where σ_η^2 represents a variance of microstructure noise η and is estimated by $\hat{\sigma}_\eta^2 = \frac{1}{2n}\sum_{i=1}^n r_{t-1+i/n}^2$ at the highest frequencies. Hence, $BC(ZMA)$ with $\tilde{K}^*(BR)$ leads to $BC(ZMA, BR)$.

- The flat-top Bartlett kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. (2008), $FBK(BNHLS)$.

Barndorff-Nielsen et al. (2008) examine the following unbiased flat-top kernel type estimator (called the realized kernel)

$$RK = \gamma_0 + \sum_{h=1}^{H} k(x)(\gamma_h + \gamma_{-h}), \tag{A.12}$$

where $\gamma_h = \sum_{i=1}^{n} r_{t-1+i/n} r_{t-1+(i-h)/n}$ with $h = -H, \cdots, H$ and the non-stochastic $k(x) \in [0,1]$ for $x = \frac{h-1}{H}$ is a weight function. The flat-top Bartlett kernel estimator is equivalent to RK in case where $k(x) = 1 - x$. For this class of kernels, Barndorff-Nielsen et al. (2008) show that the asymptotic distribution of $RK - IV$ is mixed normal with zero mean and rate of convergence $n^{1/6}$ when $H = cn^{2/3}$ where c is a constant. Then, the asymptotically optimal value of c which minimizes the asymptotic variance is given by

$$c^* \approx 2.28\zeta^{\frac{4}{3}}, \tag{A.13}$$

where $\zeta^2 = \sigma_\eta^2/\sqrt{IQ}$. Hence, RK with $k(x) = 1-x$ and $H = c^* n^{2/3}$ corresponds to $FBK(BNHLS)$.

- The flat-top cubic kernel estimator and the flat-top modified Tukey-Hanning kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. (2008), $FCK(BNHLS)$ and $FMTH(BNHLS)$.

The estimators based on the cubic kernel and the modified Tukey-Hanning kernel are equivalent to RK with $k(x) = 1 - 3x^2 + 2x^3$ and $k(x) = \{1 - \cos\pi(1-x)^2\}/2$, respectively. When $H = c\zeta n^{1/2}$, RK for this class of kernels is consistent at the rate of convergence $n^{1/4}$ as shown in Barndorff-Nielsen et al. (2008). The asymptotically optimal value of c is expressed as

$$c^* = \sqrt{\rho \frac{k_\bullet^{1,1}}{k_\bullet^{0,0}} \left\{ 1 + \sqrt{1 + \frac{3k_\bullet^{0,0} k_\bullet^{2,2}}{\rho (k_\bullet^{1,1})^2}} \right\}}, \tag{A.14}$$

where $\rho = IV/\sqrt{IQ}$, $k_\bullet^{0,0} = \int_0^1 k(x)^2 dx$, $k_\bullet^{1,1} = \int_0^1 k'(x)^2 dx$ and $k_\bullet^{2,2} = \int_0^1 k''(x)^2 dx$, where the primes represent derivatives. The values of ($k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2}$) amount to $(k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2}) = (0.371, 1.20, 12.0)$ for the cubic kernel and $(k_\bullet^{0,0}, k_\bullet^{1,1}, k_\bullet^{2,2}) = (0.219, 1.71, 41.7)$ for the modified Tukey-Hanning kernel. We define $FCK(BNHLS)$ and $FMTH(BNHLS)$ as RK with $H = c^*\zeta n^{1/2}$ at $k(x) = 1 - 3x^2 + 2x^3$ and $k(x) = \{1 - \cos\pi(1-x)^2\}/2$.

- The flat-top Bartlett kernel estimator, the flat-top cubic kernel estimator and the flat-top modified Tukey-Hanning kernel estimator with a finite sample optimal number of autocovariances proposed by Bandi and Russell (2011), $FBK(BR)$, $FCK(BR)$ and $FMTH(BR)$.

Bandi and Russell (2011) provide an alternative way to choose the number of autocovariances in finite samples. Denote H as δn with $0 < \delta \leq 1$. The optimal value of δ is defined in Theorem 3 of Bandi and Russell (2011) as follows

$$\delta^* = \underset{0<\delta\leq 1}{\arg\min} \left[(\text{bias}(RK))^2 + \text{Var}(RK) \right], \tag{A.15}$$

where bias$(RK) = 0$ and

$$\text{Var}(RK) = \frac{IQ}{n} \omega^{\mathrm{T}} \Omega_1 \omega + 4\sigma_\eta^4 n(\omega^{\mathrm{T}} \Omega_2 \omega) + 4\sigma_\eta^4 (\omega^{\mathrm{T}} \Omega_3 \omega) + (2\sigma_\eta^2 IV) 4(\omega^{\mathrm{T}} \Omega_4 \omega), \tag{A.16}$$

with $\omega = \left(1, 1, k\left(\frac{1}{\delta n}\right), \cdots, k\left(\frac{\delta n - 1}{\delta n}\right)\right)^{\mathrm{T}}$ and Ω_a $a = 1, \cdots, 4$ are ($\delta n + 1, \delta n + 1$) square matrices. For $j \leq \delta n$, the matrices Ω_1 and Ω_4 are defined as

$$\begin{aligned}
&\Omega_1[1,1] = 2, \quad \Omega_1[1+j, 1+j] = 4, \\
&\Omega_4[1,1] = 1, \quad \Omega_4[2,1] = -1, \quad \Omega_4[1,2] = -1, \quad \Omega_4[2,2] = 2, \\
&\Omega_4[1+j, 1+j] = 2, \quad \Omega_4[1+j, j] = -1, \quad \Omega_4[j, j+1] = -1,
\end{aligned} \tag{A.17}$$

and zeros everywhere else. For $j \leq \delta n - 1$, the matrices Ω_2 and Ω_3 are defined as

$$\begin{aligned}
&\Omega_2[1,1] = 3, \quad \Omega_2[1,2] = -4, \quad \Omega_2[2,1] = -4, \quad \Omega_2[2,2] = 7, \\
&\Omega_2[2+j, 2+j] = 6, \quad \Omega_2[2+j, 1+j] = -4, \quad \Omega_2[1+j, 2+j] = -4, \quad \Omega_2[2+j, j] = 1, \\
&\Omega_2[j, 2+j] = 1, \quad \Omega_3[1,1] = -1, \quad \Omega_3[1,2] = 2, \quad \Omega_3[2,1] = 2, \quad \Omega_3[2,2] = -4.5, \\
&\Omega_3[j+2, j+2] = -3(j+1) - 1, \quad \Omega_3[2+j, 1+j] = 2(j+1), \quad \Omega_3[1+j, 2+j] = 2(j+1), \\
&\Omega_3[2+j, j] = -(j+1)/2, \quad \Omega_3[j, 2+j] = -(j+1)/2,
\end{aligned} \tag{A.18}$$

and zeros everywhere else. Thus, RK with $H = \delta^* n$ for the Bartlett kernel, cubic kernel and modified Tukey-Hanning kernel leads to $FBK(BR)$, $FCK(BR)$ and $FMTH(BR)$, respectively.

References

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005), "How often to sample a continuous-time process in the presence of market microstructure noise," *Review of Financial Studies*, Vol.18, No.2, pp.351–416.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007), "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," *Review of Economics and Statistics*, Vol.89, No.4, pp.701–720.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. and Ebens, H. (2001), "The distribution of realized stock return volatility," *Journal of Financial Economics*, Vol.61, No.1, pp.43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), "Modeling and forecasting realized volatility," *Econometrica*, Vol.71, No.2, pp.579–625.

Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, Vol.59, No.3, 817-858.

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, Vol.74, No.1, pp.3–30.

Bakshi, G., Cao, C. and Chen, Z. (1997), "Empirical performance of alternative option pricing models," *Journal of Finance*, Vol.52, No.5, pp.2003–2049.

Bandi, F. M. and Russell, J. R. (2006), "Separating microstructure noise from volatility," *Journal of Financial Economics*, Vol.79, No.3, pp.655–692.

Bandi, F. M. and Russell, J. R. (2008), "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, Vol.75, No.2, pp.339–369.

Bandi, F. M. and Russell, J. R. (2011), "Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations," *Journal of Econometrics*, Vol.160, No.1, pp.145–159.

Bandi, F. M., Russell, J. R. and Yang, C. (2008), "Realized volatility forecasting and option pricing," *Journal of Econometrics*, Vol.147, No.1, pp.34–46.

Barndorff-Nielsen, O. E., Nielsen, B., Shephard, N. and Ysusi, C. (2004), "Measuring and forecasting financial variability using realized variance," in A. C. Harvey, S. J. Koopman and N. Shephard, (eds), *State Space and Unobserved Components Models: Theory and Applications*, Cambridge: Cambridge University Press, pp.205–235.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), "Non-Gaussian OU based models and some of their uses in financial economics (with Discussions)," *Journal of the Royal Statistical Society, Series B*, Vol.63, No.2, pp.167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a), "Estimating quadratic variation using realized volatility," *Journal of Applied Econometrics*, Vol.17, No.5, pp.457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society, Series B*, Vol.64, No.2, pp.253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004), "Power and bipower variation with stochastic volatility and jumps (with Discussions)," *Journal of Financial Econometrics*, Vol.2, No.1, pp.1–48.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2004), "Regular and modified kernel-based estimators of integrated variance: The case with independent noise?" Economics Papers 2004-W28, Economics Group, Nuffield College, University of Oxford.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008), "Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, Vol.76, No.6, pp.1481-1536.

Beran, J. (1994), *Statistics for Long-Memory Processes*, New York: Chapman & Hall.

Beran, J. (1995), "Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models," *Journal of the Royal Statistical Society, Series B*, Vol.57, No.4, pp.659–672.

Black, F. and Scholes, M. (1973), "The pricing of options and corporate liabilities," *Journal of Political Economy*, Vol.81, No.3, pp.637–654.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, Vol.31, No.3, pp.307–327.

Bollerslev, T. and Mikkelsen, H. O. (1996), "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, Vol.73, No.1, pp.151–184.

Bollerslev, T. and Mikkelsen, H. O. (1999), "Long-term equity anticipation securities and stock market volatility dynamics," *Journal of Econometrics*, Vol.92, No.1, pp.75–99.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton: Princeton University Press.

Carlstein, E. (1986), "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *Annals of Statistics*, Vol.14, No.3, 1171-1179.

Clements, M. P., Galvão, A. B. and Kim, J. H. (2008), "Quantile forecasts of daily exchange rate returns from forecasts of realized volatility," *Journal of Empirical Finance*, Vol.15, No.4, pp.729–750.

Conrad, C. and Haag, B. R. (2006), "Inequality constraints in the fractionally integrate GARCH model," *Journal of Financial Econometrics*, Vol.4, No.3, pp.413–449.

Corsi, F. (2009), "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, Vol.7, No.2, pp.174–196.

Diebold, F. X. (1988), *Empirical Modeling of Exchange Rate Dynamics*, Berlin: Springer-Verlag.

Duan, J.-C. (1995), "The GARCH option pricing model," *Mathematical Finance*, Vol.5, No.1, pp.13–32.

Duan, J.-C. and Shimonato, J.-G. (1998), "Empirical Martingale simulation for asset prices," *Management Science*, Vol.44, No.9, pp.1218–1233.

Engle, R. F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, Vol.50, No.4, pp.987–1007.

Giot, P. and Laurent, S. (2004), "Modelling daily Value-at-Risk using realized volatility and ARCH type models," *Journal of Empirical Finance*, Vol.11, No.3, pp.379–398.

Hansen, P. R., Huang, Z. and Shek, H. (2010), "Realized GARCH: A joint model of returns and realized measures of volatility," *Journal of Applied Econometrics*, forthcoming.

Hansen, P. R. and Lunde, A. (2005a), "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?" *Journal of Applied Econometrics*, Vol.20, No.7, pp.873–889.

Hansen, P. R. and Lunde, A. (2005b), "A realized variance for the whole day based on intermittent high-frequency data," *Journal of Financial Econometrics*, Vol.3, No.4, pp.525–554.

Hansen, P. R. and Lunde, A. (2006), "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, Vol.24, No.2, pp.127-161.

Harris, L. (1990), "Estimation of stock price variances and serial covariances from discrete observations," *Journal of Financial and Quantitative Analysis*, Vol.25, No.3, pp.291–306.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005), "Forecasting daily variability of the S&P 100 stock index using historical, realized and implied volatility measurements," *Journal of Empirical Finance*, Vol.12, No.3, pp.445–475.

Kunitomo, N. and Sato, S. (2008), "Separating information maximum likelihood estimation of realized volatility and covariance with micro-market noise," Discussion Paper CIRJE-F-581, Graduate School of Economics, University of Tokyo.

Martens, M. (2002), "Measuring and forecasting S&P 500 index-futures volatility using high-Frequency data," *Journal of Futures Markets*, Vol.22, No.6, pp.497–518.

Meddahi, N. (2002), "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, Vol.17, No.5, pp.479–508.

Nagakura, D. and Watanabe, T. (2010), "A state space approach to estimating integrated variance under the existence of microstructure noise," Global COE Hi-Stat Discussion Paper Series 115, Hitotsubashi University.

Nelson, D. B. (1991), "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, Vol.59, No.2, pp.247–370.

Newey, W. K. and West, K. D. (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, Vol.55, No.3, pp.703-708.

Nishina, K. and Nabil, M. N. (1997), "Return dynamics of Japanese stock index options," *Japanese Economic Review*, Vol.48, No.1, pp.43–64.

Oya, K. (2011), "Bias corrected realized variance under dependent microstructure noise," *Mathematics and Computers in Simulation*, Vol.81, No.7, pp.1290–1298.

Roll, R. (1984), "A simple implicit measure of the effective bid-ask spread in an efficient market," *Journal of Finance*, Vol.39, No.4, pp.1127–1139.

Schoffer, O. (2003), "HY-A-PARCH: A stationary A-PARCH model with long memory," Technical Report 4003, Fachbereich Statitik, Universität Dortmunt.

Takahashi, M., Omori, Y. and Watanabe, T. (2009), "Estimating stochastic volatility models using daily returns and realized volatility simultaneously," *Computational Statistics & Data Analysis*, Vol.53, No.6, pp.2404–2426.

Taylor, S. J. (2001), "Consequences for option pricing for a long memory in volatility," Working Paper 2001/017, Lancaster University Management School.

Tse, Y. K. (1998), "The conditional heteroskedasticity of the Yen-Dollar exchange rate," *Journal of Applied Econometrics*, Vol.13, No.1, pp.49–55.

Ubukata, M. and Oya, K. (2009), "Estimation and Testing for Dependence in Market Microstructure Noise," *Journal of Financial Econometrics*, Vol.7, No.2, pp.106–151.

Zhang, L. (2006), "Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach," *Bernoulli*, Vol.12, No.6, pp.1019-1043.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005), "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, Vol.100, No.472, pp.1394–1411.

Zhou, B. (1996), “High-frequency data and volatility in foreign-exchange rates,” *Journal of Business & Economic Statistics*, Vol.14, No.1, pp.45–52.

Table 1: List of daily realized volatilities

Methods of calculating realized volatility	Methods of daily adjustment: Hansen and Lunde (2005a) adjustment	Methods of daily adjustment: Squares of overnight and lunch-time returns
Realized volatility with returns sampled at the highest frequencies	$RV(\text{1min})^{HL}$	$RV(\text{1min})^{SR}$
Realized volatility with 5-minute returns	$RV(\text{5min})^{HL}$	$RV(\text{5min})^{SR}$
Realized volatility with 15-minute returns	$RV(\text{15min})^{HL}$	$RV(\text{15min})^{SR}$
Optimally-sampled realized volatility as proposed in Bandi and Russell (2008)	$RV(BR)^{HL}$	$RV(BR)^{SR}$
Bartlett kernel estimator in Barndorff-Nielsen et al. (2004) with a finite sample optimal number of autocovariances in Bandi and Russell (2011)	$BK(BR)^{HL}$	$BK(BR)^{SR}$
Two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang et al. (2005)	$ZMA(ZMA)^{HL}$	$ZMA(ZMA)^{SR}$
Two-scale estimator with a finite sample optimal number of subsamples proposed by Bandi and Russell (2011)	$ZMA(BR)^{HL}$	$ZMA(BR)^{SR}$
Bias-corrected two-scale estimator with an asymptotically optimal number of subsamples proposed by Zhang et al. (2005)	$BC(ZMA, ZMA)^{HL}$	$BC(ZMA, ZMA)^{SR}$
Bias-corrected two-scale estimator with a finite sample optimal number of subsamples proposed by Bandi and Russell (2011)	$BC(ZMA, BR)^{HL}$	$BC(ZMA, BR)^{SR}$
Flat-top Bartlett kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. (2008)	$FBK(BNHLS)^{HL}$	$FBK(BNHLS)^{SR}$
Flat-top Bartlett kernel estimator with a finite sample optimal number of autocovariances proposed by Bandi and Russell (2011)	$FCK(BNHLS)^{HL}$	$FCK(BNHLS)^{SR}$
Flat-top cubic kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. (2008)	$FMTH(BNHLS)^{HL}$	$FMTH(BNHLS)^{SR}$
Flat-top cubic kernel estimator with a finite sample optimal number of autocovariances proposed by Bandi and Russell (2011)	$FBK(BR)^{HL}$	$FBK(BR)^{SR}$
Flat-top modified Tukey-Hanning kernel estimator with an asymptotically optimal number of autocovariances proposed by Barndorff-Nielsen et al. (2008)	$FCK(BR)^{HL}$	$FCK(BR)^{SR}$
Flat-top modified Tukey-Hanning kernel estimator with a finite sample optimal number of autocovariances Bandi and Russell (2011)	$FMTH(BR)^{HL}$	$FMTH(BR)^{SR}$

Table 2: Descriptive statistics of daily returns

Mean	-0.0095
	(0.0270)
Standard Deviation	1.4261
Min	-7.2340
Max	7.6605
Skewness	-0.0616
	(0.0464)
Kurtosis	4.9003
	(0.0927)
LB(10)	18.69

The numbers in parentheses are standard errors. LB(10) is the Ljung-Box statistic adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelations up to 10 lags.

Table 3: Descriptive statistics of daily realized volatilities

(a) Daily realized volatilities

	Mean		Std	Min	Max	Skewness		Kurtosis		LB(10)
$RV(\text{1min})^{HL}$	1.9921	(0.0302)	1.5963	0.1634	25.9768	3.6327	(0.0464)	34.4942	(0.0927)	1832.50
$RV(\text{5min})^{HL}$	1.9432	(0.0333)	1.7588	0.0819	28.6264	4.2576	(0.0464)	40.7168	(0.0927)	1467.57
$RV(\text{15min})^{HL}$	1.8471	(0.0400)	2.1114	0.0635	35.9133	5.2932	(0.0464)	54.5825	(0.0927)	822.70
$RV(BR)^{HL}$	1.9565	(0.0327)	1.7271	0.0864	30.4920	4.0137	(0.0464)	40.1922	(0.0927)	1634.19
$BK(BR)^{HL}$	1.9270	(0.0330)	1.7420	0.1049	23.7380	3.7378	(0.0464)	28.5100	(0.0927)	1451.33
$ZMA(ZMA)^{HL}$	1.9489	(0.0344)	1.8156	0.0918	27.9262	3.5331	(0.0464)	28.0163	(0.0927)	1257.45
$ZMA(BR)^{HL}$	1.9380	(0.0350)	1.8513	0.0979	26.7405	4.1267	(0.0464)	34.1826	(0.0927)	1303.66
$BC(ZMA, ZMA)^{HL}$	2.0492	(0.0324)	1.7094	0.1144	26.0474	3.4281	(0.0464)	29.0140	(0.0927)	1792.37
$BC(ZMA, BR)^{HL}$	1.9838	(0.0331)	1.7485	0.1087	23.3101	3.5065	(0.0464)	26.3526	(0.0927)	1617.98
$FBK(BNHLS)^{HL}$	2.0024	(0.0331)	1.7496	0.0940	24.6444	3.4001	(0.0464)	25.3333	(0.0927)	1704.18
$FBK(BR)^{HL}$	1.9992	(0.0331)	1.7485	0.1027	24.4031	3.3623	(0.0464)	24.7373	(0.0927)	1705.63
$FCK(BNHLS)^{HL}$	1.9916	(0.0337)	1.7808	0.0989	23.8094	3.5342	(0.0464)	26.5715	(0.0927)	1606.44
$FCK(BR)^{HL}$	2.0001	(0.0331)	1.7475	0.1001	24.4053	3.3591	(0.0464)	24.7422	(0.0927)	1714.44
$FMTH(BNHLS)^{HL}$	1.9987	(0.0341)	1.8030	0.1017	24.8023	3.5003	(0.0464)	25.9096	(0.0927)	1578.92
$FMTH(BR)^{HL}$	2.0036	(0.0338)	1.7883	0.0978	24.8909	3.4800	(0.0464)	26.0213	(0.0927)	1638.10
$RV(\text{1min})^{SR}$	1.1362	(0.0163)	0.8608	0.0745	11.5340	2.8392	(0.0464)	21.3989	(0.0927)	1731.78
$RV(\text{5min})^{SR}$	1.4284	(0.0225)	1.1889	0.0432	17.6924	3.5080	(0.0464)	29.3557	(0.0927)	1643.41
$RV(\text{15min})^{SR}$	1.5417	(0.0306)	1.6183	0.0452	26.6608	4.9384	(0.0464)	49.6731	(0.0927)	926.81
$RV(BR)^{SR}$	1.3780	(0.0210)	1.1122	0.0584	17.6924	3.2715	(0.0464)	28.0191	(0.0927)	1775.28
$BK(BR)^{SR}$	1.3890	(0.0218)	1.1503	0.0691	14.2190	3.0944	(0.0464)	20.7172	(0.0927)	1661.81
$ZMA(ZMA)^{SR}$	0.9183	(0.0148)	0.7830	0.0277	11.5340	2.9299	(0.0464)	21.7484	(0.0927)	1301.80
$ZMA(BR)^{SR}$	1.2602	(0.0205)	1.0810	0.0442	14.0378	3.3175	(0.0464)	23.6746	(0.0927)	1564.40
$BC(ZMA, ZMA)^{SR}$	1.2315	(0.0179)	0.9473	0.0554	11.9358	2.7099	(0.0464)	18.8033	(0.0927)	1911.19
$BC(ZMA, BR)^{SR}$	1.3778	(0.0210)	1.1120	0.0570	13.2960	2.8821	(0.0464)	18.7785	(0.0927)	1809.61
$FBK(BNHLS)^{SR}$	1.3433	(0.0203)	1.0748	0.0487	13.3620	2.7856	(0.0464)	17.8970	(0.0927)	1890.93
$FBK(BR)^{SR}$	1.3504	(0.0205)	1.0818	0.0533	13.3620	2.7732	(0.0464)	17.7268	(0.0927)	1879.32
$FCK(BNHLS)^{SR}$	1.3694	(0.0211)	1.1154	0.0521	13.3620	2.8910	(0.0464)	18.7753	(0.0927)	1817.13
$FCK(BR)^{SR}$	1.3508	(0.0205)	1.0813	0.0521	13.3620	2.7695	(0.0464)	17.7108	(0.0927)	1891.13
$FMTH(BNHLS)^{SR}$	1.3303	(0.0206)	1.0884	0.0519	13.3336	2.8674	(0.0464)	18.3983	(0.0927)	1797.91
$FMTH(BR)^{SR}$	1.3316	(0.0205)	1.0834	0.0501	13.3336	2.8447	(0.0464)	18.3383	(0.0927)	1836.58

The numbers in parentheses are standard errors. LB(10) is the Ljung-Box statistic adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelations up to 10 lags.

(b) Log daily realized volatilities

	Mean		Std	Min	Max	Skewness		Kurtosis		LB(10)
$\ln(RV(1\text{min})^{HL})$	0.4379	(0.0137)	0.7230	−1.8116	3.2572	−0.1904	(0.0464)	3.0401	(0.0927)	6297.57
$\ln(RV(5\text{min})^{HL})$	0.3765	(0.0145)	0.7675	−2.5028	3.3543	−0.1674	(0.0464)	3.3543	(0.0927)	4990.64
$\ln(RV(15\text{min})^{HL})$	0.2375	(0.0162)	0.8554	−2.7563	3.5811	0.0143	(0.0464)	3.4068	(0.0927)	3945.17
$\ln(RV(BR)^{HL})$	0.3804	(0.0148)	0.7805	−2.4487	3.4175	−0.2455	(0.0464)	3.3095	(0.0927)	5090.36
$\ln(BK(BR)^{HL})$	0.3681	(0.0144)	0.7632	−2.2548	3.1671	−0.1109	(0.0464)	3.2900	(0.0927)	5051.03
$\ln(ZMA(ZMA)^{HL})$	0.3516	(0.0152)	0.8018	−2.3882	3.3296	−0.1073	(0.0464)	3.2190	(0.0927)	4562.03
$\ln(ZMA(BR)^{HL})$	0.3552	(0.0149)	0.7848	−2.3243	3.2862	−0.0906	(0.0464)	3.3077	(0.0927)	4867.34
$\ln(BC(ZMA,ZMA)^{HL})$	0.4395	(0.0146)	0.7691	−2.1680	3.2599	−0.2957	(0.0464)	3.2360	(0.0927)	5495.76
$\ln(BC(ZMA,BR)^{HL})$	0.3979	(0.0146)	0.7691	−2.2195	3.1489	−0.1779	(0.0464)	3.2967	(0.0927)	5063.47
$\ln(FBK(BNHLS)^{HL})$	0.4071	(0.0146)	0.7712	−2.3644	3.2045	−0.1934	(0.0464)	3.2797	(0.0927)	5170.08
$\ln(FBK(BR)^{HL})$	0.4041	(0.0146)	0.7730	−2.2764	3.1947	−0.1872	(0.0464)	3.2484	(0.0927)	5183.89
$\ln(FCK(BNHLS)^{HL})$	0.3950	(0.0147)	0.7779	−2.3141	3.1701	−0.1704	(0.0464)	3.2675	(0.0927)	5085.94
$\ln(FCK(BR)^{HL})$	0.4049	(0.0146)	0.7726	−2.3013	3.1948	−0.1896	(0.0464)	3.2596	(0.0927)	5160.47
$\ln(FMTH(BNHLS)^{HL})$	0.3938	(0.0148)	0.7834	−2.2858	3.2109	−0.1542	(0.0464)	3.2270	(0.0927)	5082.26
$\ln(FMTH(BR)^{HL})$	0.4000	(0.0148)	0.7800	−2.3250	3.2145	−0.1765	(0.0464)	3.2717	(0.0927)	5128.36
$\ln(RV(1\text{min})^{SR})$	−0.1099	(0.0133)	0.7040	−2.5963	2.4453	−0.2092	(0.0464)	3.1449	(0.0927)	4730.86
$\ln(RV(5\text{min})^{SR})$	0.0924	(0.0140)	0.7389	−3.1420	2.8731	−0.1945	(0.0464)	3.3512	(0.0927)	4501.35
$\ln(RV(15\text{min})^{SR})$	0.0919	(0.0156)	0.8229	−3.0959	3.2832	−0.0722	(0.0464)	3.5318	(0.0927)	3893.94
$\ln(RV(BR)^{SR})$	0.0599	(0.0140)	0.7402	−2.8408	2.8731	−0.2618	(0.0464)	3.3588	(0.0927)	4459.41
$\ln(BK(BR)^{SR})$	0.0643	(0.0140)	0.7380	−2.6718	2.6546	−0.1765	(0.0464)	3.3065	(0.0927)	4531.69
$\ln(ZMA(ZMA)^{SR})$	−0.3783	(0.0148)	0.7817	−3.5858	2.4453	−0.1781	(0.0464)	3.1648	(0.0927)	3603.02
$\ln(ZMA(BR)^{SR})$	−0.0447	(0.0142)	0.7523	−3.1196	2.6418	−0.1564	(0.0464)	3.3117	(0.0927)	4392.61
$\ln(BC(ZMA,ZMA)^{SR})$	−0.0402	(0.0137)	0.7232	−2.8926	2.4795	−0.2516	(0.0464)	3.2377	(0.0927)	4637.04
$\ln(BC(ZMA,BR)^{SR})$	0.0616	(0.0139)	0.7320	−2.8646	2.5875	−0.1861	(0.0464)	3.2821	(0.0927)	4572.67
$\ln(FBK(BNHLS)^{SR})$	0.0371	(0.0138)	0.7318	−3.0218	2.5924	−0.1948	(0.0464)	3.2761	(0.0927)	4624.34
$\ln(FBK(BR)^{SR})$	0.0414	(0.0139)	0.7334	−2.9311	2.5924	−0.1923	(0.0464)	3.2478	(0.0927)	4642.55
$\ln(FCK(BNHLS)^{SR})$	0.0513	(0.0140)	0.7377	−2.9554	2.5924	−0.1825	(0.0464)	3.2660	(0.0927)	4614.31
$\ln(FCK(BR)^{SR})$	0.0420	(0.0139)	0.7330	−2.9554	2.5924	−0.1946	(0.0464)	3.2636	(0.0927)	4625.57
$\ln(FMTH(BNHLS)^{SR})$	0.0202	(0.0140)	0.7402	−2.9580	2.5903	−0.1724	(0.0464)	3.2446	(0.0927)	4613.85
$\ln(FMTH(BR)^{SR})$	0.0225	(0.0140)	0.7391	−2.9933	2.5903	−0.1813	(0.0464)	3.2580	(0.0927)	4614.36

The numbers in parentheses are standard errors. LB(10) is the Ljung-Box statistic adjusted for heteroskedasticity following Diebold (1988) to test the null hypothesis of no autocorrelations up to 10 lags.

Table 4: Moneyness of put options

	S/K	< 0.91	deep-in-the-money (DITM)
$0.91 <$	S/K	< 0.97	in-the-money (ITM)
$0.97 <$	S/K	< 1.03	at-the-money (ATM)
$1.03 <$	S/K	< 1.09	out-of-the-money (OTM)
$1.09 <$	S/K		deep-out-of-the-money (DOTM)

$S =$ price of underlying asset and $K =$ exercise price.

Table 5: Put option pricing performance using different models

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	269	102	115	92	68	646
RMSE						
GARCH	26.1499	54.4269	73.8162	68.3135	50.1119*	51.4919
EGARCH	23.7247	57.7130	77.4830	67.4678	53.2560	52.6864
FIEGARCH	22.3849*	50.2646	67.4075	62.6070	53.6379	47.7233
ARFIMA	26.0913	48.6655	65.0287	63.3309	53.8862	47.8233
ARFIMAX	25.5555	47.7278*	63.8177	61.9602*	54.0913	47.0252
HAR	26.7846	49.7439	64.2658	64.7635	52.0057	48.0281
HARX	25.0283	47.8229	62.6602*	62.4055	52.8805	46.5820*
BS	32.5314	68.6507	96.1012	77.3699	57.2586	63.4549
MAE						
GARCH	11.1874	35.9078	59.9180	47.6937	37.6492*	31.7501
EGARCH	11.4961	43.4651	65.4029	47.3676	40.4139	34.2929
FIEGARCH	9.7700*	35.2912	55.6893	42.1285	40.5887	29.8266
ARFIMA	10.4200	27.1600*	48.5898	41.8259	40.6119	27.5089
ARFIMAX	10.2684	27.1664	48.0434	41.1390	40.8868	27.2806
HAR	10.7746	28.7815	48.4673	41.5826	39.1654	27.7038
HARX	10.0268	29.0455	47.8669*	39.8116*	39.9925	27.1621*
BS	13.9732	45.0204	68.1068	49.1433	42.7033	36.5451
RMSPE						
GARCH	0.8413	0.6187	0.2901	0.0904	0.0176*	0.6094
EGARCH	1.6894	0.8511	0.3181	0.0878	0.0190	1.1497
FIEGARCH	1.5059	0.6431	0.2685	0.0805	0.0193	1.0116
ARFIMA	0.5101	0.3344	0.2104	0.0770	0.0196	0.3671
ARFIMAX	0.5052*	0.3302*	0.2068*	0.0754*	0.0197	0.3633*
HAR	0.5254	0.4004	0.2193	0.0795	0.0185	0.3870
HARX	0.5882	0.4214	0.2204	0.0771	0.0190	0.4262
BS	0.8050	0.5275	0.2632	0.0890	0.0265	0.5721
MAPE						
GARCH	0.5723	0.4311	0.2176	0.0635	0.0134*	0.3556
EGARCH	0.9976	0.6029	0.2441	0.0620	0.0144	0.5644
FIEGARCH	0.7894	0.4626	0.2061	0.0551	0.0145	0.4478
ARFIMA	0.4141	0.2552	0.1598	0.0532	0.0147	0.2503
ARFIMAX	0.4100*	0.2547*	0.1578*	0.0521	0.0149	0.2480*
HAR	0.4350	0.2989	0.1671	0.0540	0.0140	0.2673
HARX	0.4407	0.3140	0.1670	0.0520*	0.0144	0.2717
BS	0.7293	0.4422	0.2090	0.0615	0.0169	0.4213

The values of loss functions for the ARFIMA(X) and HAR(X) models are calculated using $RV(\text{1min})^{HL}$.
* indicates the best model which minimizes the loss function.

Table 6: Put option pricing performance using different realized volatilities without the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	269	102	115	92	68	646
RMSE						
$RV(\text{1min})^{SR}$	33.7772	72.3935	75.0869	69.7993	58.3730	57.9549
$RV(\text{5min})^{SR}$	30.9317	61.2178*	62.7097*	63.0628	56.9214	50.9669*
$RV(\text{15min})^{SR}$	30.5734*	61.5178	63.3914	62.4719*	57.4031	51.0372
$RV(BR)^{SR}$	31.6043	63.4804	65.2075	65.3068	57.4180	52.5696
$BK(BR)^{SR}$	31.7665	64.6433	66.3408	65.0087	57.5804	53.0504
$ZMA(ZMA)^{SR}$	36.3328	86.0197	96.5077	80.4540	59.5748	68.3481
$ZMA(BR)^{SR}$	32.9693	69.6890	72.3350	67.8439	57.9567	56.2138
$BC(ZMA, ZMA)^{SR}$	32.4387	65.5644	66.7040	65.6463	57.3477	53.5622
$BC(ZMA, BR)^{SR}$	31.2988	62.1974	63.9627	63.9701	56.6031*	51.6439
$FBK(BNHLS)^{SR}$	31.4696	62.3843	64.0686	64.1482	56.9261	51.8146
$FBK(BR)^{SR}$	31.4152	62.3076	64.1742	64.1708	57.1751	51.8424
$FCK(BNHLS)^{SR}$	31.3722	62.0590	63.8987	63.8778	56.8587	51.6354
$FCK(BR)^{SR}$	31.3220	62.1602	63.6578	63.6128	56.6955	51.5233
$FMTH(BNHLS)^{SR}$	31.5412	62.9802	64.5812	64.1298	57.0168	52.0665
$FMTH(BR)^{SR}$	31.3468	62.5522	64.1934	63.9274	57.0100	51.8140
MAE						
$RV(\text{1min})^{SR}$	16.2666	48.1494	48.7068	38.3192	45.0849	33.2498
$RV(\text{5min})^{SR}$	14.0349	34.8349*	37.9625	35.0266*	43.1779	27.6359*
$RV(\text{15min})^{SR}$	13.9982*	35.8372	37.8965*	35.0757	43.5465	27.8129
$RV(BR)^{SR}$	14.4418	36.4298	38.3761	35.7835	43.7674	28.3006
$BK(BR)^{SR}$	14.5960	37.6658	39.1251	35.2227	43.9036	28.6278
$ZMA(ZMA)^{SR}$	17.8797	62.9990	71.9368	46.1453	46.2418	41.6379
$ZMA(BR)^{SR}$	15.5312	43.9368	43.6062	36.4866	44.4540	31.0430
$BC(ZMA, ZMA)^{SR}$	15.1487	39.9710	40.9115	36.2101	43.8037	29.6700
$BC(ZMA, BR)^{SR}$	14.2087	35.3617	38.6568	36.0104	43.0188*	28.0384
$FBK(BNHLS)^{SR}$	14.3644	36.0563	38.7287	36.1495	43.4338	28.2892
$FBK(BR)^{SR}$	14.3468	35.7935	38.7137	35.9563	43.4892	28.2160
$FCK(BNHLS)^{SR}$	14.2012	35.1550	38.3954	35.7400	43.3019	27.9474
$FCK(BR)^{SR}$	14.3420	35.7243	38.4458	35.6161	43.1863	28.0751
$FMTH(BNHLS)^{SR}$	14.4396	36.6320	38.8405	35.5529	43.4646	28.3496
$FMTH(BR)^{SR}$	14.4308	36.3346	38.6767	35.5519	43.4725	28.2705

This is calculated using the ARFIMAX model. * indicates the best model which minimizes the loss function.

Table 6: (Continued) Put option pricing performance using different realized volatilities without the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	269	102	115	92	68	646
RMSPE						
$RV(\text{1min})^{SR}$	0.6632	0.4449	0.1821	0.0750	0.0235	0.4703
$RV(\text{5min})^{SR}$	0.5935*	0.3372*	0.1511*	0.0678	0.0217	0.4116*
$RV(\text{15min})^{SR}$	0.5959	0.3453	0.1554	0.0674*	0.0219	0.4143
$RV(BR)^{SR}$	0.6038	0.3497	0.1566	0.0702	0.0223	0.4198
$BK(BR)^{SR}$	0.6106	0.3594	0.1583	0.0694	0.0223	0.4253
$ZMA(ZMA)^{SR}$	0.7078	0.5857	0.2532	0.0891	0.0247	0.5248
$ZMA(BR)^{SR}$	0.6392	0.4117	0.1712	0.0724	0.0229	0.4505
$BC(ZMA, ZMA)^{SR}$	0.6249	0.3753	0.1586	0.0701	0.0224	0.4360
$BC(ZMA, BR)^{SR}$	0.5965	0.3402	0.1526	0.0684	0.0217*	0.4139
$FBK(BNHLS)^{SR}$	0.6016	0.3454	0.1528	0.0685	0.0220	0.4176
$FBK(BR)^{SR}$	0.6013	0.3427	0.1523	0.0684	0.0220	0.4171
$FCK(BNHLS)^{SR}$	0.5944	0.3385	0.1524	0.0682	0.0219	0.4124
$FCK(BR)^{SR}$	0.5984	0.3422	0.1511	0.0679	0.0217	0.4152
$FMTH(BNHLS)^{SR}$	0.6038	0.3510	0.1533	0.0683	0.0219	0.4197
$FMTH(BR)^{SR}$	0.6032	0.3483	0.1524	0.0680	0.0220	0.4189
MAPE						
$RV(\text{1min})^{SR}$	0.5909	0.4063	0.1400	0.0461	0.0172	0.3435
$RV(\text{5min})^{SR}$	0.5173*	0.2776*	0.1108*	0.0425*	0.0160	0.2867*
$RV(\text{15min})^{SR}$	0.5211	0.2908	0.1116	0.0428	0.0161	0.2906
$RV(BR)^{SR}$	0.5301	0.2897	0.1119	0.0433	0.0163	0.2943
$BK(BR)^{SR}$	0.5371	0.3004	0.1137	0.0425	0.0164	0.2991
$ZMA(ZMA)^{SR}$	0.6342	0.5635	0.2123	0.0563	0.0180	0.4008
$ZMA(BR)^{SR}$	0.5672	0.3632	0.1241	0.0437	0.0168	0.3236
$BC(ZMA, ZMA)^{SR}$	0.5509	0.3244	0.1175	0.0436	0.0164	0.3095
$BC(ZMA, BR)^{SR}$	0.5206	0.2803	0.1123	0.0434	0.0160*	0.2889
$FBK(BNHLS)^{SR}$	0.5258	0.2876	0.1124	0.0435	0.0162	0.2923
$FBK(BR)^{SR}$	0.5253	0.2839	0.1121	0.0432	0.0161	0.2914
$FCK(BNHLS)^{SR}$	0.5196	0.2788	0.1115	0.0430	0.0161	0.2881
$FCK(BR)^{SR}$	0.5237	0.2838	0.1114	0.0429	0.0160	0.2905
$FMTH(BNHLS)^{SR}$	0.5293	0.2913	0.1123	0.0428	0.0161	0.2942
$FMTH(BR)^{SR}$	0.5294	0.2902	0.1118	0.0427	0.0162	0.2940

This is calculated using the ARFIMAX model. * indicates the best model which minimizes the loss function.

Table 7: Put option pricing performance using different realized volatilities with the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	269	102	115	92	68	646
RMSE						
$RV(\text{1min})^{HL}$	25.5555	47.7278*	63.8177	61.9602	54.0913	47.0252
$RV(\text{5min})^{HL}$	26.5445	50.8029	63.8614	62.7478	54.9065	48.0104
$RV(\text{15min})^{HL}$	27.7902	52.4171	59.1257*	60.3301*	55.4357	47.1124
$RV(BR)^{HL}$	26.7471	51.8100	65.3036	64.0321	54.6332	48.7751
$BK(BR)^{HL}$	27.2496	53.4551	65.5017	64.3414	55.3272	49.3554
$ZMA(ZMA)^{HL}$	25.9568	48.0991	62.1756	62.0206	53.5106*	46.7253*
$ZMA(BR)^{HL}$	26.8309	53.3002	66.0326	64.0710	55.1837	49.2923
$BC(ZMA,ZMA)^{HL}$	25.0585*	49.1121	68.9332	65.0721	53.5186	48.9159
$BC(ZMA,BR)^{HL}$	25.9226	50.8500	66.7806	64.6737	54.7105	48.9157
$FBK(BNHLS)^{HL}$	25.6890	50.0508	66.9318	64.2957	54.1025	48.6280
$FBK(BR)^{HL}$	25.7441	50.2870	66.9377	64.2170	53.9162	48.6435
$FCK(BNHLS)^{HL}$	25.9446	50.9753	67.0718	64.3001	53.9914	48.8578
$FCK(BR)^{HL}$	25.5931	49.7899	66.4022	63.8327	54.3421	48.3760
$FMTH(BNHLS)^{HL}$	25.5855	50.2891	67.1410	63.9189	54.3706	48.6562
$FMTH(BR)^{HL}$	25.3491	49.5069	66.5417	63.6452	54.1618	48.2542
MAE						
$RV(\text{1min})^{HL}$	10.2684	27.1664	48.0434	41.1390	40.8868	27.2806
$RV(\text{5min})^{HL}$	10.4919	27.9590	46.9035	40.8223	41.5855	27.3243
$RV(\text{15min})^{HL}$	11.5637	28.2217	40.6678*	37.2229*	41.9139	26.2240*
$RV(BR)^{HL}$	10.5037	28.6609	48.0673	41.8572	41.4638	27.7818
$BK(BR)^{HL}$	10.7010	29.2284	46.5957	40.9401	41.8132	27.5978
$ZMA(ZMA)^{HL}$	10.3023	26.9121*	45.2253	39.8382	39.9880*	26.4730
$ZMA(BR)^{HL}$	10.5920	29.6355	47.8792	41.5936	41.6622	27.9223
$BC(ZMA,ZMA)^{HL}$	9.9801	29.1722	53.1619	44.7374	40.6054	28.8713
$BC(ZMA,BR)^{HL}$	10.2106	28.9566	49.8740	43.4654	41.5478	28.2659
$FBK(BNHLS)^{HL}$	10.2475	29.0867	50.6716	43.2955	40.8254	28.3436
$FBK(BR)^{HL}$	10.1695	29.0718	50.6228	43.3869	40.7404	28.3041
$FCK(BNHLS)^{HL}$	10.1297	29.3596	50.6263	43.5568	40.8991	28.3745
$FCK(BR)^{HL}$	10.0080	28.7132	50.2599	43.0593	41.2015	28.1176
$FMTH(BNHLS)^{HL}$	10.0540	29.3243	51.0311	43.5582	41.1704	28.4383
$FMTH(BR)^{HL}$	9.9732*	29.0008	50.7100	43.4861	41.0822	28.2768

This is calculated using the ARFIMAX model. * indicates the best model which minimizes the loss function.

Table 7: (Continued) Put option pricing performance using different realized volatilities with the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	269	102	115	92	68	646
RMSPE						
RV(1min)^{HL}	0.5052^{*}	0.3302	0.2068	0.0754	0.0197	0.3633^{*}
RV(5min)^{HL}	0.5258	0.3434	0.2013	0.0745	0.0201	0.3766
RV(15min)^{HL}	0.5163	0.3263^{*}	0.1731^{*}	0.0687^{*}	0.0204	0.3659
$RV(BR)^{HL}$	0.5106	0.3542	0.2082	0.0773	0.0200	0.3701
$BK(BR)^{HL}$	0.5080	0.3581	0.2059	0.0764	0.0206	0.3690
$ZMA(ZMA)^{HL}$	0.5293	0.3456	0.1968	0.0727	0.0193^{*}	0.3784
$ZMA(BR)^{HL}$	0.5083	0.3667	0.2066	0.0762	0.0203	0.3705
$BC(ZMA, ZMA)^{HL}$	0.5286	0.3814	0.2255	0.0799	0.0197	0.3864
$BC(ZMA, BR)^{HL}$	0.5239	0.3735	0.2118	0.0772	0.0200	0.3811
$FBK(BNHLS)^{HL}$	0.7739	0.3759	0.2148	0.0772	0.0197	0.5299
$FBK(BR)^{HL}$	0.5774	0.3793	0.2155	0.0773	0.0197	0.4132
$FCK(BNHLS)^{HL}$	0.5245	0.3714	0.2134	0.0773	0.0198	0.3812
$FCK(BR)^{HL}$	0.5172	0.3772	0.2134	0.0769	0.0199	0.3779
$FMTH(BNHLS)^{HL}$	0.5223	0.3769	0.2153	0.0772	0.0200	0.3810
$FMTH(BR)^{HL}$	0.5139	0.3750	0.2138	0.0767	0.0198	0.3758
MAPE						
RV(1min)^{HL}	0.4100	0.2547	0.1578	0.0521	0.0149	0.2480^{*}
RV(5min)^{HL}	0.4173	0.2618	0.1520	0.0512	0.0151	0.2511
RV(15min)^{HL}	0.4312	0.2492^{*}	0.1294^{*}	0.0462^{*}	0.0153	0.2501
$RV(BR)^{HL}$	0.4131	0.2672	0.1562	0.0531	0.0151	0.2511
$BK(BR)^{HL}$	0.4156	0.2725	0.1520	0.0515	0.0153	0.2521
$ZMA(ZMA)^{HL}$	0.4182	0.2597	0.1472	0.0497	0.0144^{*}	0.2500
$ZMA(BR)^{HL}$	0.4096	0.2798	0.1547	0.0523	0.0152	0.2513
$BC(ZMA, ZMA)^{HL}$	0.4214	0.2884	0.1730	0.0567	0.0148	0.2614
$BC(ZMA, BR)^{HL}$	0.4156	0.2834	0.1614	0.0544	0.0151	0.2559
$FBK(BNHLS)^{HL}$	0.4474	0.2860	0.1644	0.0545	0.0148	0.2701
$FBK(BR)^{HL}$	0.4284	0.2871	0.1644	0.0546	0.0148	0.2623
$FCK(BNHLS)^{HL}$	0.4092	0.2847	0.1635	0.0548	0.0149	0.2538
$FCK(BR)^{HL}$	0.4101	0.2829	0.1631	0.0542	0.0150	0.2538
$FMTH(BNHLS)^{HL}$	0.4138	0.2885	0.1655	0.0549	0.0150	0.2567
$FMTH(BR)^{HL}$	0.4085^{*}	0.2855	0.1641	0.0547	0.0149	0.2538

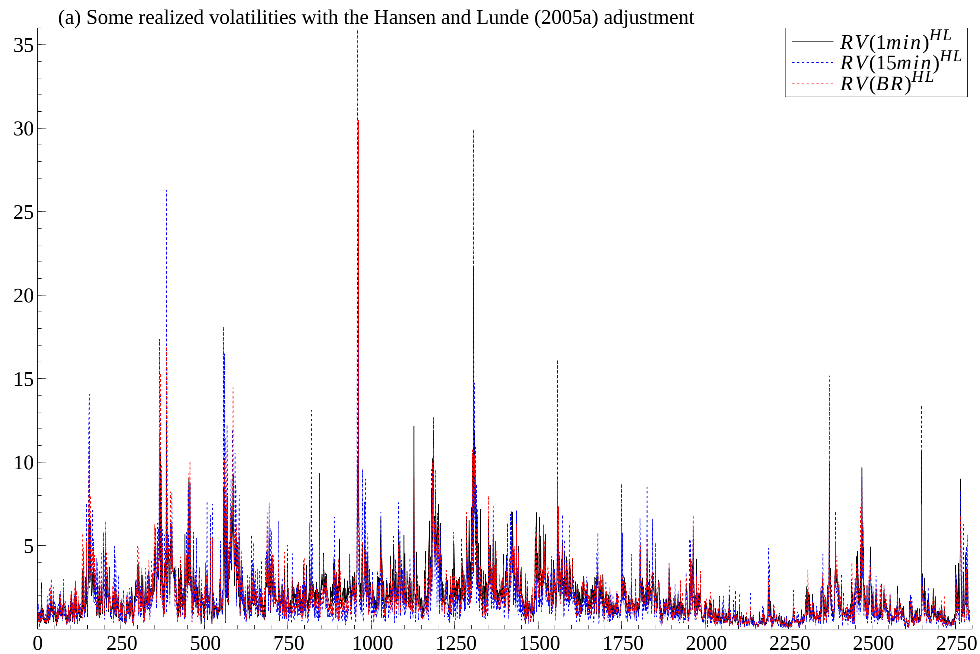
This is calculated using the ARFIMAX model. * indicates the best model which minimizes the loss function.

Table 8: Put option pricing performance of ARCH type models assuming the risk-neutrality and using the Duan (1995) method

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	269	102	115	92	68	646
RMSE						
GARCH						
Risk neutral	26.1499	54.4269	73.8162	68.3135	50.1119	51.4919
Duan	25.8386	53.8704	73.3938	67.9316	50.0423	51.1464
EGARCH						
Risk neutral	23.7247	57.7130	77.4830	67.4678	53.2560	52.6864
Duan	23.9183	57.7392	77.5168	67.6288	53.3418	52.7747
FIEGARCH						
Risk neutral	22.3849	50.2646	67.4075	62.6070	53.6379	47.7233
Duan	22.4163	49.7467	65.5617	61.4178	58.0791	47.5127
MAE						
GARCH						
Risk neutral	11.1874	35.9078	59.9180	47.6937	37.6492	31.7501
Duan	11.0551	35.6945	59.7272	47.0229	37.6450	31.5313
EGARCH						
Risk neutral	11.4961	43.4651	65.4029	47.3676	40.4139	34.2929
Duan	11.6524	43.3410	65.3519	47.3802	40.4712	34.3371
FIEGARCH						
Risk neutral	9.7700	35.2912	55.6893	42.1285	40.5887	29.8266
Duan	9.7211	34.5632	53.4257	40.2707	45.1972	29.5088
RMSPE						
GARCH						
Risk neutral	0.8413	0.6187	0.2901	0.0904	0.0176	0.6094
Duan	0.8694	0.6229	0.2907	0.0902	0.0176	0.6263
EGARCH						
Risk neutral	1.6894	0.8511	0.3181	0.0878	0.0190	1.1497
Duan	1.7832	0.8493	0.3187	0.0878	0.0191	1.2072
FIEGARCH						
Risk neutral	1.5059	0.6431	0.2685	0.0805	0.0193	1.0116
Duan	1.3165	0.6244	0.2580	0.0778	0.0207	0.8922
MAPE						
GARCH						
Risk neutral	0.5723	0.4311	0.2176	0.0635	0.0134	0.3556
Duan	0.5866	0.4318	0.2179	0.0629	0.0134	0.3616
EGARCH						
Risk neutral	0.9976	0.6029	0.2441	0.0620	0.0144	0.5644
Duan	1.0362	0.5993	0.2440	0.0619	0.0145	0.5799
FIEGARCH						
Risk neutral	0.7894	0.4626	0.2061	0.0551	0.0145	0.4478
Duan	0.7594	0.4488	0.1967	0.0525	0.0161	0.4313

"Risk neutral" shows the results assuming the risk-neutrality, which are the same as those in Table 5. "Duan" shows the ones using the Duan (1995) method without assuming the risk-neutrality.

Figure 1: Realized volatility



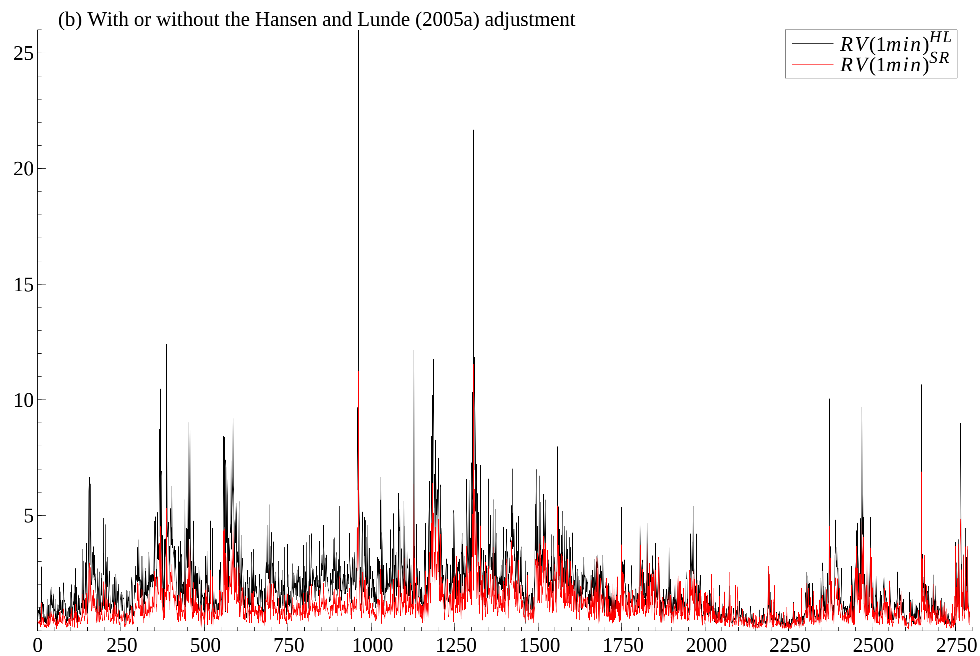
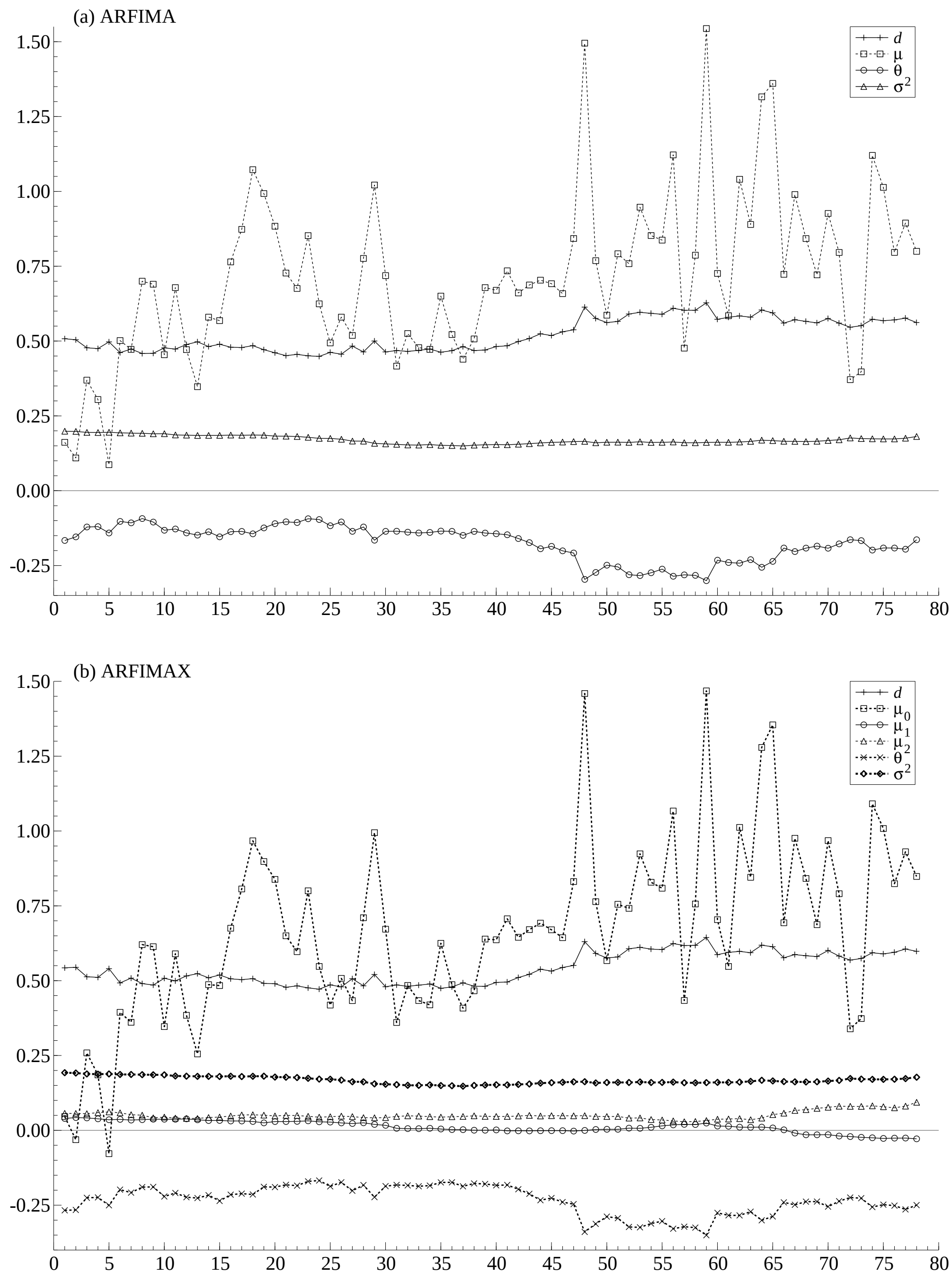
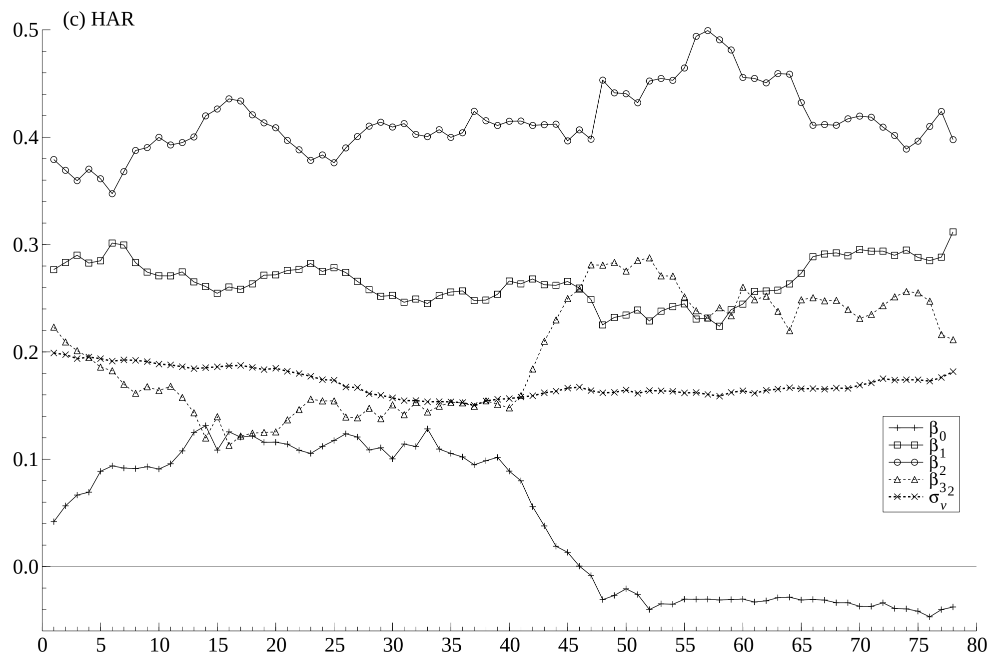
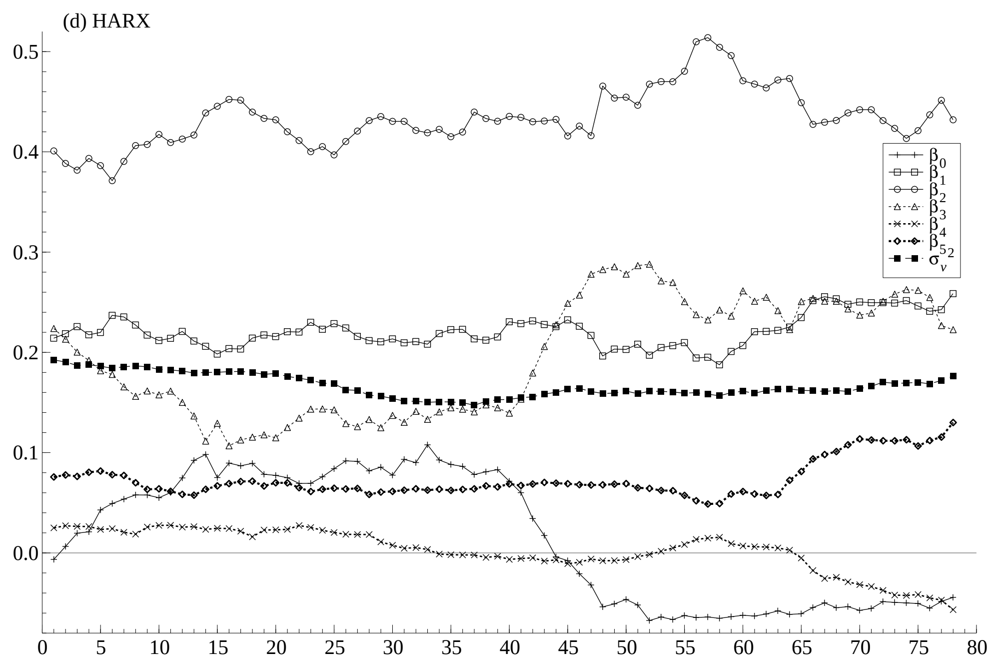
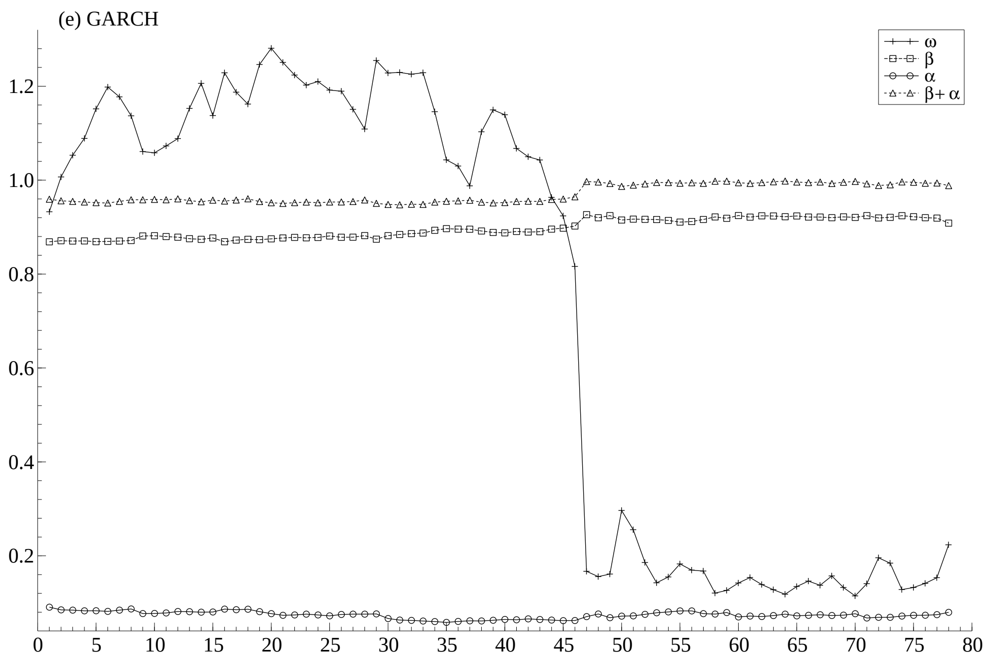


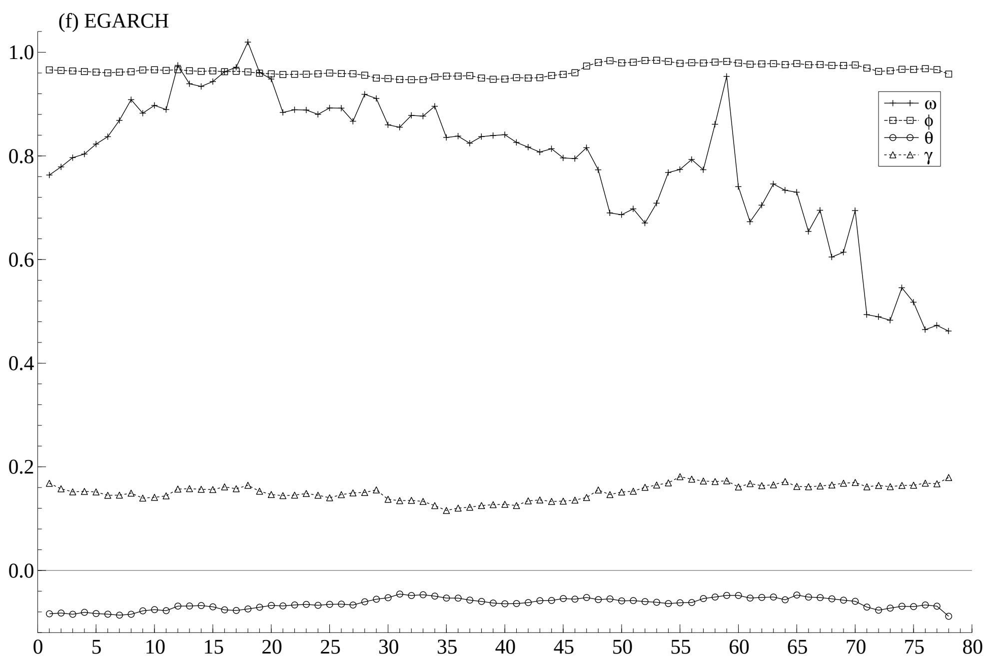
Figure 2: Parameter estimates

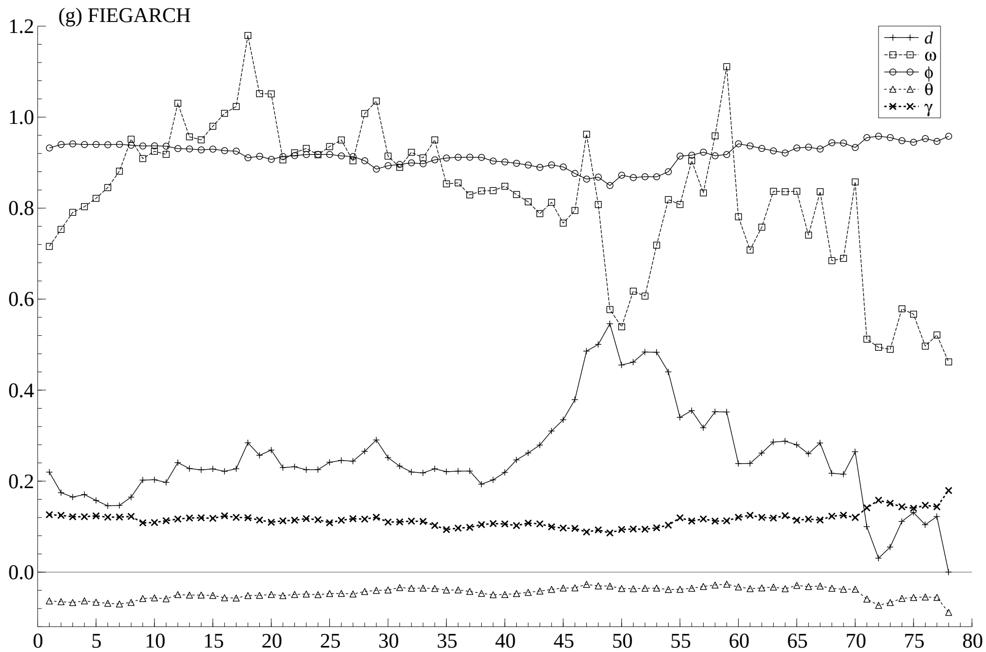



(c) HAR
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80


(d) HARX
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80


(e) GARCH
ω
β
α
β+ α
1.2
1.0
0.8
0.6
0.4
0.2
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80


(f) EGARCH
ω
ϕ
θ
γ
1.0
0.8
0.6
0.4
0.2
0.0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80


(g) FIEGARCH
d
ω
ϕ
θ
γ
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80