Hitotsubashi University

Global COE Hi-Stat Discussion Paper Series 273

Research Unit for Statistical and Empirical Analysis in Social Sciences (Hi-Stat)

Pricing Nikkei 225 Options Using Realized Volatility

Masato Ubukata
Toshiaki Watanabe

January 2013

Hi-Stat Discussion Paper

Hi-Stat
Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitatchi Tokyo, 186-8601 Japan
http://gcoe.ier.hit-u.ac.jp

Pricing Nikkei 225 Options Using Realized Volatility*

MASATO UBUKATA

Department of Economics, Kushiro Public University of Economics

TOSHIAKI WATANABE

Institute of Economic Research, Hitotsubashi University

Abstract

This article examines option pricing performance using realized volatilities with or without handling microstructure noise, non-trading hours and large jumps. The dynamics of realized volatility is specified by ARFIMA(X) and HAR(X) models. Main results using put options on the Nikkei 225 index are: (1) ARFIMAX model performs best, (2) the Hansen and Lunde (2005a) adjustment for non-trading hours improves the performance, (3) methods for reducing microstructure noise-induced bias yield better performance, while if the Hansen-Lunde adjustment is used, the other methods are not necessarily needed and (4) the performance is unaffected by removing large jumps from realized volatility.

JEL Classification Numbers: C13, C22, C52.

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

* This is a revised version of Ubukata and Watanabe (2011). 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 the Joint Usage and Research Center, Institute of Economic Research, Hitotsubashi University (IERPK1109, 1206) is gratefully acknowledged. We would also like to 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. All remaining errors are solely our responsibility.

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 by Duan (1995), Bollerslev and Mikkelsen (1999), Heston and Nandi (2000) and Barone-Adesi et al. (2008). More recently, some authors have applied realized volatility to option pricing as well as the applications to volatility forecasting (Koopman et al. 2005; Andersen et al. 2007) and Value-at-Risk (Giot and Laurent, 2004; Clements et al., 2008; Watanabe, 2012). Stentoft (2008) examine an option pricing model with realized volatility using inverse Gaussian distribution. Christoffersen et al. (2010) propose generalized expected realized volatility model to incorporate realized volatility into the Heston-Nandi (2000) GARCH model and derive closed-form option valuation formulas. Corsi et al. (2011) develop the HAR model (heterogeneous interval autoregressive) by Corsi (2009) and applies it to option pricing.

This article investigates how different option pricing performance would be if we take account of some practical issues involved in calculating realized volatility. The high-frequency financial literature has developed to measure realized volatility associated with microstructure noise-induced bias, the presence of non-trading hours and separate estimates of the continuous and discontinuous (jump) components of the volatility process. This raises the question of what calculation method for realized volatility is reasonably useful in option pricing. As far as we know, there are few that have tackled this problem. One exception is Bandi et al. (2008), which apply different realized volatilities to the pricing of S&P 500 index options and compare their option pricing performance. Their method is, however, different from ours as follows. First, they compare the profits from the straddle trading strategy obtained by substituting the volatility forecasts from a time-series model for realized volatility into the Black-Scholes option pricing formula. We employ a realized volatility option pricing model without the Black-Scholes formula. Second, they compare the performance of realized volatilities with or without taking account of microstructure noise, while our analysis further includes realized volatilities with or without taking account of non-trading hours and asset price jumps and comparison with ARCH-type models.

There are several problems in calculating realized volatility. First, realized volatility is influenced by market microstructure noise induced by various market frictions such as bid-ask spread and nonsynchronous 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., 2004b, 2008; Hansen and Lunde, 2006; Jacod et al., 2009; 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 of Nikkei 225 stock index options traded at Osaka Securities Exchange. 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. Third, financial markets sometimes display asset price discontinuities, so-called jumps. Under jump-diffusion processes for the underlying asset prices, the realized volatility includes variation due to jumps. We also construct realized volatility removing significant large jumps and investigate its contribution to option pricing performance.

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 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 for the examination of realized volatility option pricing model, while it would be important to relax this assumption. However, a direct test for risk premium under some specifications of conditional expectation of the Nikkei 225 returns with realized volatility implies that the data used here may be insensitive to estimated parameters of the market price of risk. Thus, we only consider the case where the risk neutral and physical dynamics of realized volatility are identical. 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 by assuming the risk neutrality and by using the Duan (1995) method.

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) the performance is improved by mitigating the effect of microstructure noise on realized volatility, while 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, (4) the option-pricing performance is not necessarily improved by removing significant large jumps from realized volatility and (5) 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 under these settings. 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 hours on option pricing, we use as many as 34 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 seventeenth 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 17 kinds of realized volatilities, which are denoted as $RV(\text{1min})^{HL}$, $RV(\text{5min})^{HL}$, $RV(\text{10min})^{HL}$, $RV(\text{15min})^{HL}$, $RV(\text{20min})^{HL}$, $RV(BR)^{HL}$, $BK(BR)^{HL}$, $ZMA(ZMA)^{HL}$, $ZMA(BR)^{HL}$, $BC(ZMA, ZMA)^{HL}$, $BC(ZMA, BR)^{HL}$, $FBK(BNHLS)^{HL}$, $FCK(BNHLS)^{HL}$, $FMTH(BNHLS)^{HL}$, $FBK(BR$

[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.

)^{HL}, $FCK(BR)^{HL}$, $FMTH(BR)^{HL}$. For comparison, we also calculate 17 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 $RV(\text{1min})^{SR}$, $RV(\text{5min})^{SR}$, $RV(\text{10min})^{SR}$, $RV(\text{15min})^{SR}$, $RV(\text{20min})^{SR}$, $RV(BR)^{SR}$, $BK(BR)^{SR}$, $ZMA(ZMA)^{SR}$, $ZMA(BR)^{SR}$, $BC(ZMA, ZMA)^{SR}$, $BC(ZMA, BR)^{SR}$, $FBK(BNHLS)^{SR}$, $FCK(BNHLS)^{SR}$, $FMTH(BNHLS)^{SR}$, $FBK(BR)^{SR}$, $FCK(BR)^{SR}$, $FMTH(BR)^{SR}$.

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.

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

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 the Schwartz information criterion (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

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

model of realized volatility. The model consists of three realized volatility components defined over different time periods as follows.

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

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

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

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.

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

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. (2004a), Barndorff-Nielsen and Shephard (2001, 2002) and Nagakura and Watanabe (2011) 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.

[4]Nagakura and Watanabe (2011) consider microstructure noise while Barndorff-Nielsen et al. (2004a) and Barndorff-Nielsen and Shephard (2001, 2002) neglect it.

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

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

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, 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.[6]

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

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

[6]It would be important to relax the assumption of risk neutrality if the risk is priced in the market. For example, the option pricing models with realized volatility proposed by Christoffersen et al. (2010) and Corsi et al. (2011) allow for more flexible specification of the expected return such that

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

where λ denotes the market price of risk and $\lambda = 0$ corresponds to risk neutrality. Christoffersen et al. (2010) define $\bar{\sigma}_t^2$ as a weighted average of two components by daily return innovation and realized volatility such that $\bar{\sigma}_t^2 = m\sigma_t^2 + (1-m)RV_t$ and Corsi et al. (2011) adopt an estimates of integrated volatility as $\bar{\sigma}_t^2$. As introduced in Stentoft (2008), this specification allows us to directly test statistical significance of λ using realized volatility, i.e.,

$$\begin{aligned} R_t &= r - d - \frac{1}{2}RV_t + \lambda RV_t + \sqrt{RV_t}z_t, \\ \frac{R_t - (r-d) + 0.5RV_t}{\sqrt{RV_t}} &= \lambda\sqrt{RV_t} + z_t. \end{aligned}$$

We estimate the parameter λ by the linear regression procedure with Newey and West (1987) standard errors and different realized volatilities, while its statistical significance is not obtained. Even in the other specification like $R_t = r - d - \frac{1}{2}RV_t + \lambda\sqrt{RV_t} + \sqrt{RV_t}z_t$, the parameter λ is not significant at the 5% level, although it is significant at the 10% level in some regressions with different realized volatilities. These results imply that the data used here may be insensitive to the market price of risk.

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.

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

We set $m = 10000$. For variance reduction, we used the control variate and the Empirical Martingale Simulation proposed by Duan and Simonato (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

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

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. $\text{E}^Q[\exp(R_t)|I_{t-1}] = \exp(r - d)$.
3. $\text{Var}^Q[R_t|I_{t-1}] = \text{Var}^P[R_t|I_{t-1}]$ a.s.

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

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

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 $(S^{(1)}_{T+\tau}, \ldots, S^{(m)}_{T+\tau})$ 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 34 daily different realized volatilities listed in Table 1 are calculated with or without using the adjustment coefficient c defined by equation (5).

Before the computation of the 34 daily different realized volatilities, we provide the realized volatility signature plots in Figure 1 to roughly gauge the impact of microstructure frictions contained in the high-frequency returns of the Nikkei 225 index. The signature plots are generated by the sample mean of $RV(\text{1min})$, $RV(\text{2min})$, $\cdots$, $RV(\text{20min})$ based on equation (3) as a function of the sampling frequency of the Nikkei 225 intraday returns. RV^{SR} (solid line) and RV^{HL} (dotted line) correspond to the realized volatility with the squares of overnight and lunch-time returns or the Hansen and Lunde (2005a) adjustment, respectively. If there is no severe microstructure noise, both plots should be leveled off at the frequencies. However, we can find the large impact of microstructure noise, as evidenced by a rapid decline in the plot of RV^{SR} with short measurement intervals. On the other hand, RV^{HL} applying the Hansen and Lunde (2005a) adjustment comparatively stabilizes for all sampling frequencies, although there is a gradual increase in the range of 1-7 minutes. This result implies that the Hansen and Lunde (2005a) adjustment, where the mean of realized volatility is equal to the sample variance of daily returns, may play a role to partially offset the bias caused by the microstructure noise as well as removing the discretization noise attributed to non-trading hours.

Figure 2 plots some kinds of realized volatilities and Table 3 summarizes the descriptive statistics of the 34 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 17 realized volatilities with the Hansen and Lunde (2005a) adjustment, $RV(\text{1min})^{HL}$ has the smallest standard deviation. $RV(\text{20min})^{HL}$ has the largest standard deviation of them as induced by the range from the minimum at 0.0788 to the maximum at 37.3182. The standard deviation of $ZMA(ZMA)^{SR}$ is the smallest of all. These results are confirmed by Figure 2. Figure 2(a) shows that $RV(\text{15min})^{HL}$ is more volatile than $RV(\text{1min})^{HL}$ and $RV(BR)^{HL}$, and Figure 2(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 $\ln(RV(\text{10min})^{HL})$, $\ln(RV(\text{15min})^{HL})$, $\ln(RV(\text{20min})^{HL})$, $\ln(RV(\text{10min})^{SR})$, $\ln(RV(\text{15min})^{SR})$ and $\ln(RV(\text{20min})^{SR})$. The kurtosis of log-realized volatilities is much smaller than those of realized volatilities. The kurtosis of $\ln(RV(\text{1min})^{HL})$ and $\ln(RV(\text{1min})^{SR})$ 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 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. For the maturity, option trading seems to be more active during the week following an expiration date 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 $P_T = K\text{exp}(-r\tau)$.

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 describes the put prices calculated as the average of bid and ask prices and the Black-Scholes implied volatilities for each moneyness. The average put prices range from 21.30 yen in DOTM to 2859.70 yen in DITM. Options in DOTM and OTM account for 41% and 16% of the total sample. The implied volatilities form a smile pattern where options in DOTM and DITM are characterized by higher volatilities more than 30% compared with 22.86% for options in ATM.

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 3 plots the estimates of all parameters in all models for each of the above 78 iterations. Figure 3 (a) and (b) plot the estimates of parameters in the ARFIMA and ARFIMAX models using $RV(\text{15min})^{HL}$. 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 3 (c) and (d) plot the estimates of parameters in the HAR and HARX models using $RV(15\text{min})^{HL}$. The positive estimates of β_1, β_2 and 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 3 (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 0.2 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. MAE and RMSE, which are the absolute metrics, assign a lot of weight to options with high valuations such as DITM and ITM. For MAPE and RMSPE as the relative metrics, much more weight may be put on DOTM and OTM options with valuations close to zero.

Table 6 shows the values of loss functions for ARCH type models with daily returns, the ARFIMA (X) and HAR(X) models with $RV(15\text{min})^{HL}$ and the BS model. In total, the ARFIMAX model performs best except for RMSE, while the HARX model performs best for RMSE. In DOTM, ARFIMAX model performs best for RMSPE and MAPE while the FIEGARCH model performs best for the other loss functions. In OTM and ATM, the ARFIMAX model performs best except for RMSE and it performs best for RMSPE and MAPE in ITM. 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 7 and 8 show the values of loss functions for the ARFIMAX model with 34 different realized volatilities. Table 7 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 $RV(\text{1min})^{SR}$, which does not take account of microstructure noise at all, have larger values than those of the other realized volatilities except $ZMA(ZMA)^{SR}$. This result is intuitive because $RV(\text{1min})^{SR}$ is seriously affected by the microstructure noise as shown by the realized volatility signature plot in Figure 1. Thus we conclude that the option pricing performance is improved by applying methods to remove microstructure noise-induced bias in realized volatility.

Table 8 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, most loss functions in Table 8 are smaller than those in Table 7 regardless of realized volatilities with or without taking account of microstructure noise. The improvement could be seen for all loss functions in DOTM and DITM and for RMSE and MAE in OTM. These results indicate that the Hansen and Lunde (2005a) adjustment may improve the performance of option pricing. It is also noteworthy that the performance of $RV(\text{1min})^{HL}$ is no longer bad compared with the other realized volatilities, although $RV(\text{15min})^{HL}$, $RV(\text{20min})^{HL}$ and $ZMA(ZMA)^{HL}$ perform best in total. This 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 partially offset the bias caused by microstructure noise. This finding is also consistent with the result that the signature plot of RV^{HL} in Figure 1 comparatively stabilizes for all sampling frequencies. Judging from the results in Tables 7 and 8, we conclude that: (1) the Hansen and Lunde (2005a) adjustment for removing the discretization noise induced by non-trading hours improves the performance, (2) methods for reducing microstructure noise-induced bias yield better performance, while if the Hansen and Lunde (2005a) adjustment, which plays an additional role to partially offset the microstructure noise-induced bias, is used, they are not necessarily needed. [7] [8]

[7]We have focused on put options whose maturity is 30 days so far. Following Barone-Adesi et al. (2008), the maturity of 30 days can be classified as short maturity. We also analyze put options whose maturity is 90 days classified as medium maturity. The result, which is not reported in this article to save the space, also supports an evidence that the option pricing performance is improved by taking account of microstructure noise but it does not necessarily improve the performance when the Hansen and Lunde (2005a) adjustment is employed.

[8]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 volatility 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.

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 9 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.

Financial markets sometimes display asset price discontinuities, so-called jumps. If jump-diffusion processes are used instead of equation (1), the realized volatility measures used in this article include not only integrated volatility but also jump variation. Andersen et al. (2007) conclude that the performance of forecasting future realized volatility could be improved by using the HAR model with realized volatility separately from significant jumps. In this subsection, the realized volatility without significant jumps is constructed to assess its contribution to option pricing performance.

We employ the following procedure to remove significant jump variation. Barndorff-Nielsen and Shephard (2004) propose a consistent estimator of the integrated volatility unaffected by jumps called the realized bipower variation

$$BV_t = \mu_1^{-2} \frac{n}{n-1} \sum_{i=2}^{n} \mid r_{t-1+i/n} \mid\mid r_{t-1+(i-1)/n} \mid, \tag{27}$$

where $\mu_1 = \sqrt{2/\pi}$. The bipower variation, which is originally defined as the sum of the products of adjacent absolute returns, is influenced by microstructure noise. To mitigate the noise-induced bias, we construct the following realized bipower measure based on skip-one returns introduced in Huang and Tauchen (2005) and Andersen et al. (2007)

$$BV_{1,t} = \mu_1^{-2} \frac{n}{n-2} \sum_{i=3}^{n} \mid r_{t-1+i/n} \mid\mid r_{t-1+(i-2)/n} \mid . \tag{28}$$

We calculate $BV_{1,t}$ calculated from 15 minute high-frequency returns for which the bipower variation signature plots in equation (28) are leveled off. Identification of the significant jumps is based on an asymptotically normal test statistic using realized volatility and bipower variation

$$Z_t^* = \frac{\sqrt{n}(RV_t - BV_{1,t})RV_t^{-1}}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2} - 5)\text{Max}(1, TQ_{1,t}BV_{1,t}^{-2})}}, \tag{29}$$

where

$$\begin{aligned} TQ_{1,t} &= n\mu_{4/3}^{-3} \frac{n}{n-4} \sum_{i=5}^{n} \mid r_{t-1+i/n} \mid^{4/3}\mid r_{t-1+(i-2)/n} \mid^{4/3}\mid r_{t-1+(i-4)/n} \mid^{4/3}, \qquad (30) \\ \mu_{4/3} &= 2^{2/3}\Gamma(7/6)\Gamma(1/2)^{-1}. \end{aligned}$$

$TQ_{1,t}$ is called the realized tripower quarticity based on skip-one returns which converges to the integrated quarticity of continuous processes. RV_t in equation (29) is replaced by RV(15min) during the market open. Thus daily realized volatility without significant jumps (RV-$Jump$) is constructed by the realization of Z_t^* and some critical value Φ_{α^*}

$$RV\!-\!Jump_t \;=\; I[Z_t^* > \Phi_{\alpha^*}]BV_{1,t} + I[Z_t^* \le \Phi_{\alpha^*}]RV_t, \tag{31}$$

where $I[\,\cdot\,]$ denotes the indicator function. The proportion of day with significant jumps based on significance levels $(1-\alpha^* = 0.001, 0.01, 0.025, 0.05)$ takes 5.2%, 13.1%, 20.2% and 26.0% for the full sample, respectively. We estimate the ARFIMAX model and simulating option prices using the RV-$Jump$ based on $\alpha^* = 0.999$, which is the same setting as Andersen et al. (2007).

Table 10 summarizes the values of loss functions for $RV(\text{15min})^{HL}$ and RV-$Jump$ with the Hansen and Lunde (2005a) adjustment. The values of loss functions for RV-$Jump$ are not so much different from those for $RV(\text{15min})^{HL}$. Judging from the result in Table 10, we may conclude that the option-pricing performance is not necessarily improved by removing significant large jumps from realized volatility.

7 Conclusions

This article analyzes whether realized volatility is useful for option pricing. Different realized volatilities are calculated with or without taking account of microstructure noise, with or without using overnight and lunch-time returns and with or without separating significant large jumps from realized volatility. 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. 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) the performance is improved by removing the bias from microstructure noise, while if the Hansen and Lunde (2005a), which also plays a role to remove the microstructure noise-induced bias 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, (4) the performance is not necessarily improved by removing significant large jumps from realized volatility and (5) the Duan (1995) method does not improve the performance compared with assuming the risk neutrality.

Several extensions are possible. First, Jacod et al. (2009) propose an alternative realized volatility using preaveraging approach robust to microstructure noise. Andersen et al. (2012) propose two new jump-robust estimators of integrated volatility called the minimum or median realized volatility. It is interesting whether the performance of option pricing will also be improved by applying their estimators. Second, Hansen et al. (2012) and Takahashi et al. (2009) have proposed to model daily returns and realized volatility jointly. They extend ARCH type models and the stochastic volatility model respectively. It is also interesting to apply their methods to option pricing.

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-, 10-, 15- and 20-minute returns, RV(1min), RV(5min), RV(10min), RV(15min) and RV(20min).

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(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. We provide the realized volatility signature plots in Figure 1 to roughly gauge the impact of microstructure frictions contained the high-frequency returns of the Nikkei 225 index. In addition, evidence from previous studies suggests that it may be optimal to use 5 to 30-minute return data. Hence, we employ RV(5min), RV(10min), RV(15min) and RV(20min) which are equal to the sum of squared 5-, 10-, 15- and 20-minute returns ($n = 54, 27, 18$ and 14), 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. (2004b) with a finite sample optimal number of autocovariances proposed by Bandi and Russell (2011), $BK(BR)$.

$RV(1\text{min})$, $RV(5\text{min})$, $RV(15\text{min})$ 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. (2004b) 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

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

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

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

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 $\hat{\mathrm{E}}(e^2) = \frac{1}{n}\sum_{i=1}^{n} r^2_{t-1+i/n}$ 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. (2004b) 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 $\tilde{K}/n \leq 1/2$,

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

$$+ \left[\frac{2}{n^5}IQ + \left(\frac{-4\sigma_\eta^4 - 8\sigma_\eta^2 IV + 4IQ - 8IV^2}{n^4}\right)\right.$$
$$\left. + \left(\frac{-4\sigma_\eta^4 - 16\sigma_\eta^2 IV + 2IQ}{n^3}\right) + \left(\frac{8\sigma_\eta^4 - 8\sigma_\eta^2 IV}{n^2}\right) + \frac{8}{n}\sigma_\eta^4\right]\frac{n^2}{\tilde{K}^2}, \tag{A.11}$$

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 $\text{bias}(RK) = 0$ 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.

Andersen, T. G., Dobrev, D. and Schaumburg, E. (2012), "Jump-robust volatility estimation using nearest neighbor truncation," *Journal of Econometrics*, Vol.169, No.1, pp.75-93.

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. (2004a), "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. (2002), "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. (2004b), "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.

Barone-Adesi, G., Engle, R. and Mancini, L. (2008), "A garch option pricing with filtered historical simulation," *Review of Financial Studies*, Vol.21, No.3, pp.1223-1258.

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.

Christoffersen, P., Feunou, B., Jacobs, K. and Meddahi, N. (2010), "The economic value of realized volatility," Working paper.

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.

Corsi, F., Fusari, N. and Vecchia, D. L. (2011), "Realizing smiles: pricing options with realized volatility," Swiss Finance Institute Research Paper Series No.10-05.

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 Simonato, 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. (2012), "Realized GARCH: A joint model for returns and realized measures of volatility," *Journal of Applied Econometrics*, Vol.27, No.6, pp.877–906.

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.

Heston, S. and Nandi, S. (2000), "A closed-form garch option valuation model," *Review of Financial Studies*, Vol.13, No.3, pp.585–625.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. and Vetter, M. (2009), "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, Vol.119, No.7, pp.2249-2276.

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. (2011), "A state space approach to estimating integrated variance under the existence of microstructure noise," Global COE Hi-Stat Discussion Paper Series 200, 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.

Stentoft, L. (2008), "Option pricing using realized volatility," Working Paper at CREATES, University of Copenhagen.

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.

Ubukata, M. and Watanabe, T. (2011), “Pricing Nikkei 225 Options Using Realized Volatility,” IMES Discussion Paper Series No.2011-E-18.

Watanabe, T. (2012), “Quantile forecasts of financial returns using realized GARCH models,” *Japanese Economic Review*, Vol.63, No.1, pp.68–80.

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 10-minute returns	$RV(\text{10min})^{HL}$	$RV(\text{10min})^{SR}$
Realized volatility with 15-minute returns	$RV(\text{15min})^{HL}$	$RV(\text{15min})^{SR}$
Realized volatility with 20-minute returns	$RV(\text{20min})^{HL}$	$RV(\text{20min})^{SR}$
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. (2004b) 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)
Std.	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
$RV(\text{5min})^{HL}$	1.9543	(0.0330)	1.7433	0.1136	28.1494	4.1908	(0.0464)	39.6745	(0.0927)	1505.58
$RV(\text{10min})^{HL}$	1.9094	(0.0359)	1.8993	0.1011	25.4075	4.2299	(0.0464)	34.6156	(0.0927)	1153.27
$RV(\text{15min})^{HL}$	1.8753	(0.0395)	2.0879	0.0874	35.4308	5.1948	(0.0464)	53.0276	(0.0927)	843.09
$RV(\text{20min})^{HL}$	1.8657	(0.0407)	2.1504	0.0788	37.3182	5.3166	(0.0464)	55.4074	(0.0927)	778.89
$RV(BR)^{HL}$	1.9656	(0.0320)	1.6905	0.1123	29.9739	3.9445	(0.0464)	39.8392	(0.0927)	1685.20
$BK(BR)^{HL}$	1.9296	(0.0330)	1.7446	0.1050	23.9003	3.7535	(0.0464)	28.8679	(0.0927)	1464.14
$ZMA(ZMA)^{HL}$	1.9502	(0.0343)	1.8143	0.0900	27.5854	3.4608	(0.0464)	26.8874	(0.0927)	1205.76
$ZMA(BR)^{HL}$	1.9400	(0.0350)	1.8511	0.0993	26.8615	4.1347	(0.0464)	34.4985	(0.0927)	1320.11
$BC(ZMA, ZMA)^{HL}$	2.0474	(0.0323)	1.7042	0.1157	26.3048	3.4702	(0.0464)	30.1107	(0.0927)	1787.15
$BC(ZMA, BR)^{HL}$	1.9835	(0.0331)	1.7473	0.0938	23.3834	3.5291	(0.0464)	26.6684	(0.0927)	1618.02
$FBK(BNHLS)^{HL}$	2.0018	(0.0331)	1.7473	0.0942	24.6744	3.4010	(0.0464)	25.4522	(0.0927)	1715.00
$FBK(BR)^{HL}$	2.0001	(0.0331)	1.7501	0.0938	24.4619	3.3763	(0.0464)	24.9052	(0.0927)	1696.88
$FCK(BNHLS)^{HL}$	1.9921	(0.0337)	1.7799	0.0989	23.8425	3.5437	(0.0464)	26.7317	(0.0927)	1605.29
$FCK(BR)^{HL}$	2.0006	(0.0331)	1.7480	0.0937	24.4482	3.3721	(0.0464)	24.9008	(0.0927)	1708.12
$FMTH(BNHLS)^{HL}$	1.9992	(0.0341)	1.8032	0.1016	24.8002	3.5110	(0.0464)	25.9993	(0.0927)	1577.72
$FMTH(BR)^{HL}$	2.0044	(0.0338)	1.7877	0.1010	24.9093	3.4833	(0.0464)	26.0971	(0.0927)	1640.56
$RV(\text{1min})^{SR}$	1.1362	(0.0163)	0.8608	0.0745	11.5340	2.8392	(0.0464)	21.3989	(0.0927)	1731.78
$RV(\text{5min})^{SR}$	1.4638	(0.0227)	1.2020	0.0715	17.6948	3.4511	(0.0464)	28.4366	(0.0927)	1669.93
$RV(\text{10min})^{SR}$	1.5919	(0.0276)	1.4595	0.0770	20.0691	3.8113	(0.0464)	30.1096	(0.0927)	1263.18
$RV(\text{15min})^{SR}$	1.6049	(0.0311)	1.6419	0.0583	26.8341	4.8255	(0.0464)	47.6240	(0.0927)	942.02
$RV(\text{20min})^{SR}$	1.5968	(0.0320)	1.6902	0.0627	28.5123	4.9021	(0.0464)	48.6074	(0.0927)	889.37
$RV(BR)^{SR}$	1.4108	(0.0211)	1.1135	0.0715	17.6948	3.2057	(0.0464)	27.4401	(0.0927)	1812.63
$BK(BR)^{SR}$	1.3850	(0.0217)	1.1461	0.0734	14.2190	3.1023	(0.0464)	20.8864	(0.0927)	1670.92
$ZMA(ZMA)^{SR}$	0.9271	(0.0149)	0.7890	0.0277	11.5340	2.8987	(0.0464)	21.2233	(0.0927)	1267.43
$ZMA(BR)^{SR}$	1.2581	(0.0204)	1.0775	0.0448	14.0378	3.3175	(0.0464)	23.7880	(0.0927)	1580.37
$BC(ZMA, ZMA)^{SR}$	1.2217	(0.0178)	0.9399	0.0554	11.9358	2.7380	(0.0464)	19.2841	(0.0927)	1874.42
$BC(ZMA, BR)^{SR}$	1.3734	(0.0210)	1.1082	0.0491	13.2960	2.8986	(0.0464)	18.9612	(0.0927)	1804.78
$FBK(BNHLS)^{SR}$	1.3416	(0.0203)	1.0727	0.0487	13.3620	2.7863	(0.0464)	17.9634	(0.0927)	1896.82
$FBK(BR)^{SR}$	1.3490	(0.0205)	1.0811	0.0487	13.3620	2.7828	(0.0464)	17.8191	(0.0927)	1871.46
$FCK(BNHLS)^{SR}$	1.3685	(0.0211)	1.1142	0.0521	13.3620	2.8945	(0.0464)	18.8365	(0.0927)	1814.94
$FCK(BR)^{SR}$	1.3500	(0.0205)	1.0806	0.0487	13.3620	2.7774	(0.0464)	17.7930	(0.0927)	1885.29
$FMTH(BNHLS)^{SR}$	1.3312	(0.0206)	1.0896	0.0519	13.3336	2.8760	(0.0464)	18.4626	(0.0927)	1793.49
$FMTH(BR)^{SR}$	1.3315	(0.0205)	1.0830	0.0518	13.3336	2.8448	(0.0464)	18.3490	(0.0927)	1837.27

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
$\ln(RV(5\text{min})^{HL})$	0.3892	(0.0143)	0.7564	−2.1750	3.3375	−0.1389	(0.0464)	3.2799	(0.0927)	5146.27
$\ln(RV(10\text{min})^{HL})$	0.3232	(0.0152)	0.8024	−2.2914	3.2350	−0.0364	(0.0464)	3.2320	(0.0927)	4687.98
$\ln(RV(15\text{min})^{HL})$	0.2688	(0.0158)	0.8336	−2.4372	3.5676	0.0616	(0.0464)	3.3121	(0.0927)	4093.84
$\ln(RV(20\text{min})^{HL})$	0.2438	(0.0162)	0.8564	−2.5409	3.6195	0.0666	(0.0464)	3.2489	(0.0927)	3827.66
$\ln(RV(BR)^{HL})$	0.3976	(0.0144)	0.7606	−2.1863	3.4003	−0.2060	(0.0464)	3.2079	(0.0927)	5257.56
$\ln(BK(BR)^{HL})$	0.3692	(0.0145)	0.7637	−2.2543	3.1739	−0.1117	(0.0464)	3.2825	(0.0927)	5076.13
$\ln(ZMA(ZMA)^{HL})$	0.3502	(0.0153)	0.8070	−2.4083	3.3173	−0.1366	(0.0464)	3.2729	(0.0927)	4464.99
$\ln(ZMA(BR)^{HL})$	0.3564	(0.0149)	0.7849	−2.3098	3.2907	−0.0918	(0.0464)	3.3014	(0.0927)	4887.09
$\ln(BC(ZMA,ZMA)^{HL})$	0.4396	(0.0145)	0.7678	−2.1571	3.2698	−0.2920	(0.0464)	3.2139	(0.0927)	5528.40
$\ln(BC(ZMA,BR)^{HL})$	0.3985	(0.0145)	0.7683	−2.3661	3.1520	−0.1838	(0.0464)	3.3151	(0.0927)	5057.94
$\ln(FBK(BNHLS)^{HL})$	0.4066	(0.0146)	0.7721	−2.3619	3.2058	−0.2010	(0.0464)	3.2857	(0.0927)	5164.94
$\ln(FBK(BR)^{HL})$	0.4045	(0.0146)	0.7737	−2.3668	3.1971	−0.1976	(0.0464)	3.2797	(0.0927)	5154.57
$\ln(FCK(BNHLS)^{HL})$	0.3958	(0.0147)	0.7773	−2.3137	3.1715	−0.1713	(0.0464)	3.2680	(0.0927)	5077.37
$\ln(FCK(BR)^{HL})$	0.4052	(0.0146)	0.7732	−2.3679	3.1966	−0.1998	(0.0464)	3.2839	(0.0927)	5152.98
$\ln(FMTH(BNHLS)^{HL})$	0.3947	(0.0148)	0.7824	−2.2869	3.2109	−0.1537	(0.0464)	3.2350	(0.0927)	5082.80
$\ln(FMTH(BR)^{HL})$	0.4008	(0.0148)	0.7794	−2.2930	3.2152	−0.1752	(0.0464)	3.2651	(0.0927)	5135.07
$\ln(RV(1\text{min})^{SR})$	−0.1099	(0.0133)	0.7040	−2.5963	2.4453	−0.2092	(0.0464)	3.1449	(0.0927)	4730.86
$\ln(RV(5\text{min})^{SR})$	0.1236	(0.0138)	0.7275	−2.6374	2.8733	−0.1652	(0.0464)	3.2743	(0.0927)	4616.29
$\ln(RV(10\text{min})^{SR})$	0.1702	(0.0146)	0.7707	−2.5637	2.9992	−0.0858	(0.0464)	3.2701	(0.0927)	4403.00
$\ln(RV(15\text{min})^{SR})$	0.1471	(0.0152)	0.8012	−2.8421	3.2897	−0.0232	(0.0464)	3.4282	(0.0927)	3994.44
$\ln(RV(20\text{min})^{SR})$	0.1243	(0.0156)	0.8225	−2.7700	3.3503	−0.0123	(0.0464)	3.3291	(0.0927)	3834.81
$\ln(RV(BR)^{SR})$	0.0936	(0.0137)	0.7237	−2.6374	2.8733	−0.2330	(0.0464)	3.2829	(0.0927)	4559.45
$\ln(BK(BR)^{SR})$	0.0618	(0.0140)	0.7373	−2.6117	2.6546	−0.1744	(0.0464)	3.2970	(0.0927)	4545.05
$\ln(ZMA(ZMA)^{SR})$	−0.3687	(0.0148)	0.7825	−3.5858	2.4453	−0.1900	(0.0464)	3.1880	(0.0927)	3594.07
$\ln(ZMA(BR)^{SR})$	−0.0459	(0.0142)	0.7516	−3.1047	2.6418	−0.1565	(0.0464)	3.3066	(0.0927)	4403.81
$\ln(BC(ZMA,ZMA)^{SR})$	−0.0480	(0.0137)	0.7230	−2.8926	2.4795	−0.2490	(0.0464)	3.2206	(0.0927)	4617.91
$\ln(BC(ZMA,BR)^{SR})$	0.0590	(0.0138)	0.7311	−3.0130	2.5875	−0.1886	(0.0464)	3.2966	(0.0927)	4567.44
$\ln(FBK(BNHLS)^{SR})$	0.0358	(0.0139)	0.7323	−3.0218	2.5924	−0.1992	(0.0464)	3.2784	(0.0927)	4618.38
$\ln(FBK(BR)^{SR})$	0.0403	(0.0139)	0.7336	−3.0218	2.5924	−0.1992	(0.0464)	3.2793	(0.0927)	4623.05
$\ln(FCK(BNHLS)^{SR})$	0.0509	(0.0140)	0.7373	−2.9554	2.5924	−0.1812	(0.0464)	3.2622	(0.0927)	4607.92
$\ln(FCK(BR)^{SR})$	0.0414	(0.0139)	0.7332	−3.0218	2.5924	−0.2003	(0.0464)	3.2813	(0.0927)	4621.07
$\ln(FMTH(BNHLS)^{SR})$	0.0210	(0.0140)	0.7398	−2.9580	2.5903	−0.1707	(0.0464)	3.2474	(0.0927)	4600.12
$\ln(FMTH(BR)^{SR})$	0.0226	(0.0140)	0.7388	−2.9612	2.5903	−0.1806	(0.0464)	3.2554	(0.0927)	4615.30

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: Options data description

	DOTM	OTM	ATM	ITM	DITM
Sample size	268	101	115	92	71
Put price (yen)					
Mean	21.30	108.44	339.61	888.34	2859.70
Std.	35.01	78.73	147.60	314.48	1298.70
Implied volatility (%)					
Mean	31.53	24.83	22.86	22.55	39.79
Std.	8.91	8.20	7.98	8.52	14.11

Put price and implied volatility are the average of the bid and ask prices and the Black-Scholes implied volatility.

Table 6: Put option pricing performance using different models

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	268	101	115	92	71	647
RMSE						
GARCH	26.0156	54.6635	74.2943	69.1827	49.6875*	51.7628
EGARCH	24.2628	57.9856	77.9481	67.8789	52.6027	52.9858
FIEGARCH	22.5543*	50.5758	67.4635	62.6299	53.0315	47.8055
ARFIMA	27.9894	52.9349	61.2785	62.2902	53.8925	47.9563
ARFIMAX	27.2277	52.0127	59.7248	60.5138	54.4645	47.0068
HAR	27.7424	51.8933	60.2853	62.5671	52.8699	47.4187
HARX	25.0464	47.8573*	56.8534*	59.0927*	53.2045	44.7110*
BS	32.4432	68.7164	96.2482	77.9895	56.2863	63.4998
MAE						
GARCH	11.0711	36.4578	60.6267	48.5941	37.5032*	32.0784
EGARCH	11.8561	43.4189	65.7360	47.5408	39.9573	34.5179
FIEGARCH	9.9522*	35.5033	55.6567	42.0410	40.0367	29.9288
ARFIMA	11.3099	28.3373	42.9685	38.4628	40.4842	26.6576
ARFIMAX	11.0443	28.0901*	41.8705*	37.5631	40.9825	26.2406*
HAR	11.6949	31.0200	44.0702	39.4255	39.9260	27.5073
HARX	10.6813	29.4854	42.5278	37.3004*	40.2374	26.3057
BS	13.7432	44.8889	68.7296	50.4678	42.0261	36.7044
RMSPE						
GARCH	0.8403	0.6204	0.2918	0.0914	0.0178*	0.6074
EGARCH	1.7426	0.8541	0.3204	0.0883	0.0190	1.1795
FIEGARCH	1.2416	0.6452	0.2679	0.0805	0.0192	0.8469
ARFIMA	0.5412	0.3417	0.1890	0.0726	0.0198	0.3830
ARFIMAX	0.5018*	0.3357*	0.1821*	0.0702*	0.0201	0.3585*
HAR	0.5537	0.4346	0.2088	0.0775	0.0196	0.4064
HARX	0.5631	0.4351	0.2022	0.0740	0.0197	0.4111
BS	0.8006	0.5289	0.2635	0.0898	0.0261	0.5681
MAPE						
GARCH	0.5636	0.4363	0.2199	0.0646	0.0135*	0.3513
EGARCH	1.0230	0.6041	0.2460	0.0623	0.0145	0.5722
FIEGARCH	0.7511	0.4658	0.2059	0.0551	0.0145	0.4299
ARFIMA	0.4307	0.2612	0.1406	0.0484	0.0149	0.2527
ARFIMAX	0.4127*	0.2552*	0.1361*	0.0471*	0.0151	0.2433*
HAR	0.4523	0.3175	0.1522	0.0518	0.0146	0.2729
HARX	0.4441	0.3125	0.1481	0.0493	0.0147	0.2677
BS	0.7221	0.4432	0.2106	0.0632	0.0168	0.4165

The values of loss functions for the ARFIMA(X) and HAR(X) models are calculated using RV(15min) ^{HL}.
* indicates the best model which minimizes the loss function.

Table 7: Put option pricing performance using different realized volatilities without the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	268	101	115	92	71	647
RMSE						
$RV(\text{1min})^{SR}$	33.9227	72.1248	74.8183	70.0870	57.2975	57.8139
$RV(\text{5min})^{SR}$	30.4495	59.6870	61.8254	63.0656	55.5741	50.2400
$RV(\text{10min})^{SR}$	29.3536*	57.2161*	59.8651*	61.8367	55.1318*	48.8080
$RV(\text{15min})^{SR}$	29.3730	57.6910	59.8751	60.7149*	55.3323	48.7268*
$RV(\text{20min})^{SR}$	29.8907	59.0459	61.1204	62.3050	55.5908	49.6935
$RV(BR)^{SR}$	31.4080	61.9816	63.5877	64.5523	55.8642	51.5946
$BK(BR)^{SR}$	31.8216	64.7460	66.7455	65.6855	56.2317	53.1619
$ZMA(ZMA)^{SR}$	36.3957	85.3358	95.5640	80.5196	58.4440	67.8635
$ZMA(BR)^{SR}$	32.9158	69.2221	72.0256	68.2550	56.9409	56.0100
$BC(ZMA, ZMA)^{SR}$	32.6506	66.1236	67.4768	66.6304	56.3029	53.9682
$BC(ZMA, BR)^{SR}$	31.3256	61.9178	63.3207	63.9560	55.8048	51.3905
$FBK(BNHLS)^{SR}$	31.4899	62.4976	63.8867	64.3749	55.5842	51.7133
$FBK(BR)^{SR}$	31.3638	62.3141	63.7062	64.0400	55.9466	51.5910
$FCK(BNHLS)^{SR}$	31.1045	61.5259	63.2229	63.7776	55.4729	51.1685
$FCK(BR)^{SR}$	31.2035	61.8627	63.4931	64.1561	55.5018	51.3866
$FMTH(BNHLS)^{SR}$	31.3766	62.9582	64.5155	64.4555	55.8836	51.9598
$FMTH(BR)^{SR}$	31.4534	62.6571	63.9747	63.9862	55.4588	51.6702
MAE						
$RV(\text{1min})^{SR}$	16.3136	47.5962	48.5876	38.9868	43.4660	33.1371
$RV(\text{5min})^{SR}$	13.4821	32.3113	37.9234	35.7661	41.4337	27.0017
$RV(\text{10min})^{SR}$	12.7270*	29.8991*	37.0606	35.4503	41.1649	26.0846*
$RV(\text{15min})^{SR}$	13.0653	31.1078	36.5558*	34.6110*	41.1455*	26.2022
$RV(\text{20min})^{SR}$	13.0366	31.4907	36.5906	34.6329	41.4576	26.2937
$RV(BR)^{SR}$	14.0925	33.9555	37.9471	36.2942	41.8724	27.6387
$BK(BR)^{SR}$	14.5502	37.1796	39.2043	36.2738	42.0863	28.5756
$ZMA(ZMA)^{SR}$	17.9269	62.0901	71.0353	46.9129	44.6885	41.3191
$ZMA(BR)^{SR}$	15.5126	43.2578	43.1437	37.4512	42.8567	30.8753
$BC(ZMA, ZMA)^{SR}$	15.1992	40.2561	41.2533	36.9121	42.3911	29.8131
$BC(ZMA, BR)^{SR}$	14.2124	34.8881	37.8539	36.0624	41.8269	27.7794
$FBK(BNHLS)^{SR}$	14.2990	35.6258	38.2322	36.3791	41.8211	28.0421
$FBK(BR)^{SR}$	14.2527	35.4728	38.1690	36.2129	42.0552	27.9898
$FCK(BNHLS)^{SR}$	14.0746	34.5858	37.9700	36.0905	41.6088	27.6758
$FCK(BR)^{SR}$	14.1644	35.1162	38.2068	36.3642	41.5789	27.8735
$FMTH(BNHLS)^{SR}$	14.3573	36.2120	38.5568	36.3048	41.8786	28.2112
$FMTH(BR)^{SR}$	14.4084	36.1213	38.3852	35.9645	41.6931	28.1189

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 without the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	268	101	115	92	71	647
RMSPE						
$RV(\text{1min})^{SR}$	0.6593	0.4420	0.1822	0.0757	0.0231	0.4662
$RV(\text{5min})^{SR}$	0.5735	0.3189	0.1536	0.0685	0.0211	0.3962
$RV(\text{10min})^{SR}$	0.5538*	0.3027*	0.1553	0.0682	0.0206*	0.3825*
$RV(\text{15min})^{SR}$	0.5629	0.3104	0.1550	0.0669	0.0207	0.3889
$RV(\text{20min})^{SR}$	0.5610	0.3147	0.1506*	0.0678	0.0214	0.3880
$RV(BR)^{SR}$	0.5875	0.3282	0.1554	0.0700	0.0215	0.4060
$BK(BR)^{SR}$	0.6044	0.3564	0.1593	0.0705	0.0217	0.4200
$ZMA(ZMA)^{SR}$	0.7044	0.5806	0.2513	0.0892	0.0242	0.5202
$ZMA(BR)^{SR}$	0.6342	0.4076	0.1711	0.0731	0.0224	0.4456
$BC(ZMA, ZMA)^{SR}$	0.6223	0.3795	0.1612	0.0714	0.0221	0.4339
$BC(ZMA, BR)^{SR}$	0.5940	0.3372	0.1508	0.0685	0.0215	0.4107
$FBK(BNHLS)^{SR}$	0.5954	0.3457	0.1516	0.0689	0.0215	0.4127
$FBK(BR)^{SR}$	0.5952	0.3439	0.1519	0.0684	0.0216	0.4123
$FCK(BNHLS)^{SR}$	0.5900	0.3386	0.1511	0.0684	0.0212	0.4085
$FCK(BR)^{SR}$	0.5913	0.3409	0.1511	0.0686	0.0213	0.4096
$FMTH(BNHLS)^{SR}$	0.5986	0.3495	0.1532	0.0690	0.0215	0.4152
$FMTH(BR)^{SR}$	0.6005	0.3491	0.1527	0.0683	0.0214	0.4163
MAPE						
$RV(\text{1min})^{SR}$	0.5865	0.4036	0.1403	0.0471	0.0166	0.3394
$RV(\text{5min})^{SR}$	0.4947	0.2565	0.1139	0.0439	0.0153	0.2731
$RV(\text{10min})^{SR}$	0.4725*	0.2370*	0.1143	0.0439	0.0151	0.2609*
$RV(\text{15min})^{SR}$	0.4838	0.2476	0.1128	0.0430	0.0151*	0.2669
$RV(\text{20min})^{SR}$	0.4815	0.2481	0.1088*	0.0427*	0.0154	0.2653
$RV(BR)^{SR}$	0.5106	0.2647	0.1132	0.0444	0.0156	0.2810
$BK(BR)^{SR}$	0.5299	0.2959	0.1141	0.0439	0.0157	0.2940
$ZMA(ZMA)^{SR}$	0.6298	0.5577	0.2105	0.0573	0.0174	0.3954
$ZMA(BR)^{SR}$	0.5623	0.3583	0.1236	0.0450	0.0162	0.3190
$BC(ZMA, ZMA)^{SR}$	0.5469	0.3292	0.1193	0.0446	0.0160	0.3072
$BC(ZMA, BR)^{SR}$	0.5183	0.2767	0.1103	0.0436	0.0156	0.2854
$FBK(BNHLS)^{SR}$	0.5182	0.2867	0.1109	0.0440	0.0156	0.2871
$FBK(BR)^{SR}$	0.5189	0.2847	0.1113	0.0437	0.0157	0.2871
$FCK(BNHLS)^{SR}$	0.5137	0.2770	0.1109	0.0437	0.0155	0.2836
$FCK(BR)^{SR}$	0.5149	0.2811	0.1112	0.0439	0.0155	0.2849
$FMTH(BNHLS)^{SR}$	0.5230	0.2901	0.1119	0.0438	0.0156	0.2898
$FMTH(BR)^{SR}$	0.5246	0.2908	0.1115	0.0434	0.0156	0.2904

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

Table 8: Put option pricing performance using different realized volatilities with the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	268	101	115	92	71	647
RMSE						
$RV(\text{1min})^{HL}$	25.2978	47.7543*	64.0820	62.4539	53.4289	47.1050
$RV(\text{5min})^{HL}$	26.4284	51.5841	65.4127	63.3466	53.7522	48.5066
$RV(\text{10min})^{HL}$	27.1575	52.8379	62.8731	62.9773	54.3032	48.2857
$RV(\text{15min})^{HL}$	27.2277	52.0127	59.7248*	60.5138*	54.4645	47.0068
$RV(\text{20min})^{HL}$	27.9302	54.0083	62.5752	63.1618	54.4194	48.6489
$RV(BR)^{HL}$	26.7606	52.6679	66.6746	64.7675	54.0437	49.3641
$BK(BR)^{HL}$	27.1972	53.7206	65.6152	64.3998	54.3149	49.3519
$ZMA(ZMA)^{HL}$	25.9143	48.5860	61.7137	61.3822	52.9438*	46.5508*
$ZMA(BR)^{HL}$	26.8523	53.2114	65.9695	64.0342	54.2988	49.2017
$BC(ZMA, ZMA)^{HL}$	25.0629*	49.0419	69.2279	65.8127	53.4596	49.1629
$BC(ZMA, BR)^{HL}$	25.7274	50.3171	66.5138	64.4565	53.4167	48.5761
$FBK(BNHLS)^{HL}$	25.6558	50.1531	67.3225	64.8069	53.1501	48.7658
$FBK(BR)^{HL}$	25.5370	50.0205	67.0645	64.7963	53.5238	48.6983
$FCK(BNHLS)^{HL}$	25.8919	51.3447	67.4543	64.6765	53.5202	49.0624
$FCK(BR)^{HL}$	25.6021	50.0879	66.9533	64.4833	53.2630	48.6056
$FMTH(BNHLS)^{HL}$	25.5656	50.5696	67.6425	64.8444	53.5274	48.9441
$FMTH(BR)^{HL}$	25.4036	50.0487	67.0296	64.3883	53.5993	48.5975
MAE						
$RV(\text{1min})^{HL}$	10.0727	27.5188	48.8267	42.1931	40.6430	27.6064
$RV(\text{5min})^{HL}$	10.3591	28.7595	48.6002	41.5706	40.6569	27.7915
$RV(\text{10min})^{HL}$	10.6429	28.5464	44.3286	39.6009	40.7957	26.8517
$RV(\text{15min})^{HL}$	11.0443	28.0901	41.8705*	37.5631*	40.9825	26.2406*
$RV(\text{20min})^{HL}$	11.0477	29.1348	43.1179	38.0190	41.0241	26.6962
$RV(BR)^{HL}$	10.3585	29.0898	49.3067	42.8011	40.8908	28.1690
$BK(BR)^{HL}$	10.5398	29.2449	47.0859	41.1528	40.9414	27.6448
$ZMA(ZMA)^{HL}$	10.0743	26.9184*	45.1734	39.2185	39.6322*	26.3301
$ZMA(BR)^{HL}$	10.3499	29.5055	48.0478	41.7745	41.0417	27.8772
$BC(ZMA, ZMA)^{HL}$	9.9812	29.0488	53.2391	45.1749	40.8443	29.0377
$BC(ZMA, BR)^{HL}$	10.0627	28.8747	50.0498	43.5453	40.6527	28.2247
$FBK(BNHLS)^{HL}$	10.0934	29.4657	51.3071	44.1896	40.4658	28.6243
$FBK(BR)^{HL}$	9.9535	29.1728	50.9314	44.1319	40.6515	28.4660
$FCK(BNHLS)^{HL}$	10.0949	29.9475	51.0931	44.0451	40.4868	28.6438
$FCK(BR)^{HL}$	9.9827	29.2740	50.9210	44.0326	40.4420	28.4549
$FMTH(BNHLS)^{HL}$	10.0022	29.7642	51.7559	44.3766	40.5523	28.7489
$FMTH(BR)^{HL}$	9.9257*	29.3487	51.1854	44.2691	40.6878	28.5506

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

Table 8: (Continued) Put option pricing performance using different realized volatilities with the Hansen and Lunde (2005a) adjustment

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	268	101	115	92	71	647
RMSPE						
$RV(\text{1min})^{HL}$	0.4996*	0.3335*	0.2071	0.0760	0.0197	0.3595
$RV(\text{5min})^{HL}$	0.5192	0.3612	0.2100	0.0760	0.0197	0.3752
$RV(\text{10min})^{HL}$	0.5020	0.3369	0.1926	0.0741	0.0199	0.3599
$RV(\text{15min})^{HL}$	0.5018	0.3357	0.1821*	0.0702*	0.0201	0.3585*
$RV(\text{20min})^{HL}$	0.5026	0.3340	0.1821	0.0722	0.0204	0.3588
$RV(BR)^{HL}$	0.5116	0.3625	0.2144	0.0782	0.0200	0.3715
$BK(BR)^{HL}$	0.5060	0.3603	0.2066	0.0765	0.0201	0.3671
$ZMA(ZMA)^{HL}$	0.5046	0.3413	0.1924	0.0716	0.0192*	0.3620
$ZMA(BR)^{HL}$	0.5131	0.3694	0.2065	0.0763	0.0201	0.3726
$BC(ZMA, ZMA)^{HL}$	0.5183	0.3770	0.2241	0.0804	0.0201	0.3786
$BC(ZMA, BR)^{HL}$	0.5033	0.3644	0.2104	0.0771	0.0197	0.3666
$FBK(BNHLS)^{HL}$	0.5340	0.3799	0.2159	0.0779	0.0195	0.3871
$FBK(BR)^{HL}$	0.5372	0.3761	0.2154	0.0780	0.0197	0.3883
$FCK(BNHLS)^{HL}$	0.5182	0.3743	0.2135	0.0776	0.0197	0.3770
$FCK(BR)^{HL}$	0.5281	0.3781	0.2148	0.0776	0.0196	0.3833
$FMTH(BNHLS)^{HL}$	0.5264	0.3768	0.2159	0.0782	0.0198	0.3823
$FMTH(BR)^{HL}$	0.5211	0.3765	0.2150	0.0775	0.0197	0.3791
MAPE						
$RV(\text{1min})^{HL}$	0.4021	0.2588	0.1595	0.0534	0.0149	0.2445
$RV(\text{5min})^{HL}$	0.4068	0.2758	0.1597	0.0527	0.0149	0.2491
$RV(\text{10min})^{HL}$	0.4097	0.2607	0.1433	0.0496	0.0149	0.2446
$RV(\text{15min})^{HL}$	0.4127	0.2552*	0.1361*	0.0471*	0.0151	0.2433
$RV(\text{20min})^{HL}$	0.4069	0.2586	0.1371	0.0473	0.0152	0.2417*
$RV(BR)^{HL}$	0.4087	0.2743	0.1620	0.0543	0.0150	0.2503
$BK(BR)^{HL}$	0.4086	0.2741	0.1536	0.0519	0.0151	0.2484
$ZMA(ZMA)^{HL}$	0.4040	0.2598	0.1465	0.0491	0.0144*	0.2425
$ZMA(BR)^{HL}$	0.4035	0.2802	0.1552	0.0526	0.0151	0.2476
$BC(ZMA, ZMA)^{HL}$	0.4087	0.2850	0.1727	0.0571	0.0152	0.2543
$BC(ZMA, BR)^{HL}$	0.3976*	0.2795	0.1619	0.0547	0.0150	0.2465
$FBK(BNHLS)^{HL}$	0.4151	0.2909	0.1662	0.0556	0.0149	0.2564
$FBK(BR)^{HL}$	0.4121	0.2869	0.1652	0.0556	0.0149	0.2544
$FCK(BNHLS)^{HL}$	0.4071	0.2905	0.1645	0.0553	0.0149	0.2527
$FCK(BR)^{HL}$	0.4062	0.2888	0.1649	0.0554	0.0149	0.2522
$FMTH(BNHLS)^{HL}$	0.4043	0.2900	0.1673	0.0559	0.0149	0.2521
$FMTH(BR)^{HL}$	0.4040	0.2882	0.1656	0.0557	0.0149	0.2513

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

Table 9: 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	268	101	115	92	71	647
RMSE						
GARCH						
Risk neutral	26.0156	54.6635	74.2943	69.1827	49.6875	51.7628
Duan	25.9641	54.3667	73.9599	68.8067	49.9518	51.5745
EGARCH						
Risk neutral	24.2628	57.9856	77.9481	67.8789	52.6027	52.9858
Duan	24.1578	57.9305	77.6406	67.7803	52.7248	52.8715
FIEGARCH						
Risk neutral	22.5543	50.5758	67.4635	62.6299	53.0315	47.8055
Duan	22.7415	50.3464	66.0049	61.5208	57.0599	47.7467
MAE						
GARCH						
Risk neutral	11.0711	36.4578	60.6267	48.5941	37.5032	32.0784
Duan	11.2023	36.3791	60.2807	47.8800	37.7829	31.9881
EGARCH						
Risk neutral	11.8561	43.4189	65.7360	47.5408	39.9573	34.5179
Duan	11.8745	43.3187	65.3666	47.4544	40.1580	34.4540
FIEGARCH						
Risk neutral	9.9522	35.5033	55.6567	42.0410	40.0367	29.9288
Duan	9.9815	34.9733	53.6757	40.2149	44.0664	29.6886
RMSPE						
GARCH						
Risk neutral	0.8403	0.6204	0.2918	0.0914	0.0178	0.6074
Duan	0.8708	0.6272	0.2915	0.0911	0.0180	0.6259
EGARCH						
Risk neutral	1.7426	0.8541	0.3204	0.0883	0.0190	1.1795
Duan	1.7593	0.8501	0.3192	0.0881	0.0191	1.1892
FIEGARCH						
Risk neutral	1.2416	0.6452	0.2679	0.0805	0.0192	0.8469
Duan	1.3006	0.6339	0.2591	0.0779	0.0204	0.8810
MAPE						
GARCH						
Risk neutral	0.5636	0.4363	0.2199	0.0646	0.0135	0.3513
Duan	0.5852	0.4376	0.2194	0.0639	0.0137	0.3603
EGARCH						
Risk neutral	1.0230	0.6041	0.2460	0.0623	0.0145	0.5722
Duan	1.0346	0.6020	0.2447	0.0622	0.0146	0.5765
FIEGARCH						
Risk neutral	0.7511	0.4658	0.2059	0.0551	0.0145	0.4299
Duan	0.7562	0.4559	0.1980	0.0525	0.0158	0.4288

"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.

Table 10: Put option pricing performance with and without removing significant jumps from realized volatility

	DOTM	OTM	ATM	ITM	DITM	Total
Sample size	268	101	115	92	71	647
RMSE						
RV	27.2277	52.0127	59.7248	60.5138	54.4645	47.0068
RV-Jump	27.1354	51.6801	61.4029	61.9356	54.1492	47.5322
MAE						
RV	11.0443	28.0901	41.8705	37.5631	40.9825	26.2406
RV-Jump	10.8016	28.0453	44.1756	39.7077	40.6776	26.8143
RMSPE						
RV	0.5018	0.3357	0.1821	0.0702	0.0201	0.3585
RV-Jump	0.4976	0.3518	0.1922	0.0736	0.0200	0.3595
MAPE						
RV	0.4127	0.2552	0.1361	0.0471	0.0151	0.2433
RV-Jump	0.4052	0.2632	0.1449	0.0504	0.0149	0.2435

RV is $RV(\text{15min})^{HL}$. RV-Jump is the difference between RV and significant jump, which is detected using the test statistic in (29) at the 0.1% significance level.

Figure 1: Realized volatility signature plots

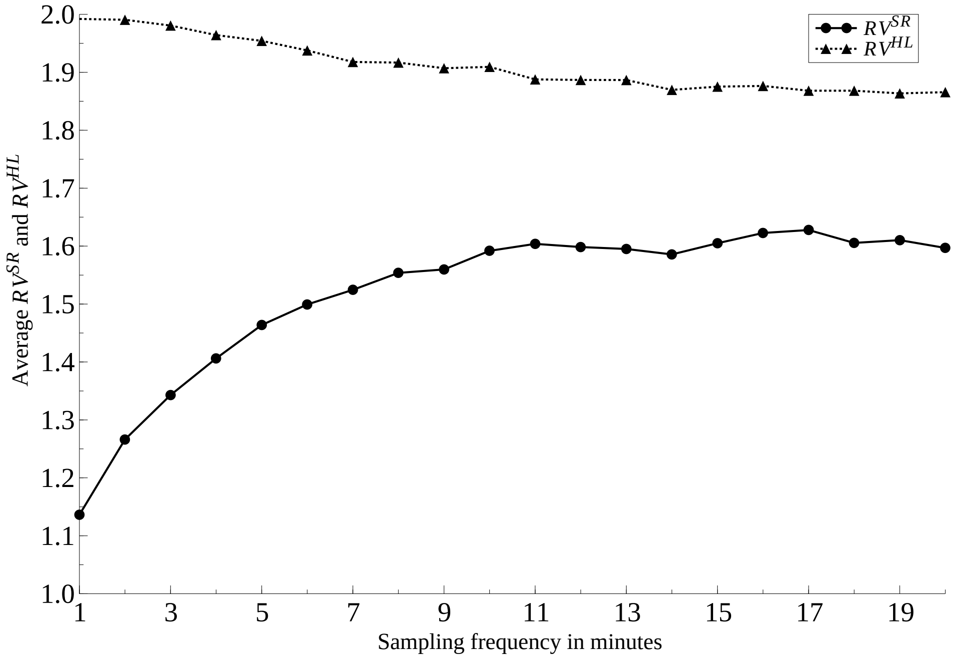
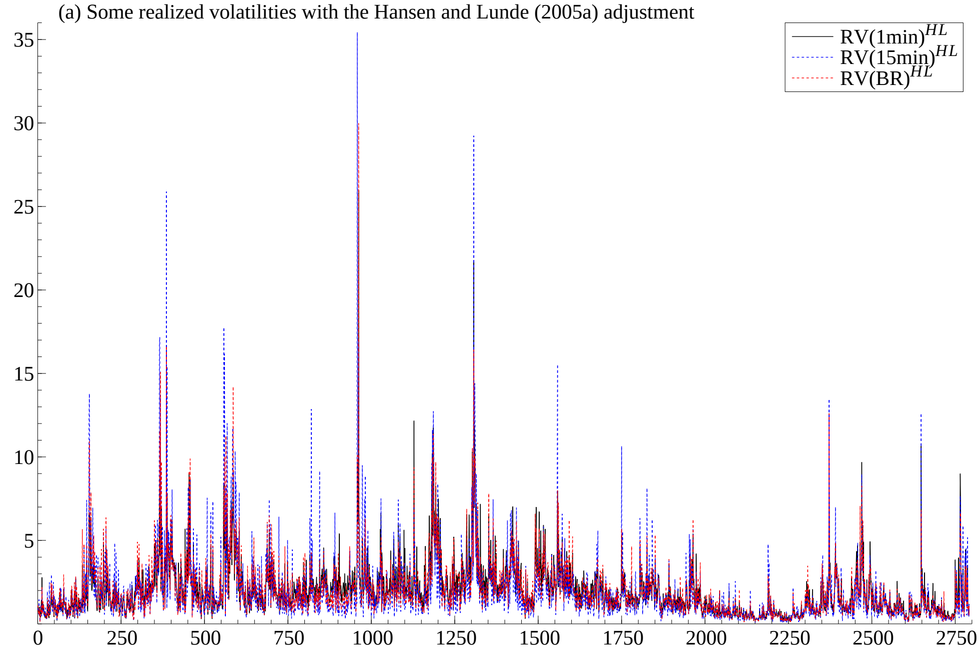


Figure 2: Realized volatility

(a) Some realized volatilities with the Hansen and Lunde (2005a) adjustment



(b) With or without the Hansen and Lunde (2005a) adjustment

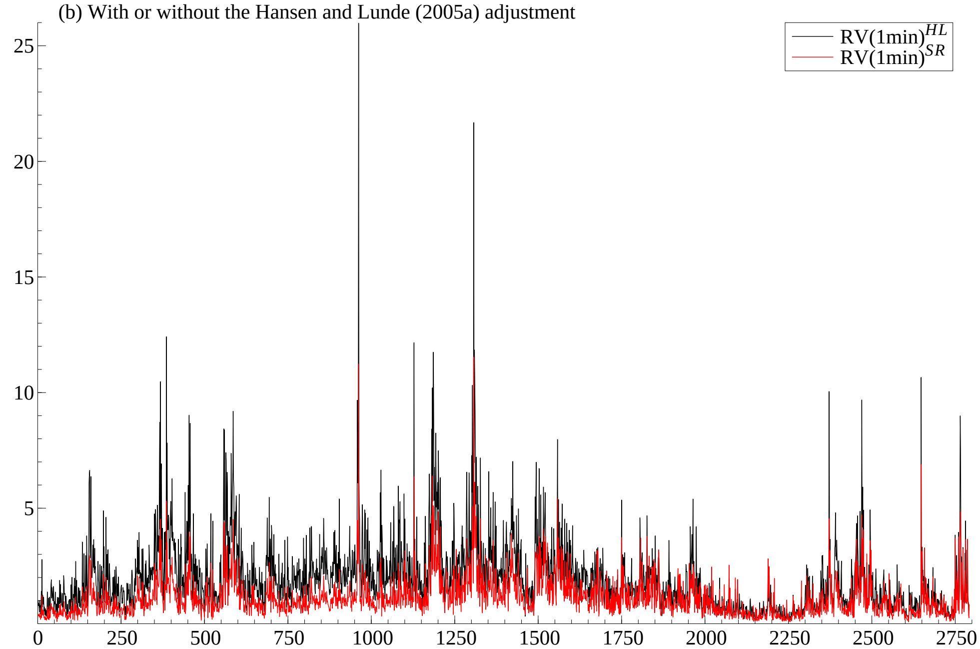
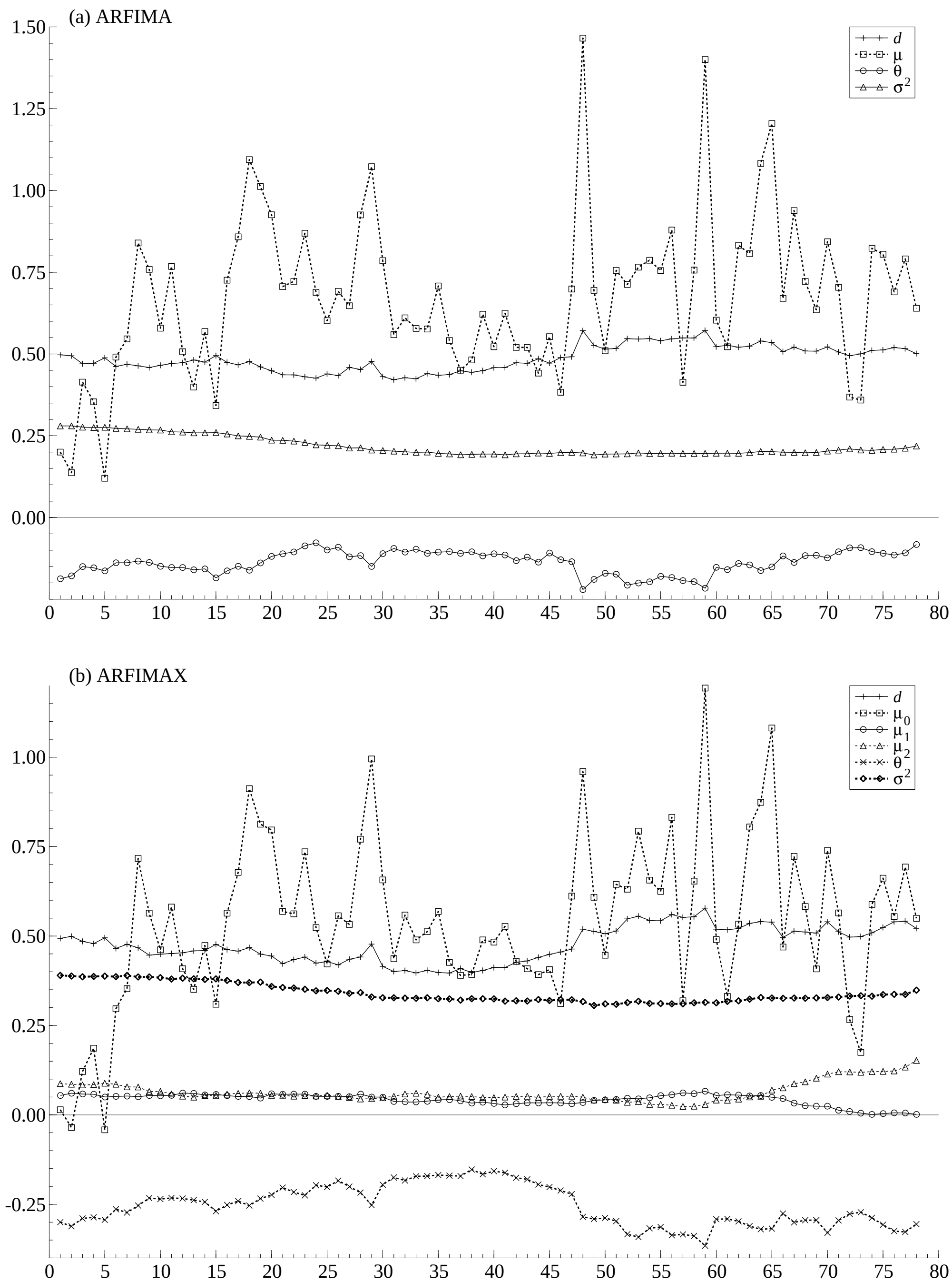
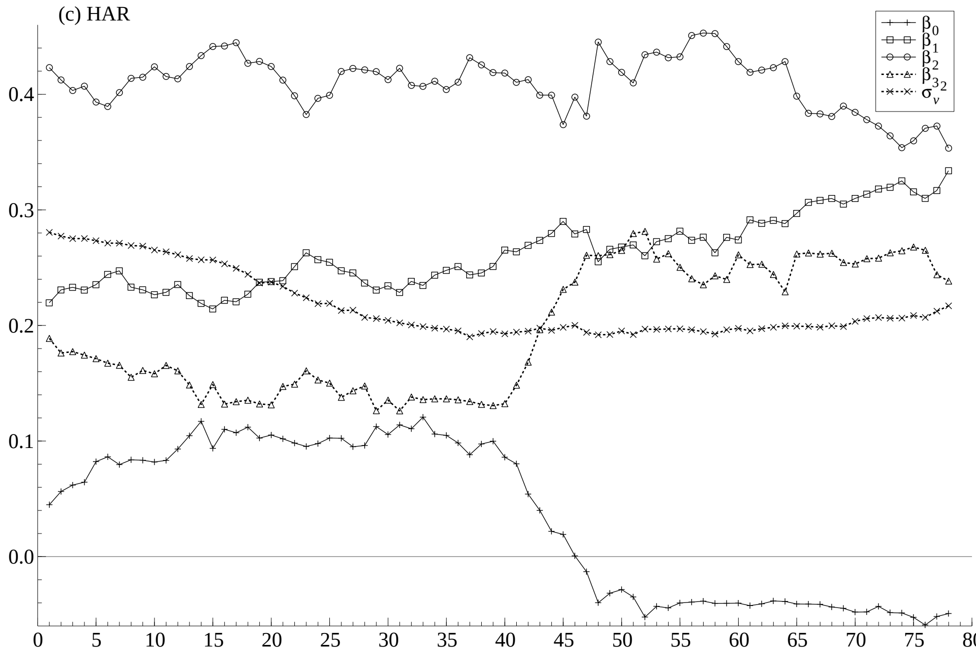
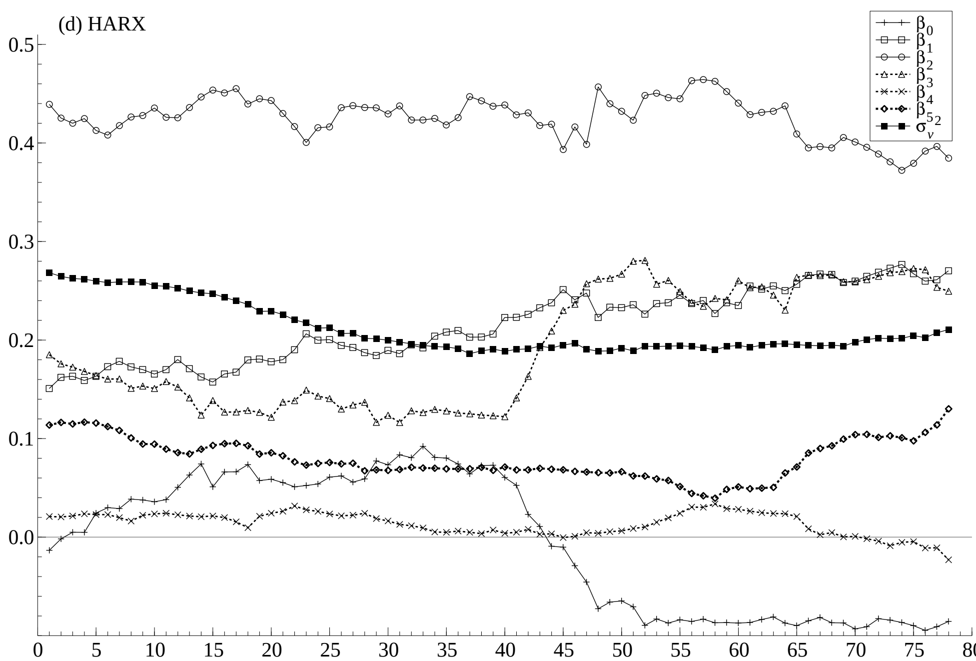


Figure 3: Parameter estimates

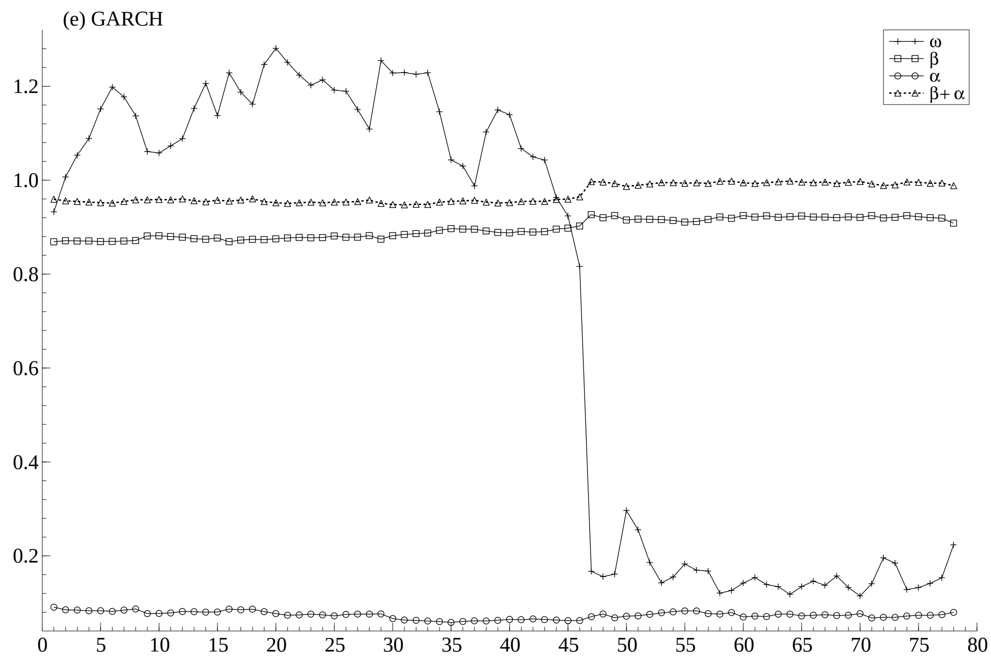


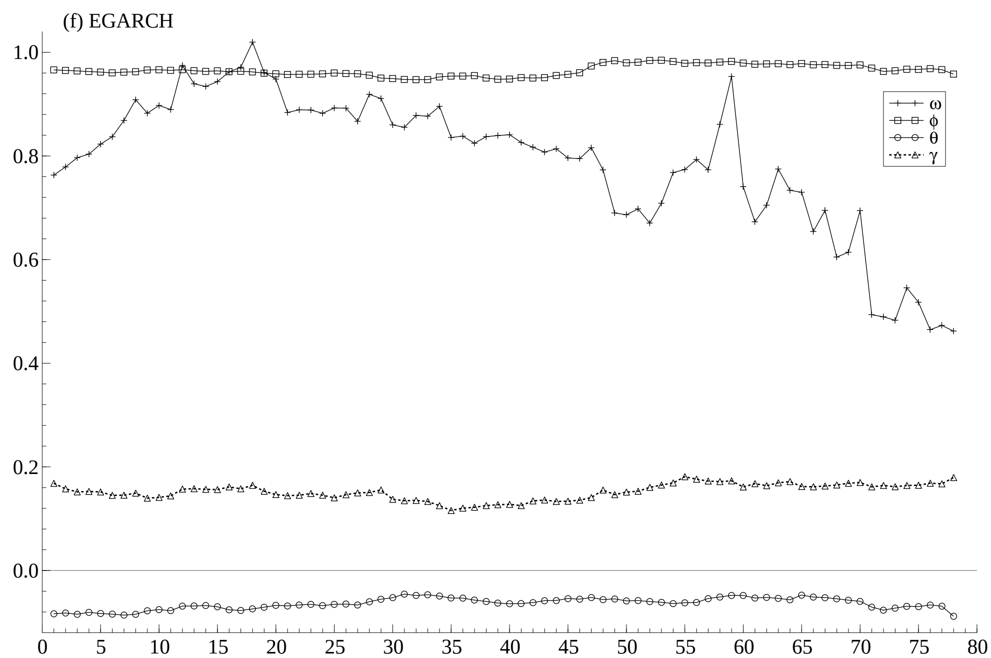


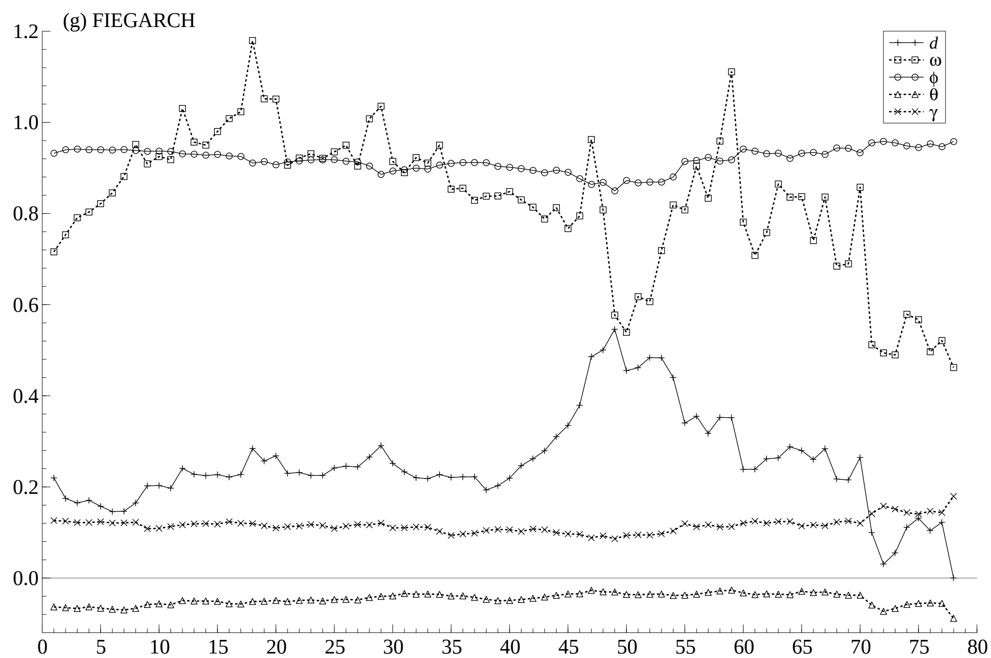
(c) HAR
β0
β1
β2
β3
σv2



(d) HARX
β0
β1
β2
β3
β4
β5
σv2


(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