CARF Working Paper

CARF-F-145

Modeling and Forecasting the Volatility of the Nikkei 225 Realized Volatility Using the ARFIMA-GARCH Model

Isao Ishida
University of Tokyo
Toshiaki Watanabe
Hitotsubashi University

January,2009

CARF is presently supported by AIG, Bank of Tokyo-Mitsubishi UFJ, Ltd., Citigroup, Dai-ichi Mutual Life Insurance Company, Meiji Yasuda Life Insurance Company, Mizuho Financial Group, Inc., Nippon Life Insurance Company, Nomura Holdings, Inc. and Sumitomo Mitsui Banking Corporation (in alphabetical order). This financial support enables us to issue CARF Working Papers.

Modeling and Forecasting the Volatility of the Nikkei 225 Realized Volatility Using the ARFIMA-GARCH Model*

ISAO ISHIDA†
Faculty of Economics and
Graduate School of Public Policy
University of Tokyo

TOSHIAKI WATANABE
Institute of Economic Research
Hitotsubashi University

This version: January 23, 2009
First draft: October 24, 2008

Abstract

In this paper, we apply the ARFIMA-GARCH model to the realized volatility and the continuous sample path variations constructed from high-frequency Nikkei 225 data. While the homoskedastic ARFIMA model performs excellently in predicting the Nikkei 225 realized volatility time series and their square-root and log transformations, the residuals of the model suggest presence of strong conditional heteroskedasticity similar to the finding of Corsi et al. (2007) for the realized S&P 500 futures volatility. An ARFIMA model augmented by a GARCH(1,1) specification for the error term largely captures this and substantially improves the fit to the data. In a multi-day forecasting setting, we also find some evidence of predictable time variation in the volatility of the Nikkei 225 volatility captured by the ARFIMA-GARCH model.

JEL classification: C22, C53, G15.

Keywords: ARFIMA-GARCH, Volatility of realized volatility, Realized bipower variation, Jump detection, BDS test, Hong-Li test, High-frequency Nikkei 225 data

*The authors are grateful to Yoshihiro Yajima, Yuji Sakurai, and the participants of the 2008 International Conference on High Frequency Data Analysis in Financial Markets at Hitotsubashi University for valuable discussions and comments. Ishida's work is partially supported by a grant-in-aid from the Zengin Foundation for Studies on Economics and Finance and Grants-in-Aid for Scientific Research 18203901 from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

†Corresponding author. Address: Faculty of Economics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033. Phone: +81-3-5841-5509. Fax: +81-3-5841-5521. Email: ishida@e.u-tokyo.ac.jp.

1 Introduction

Volatility plays key roles in the theory and applications of asset pricing, optimal portfolio allocation, and risk management. This fact, together with the development of econometric tools for volatility analysis and empirical evidence for the predictability of the volatility of numerous financial markets, spurred the phenomenal growth of the volatility literature as well as the birth of an entire financial risk management industry during the last quarter century. It is well-documented by now that time-variation in financial market volatility is highly predictable but stochastic, hence volatility itself is volatile. This paper empirically investigates whether the volatility of the Japanese stock market volatility is predictably time-varying. The obtained empirical results indicate that it is indeed time-varying with some predictable component.

The pace of progress at the frontier of volatility research has had at least four notable upsurges. The first followed the invention of the ARCH model by Engle (1982) and led to the development of the first set of econometric procedures for the empirical analysis of time-varying volatility (see, e.g., Bollerslev et al. 1994), and our deeper understanding of the empirical properties, e.g., volatility clustering, leverage effects in volatility, and fat-tails, of many financial time series. The second wave centered around the *stochastic volatility* (SV) modeling, which capitalized on and often contributed in turn to the concurrent development in the Bayesian statistical analysis using the Markov chain Monte Carlo procedure (see, e.g., Shephard 2005). This paper mainly concerns modeling and forecasting of the volatility of the *realized volatility*, and is part of the currently continuing third upsurge involving *realized volatility* measures, which was ignited by the recent availability of intraday financial data collected near or at the tick-by-tick frequency and the need to harness such high-frequency data fraught with microstructure noise and apparent short-term seasonalities before the rich information contained can be tapped into. The seminal paper by Andersen and Bollerslev (1998) defined the sum of squared intraday returns as the realized volatility (RV) for the day, and proposed to use it as a proxy for the ex post realization of the daily volatility. The squared daily return, typically used as a measure of the ex-post daily volatility in earlier volatility prediction studies, is a very noisy, albeit unbiased, proxy for the conditional variance. On the other hand, under ideal conditions in the absence of microstructure effects, the RV not only is an unbiased and much less noisy measure of the conditional variance but also converges in probability to the integrated variance over the measurement period as the sampling frequency increases to infinity if the asset price follows a diffusion process. Hence, RV may be considered an almost observable measure of volatility. A further case for the RV can be made based on the work of Hansen and Lunde (2006) and Patton (2005), who showed how the use of noisy proxies for the ex-post volatility such as the squared daily return may lead to a choice of an inferior volatility prediction model.

In a milestone paper, Andersen et al. (2003) used a Gaussian fractional VAR model, i.e., vector-ARFIMA model with MA order zero, for directly modeling and forecasting several exchange rate RV series, building on earlier empirical investigations that had found long-range dependence and approximate normality of the daily log RV time series constructed from high-frequency intraday data of stock prices (e.g., Andersen et al. 2001a) and exchange rates (e.g., Andersen et al. 2001a). They provided compelling empirical evidence for the superiority in predictive accuracy of this direct daily RV modeling approach over the daily returns approach with short- or long-memory GARCH-type models. Earlier papers including Andersen et al. (1999), Blair et al. (2001) and Martens (2001) that explored how to take advantage of intraday data within the GARCH framework also underlie the shift of the focus of the volatility literature to this approach. Ebens (1999) was also among the first to apply the ARFIMA model directly to RV time series. His ARFIMAX model for the RV of the DJIA index portfolio returns incorporated terms to capture the leverage effect, a well-documented stylized fact about equity returns. Koopman et al. (2005) conducted an extensive forecast performance comparison study of the ARFIMA model for the RV series of the S&P 100 stock index and the unobserved components RV (UC-RV) model of Barndorff-Nielsen and Shephard (2002) as well as more traditional GARCH-type and SV models based on daily returns and their implied-volatility-augmented versions, and reported that the ARFIMA model outperformed the other models although the performance of the UC-RV model was nearly as good.

Corsi (2004) proposed the heterogeneous autoregressive (HAR) model for the RV as an alternative to the ARFIMA model. The HAR-RV model employs a few predictor terms that are past daily RVs averaged over different horizons (typically a day, a week, and a month), and is capable of producing slow-decay patterns in autocorrelations exhibited by many RV series. Because of the ease in estimation and extendability of the baseline model, the HAR model has quickly become popular for modeling the dynamics of RV and other related volatility measures. Corsi et al. (2005), after finding strong conditional heteroskedasticity and non-Gaussianity in the ARFIMA and HAR residuals of the RV of the S&P 500 futures, introduced the HAR-GARCH-NIG model, which augments the basic HAR model with a GARCH volatility dynamic structure and NIG (normal inverse Gaussian) distributed standardized innovations. Subsequent papers by Bollerslev et al. (2005), Andersen et al. (2007a) among others also formulated the HAR errors as a GARCH process. Note that when the time series being modeled measures volatility, its volatility is related to the volatility of volatility of the primitive price process. The volatility of volatility of an asset price process is an important determinant of the tail property of the distribution of the asset's return, precise modeling of which is crucial, for example, for managing the extreme risk of a portfolio involving the asset, or pricing and hedging of out-of-the-money options written on the asset. Furthermore, a suite of volatility options and futures at the Chicago Board of Options Exchange (CBOE) and the CBOE Futures Exchange (CFE) written on U.S. equity market volatility indices, as well as a variety of over-the-counter volatility derivatives on major indices around the world, have been traded actively in recent years. For each of these products, the underlying is itself some measure of market volatility. As far as these volatility derivative products are concerned, the volatility of volatility is a second-moment property, rather than a fourth-moment property, of each of the respective underlying processes, and hence potential payoffs to accurate volatility-of-volatility modeling may be substantial.

In this paper, we use the ARFIMA model with the GARCH(1,1) specification for the error term (ARFIMA-GARCH), to empirically investigate the dynamic behavior of the daily RV of the Nikkei 225 index, with a particular focus on the conditional variance of the RV. The Nikkei 225 is the most widely watched indicator of the overall moves of the Japanese stock market, one of the largest in the world in terms of capitalization and trading volume. Shibata (2004, 2008), Shibata and Watanabe (2004), Watanabe (2005), Watanabe and Sasaki (2006), Watanabe and Yamaguchi (2006) among others studied the RV of the Nikkei 225 index or index futures, and reported empirical findings similar to those obtained for other major markets. Although neither a volatility index calculated and disseminated on a real-time basis by a major financial organization nor an exchange-traded volatility derivative has been introduced for any of the Japanese equity indices, Japanese-equity-related volatility derivatives have recently been traded actively over-the-counter. Several papers applied the ARFIMA-GARCH model to lower frequency macroeconomic and financial time series (e.g., Baillie et al. 1996b, Ling and Li 1997, Ooms and Doornik 1999), and a large number of papers in the RV literature employ the ARFIMA model without a conditionally heteroskedastic error specification to fit daily RV series (e.g., Oomen 2001, Giot and Laurent 2004 as well as those already referenced above). To the author's knowledge, this paper is the first to apply the ARFIMA-GARCH model to RV time series. Although several recent RV studies used the HAR-GARCH in place of the ARFIMA-GARCH model primarily for ease of estimation, it is not difficult to estimate a low-order ARFIMA or ARFIMA-GARCH(1,1) model by the maximum likelihood estimator. One advantage of the ARFIMA model is that it has the fractional integration parameter explicitly incorporated into the model, allowing one to estimate it jointly with the other parameters, a feature not shared by the HAR model, which is not formally a long-memry model.

Another notable recent development in the RV literature is the approach due to Barndorff-Nielsen and Shephard (2004, 2006) of decomposing the RV into the contribution of continuous sample path variation and that of jumps. Extending the theory of quadratic variation of semimartingales, Barndorff-Nielsen et al. (2006) provided an asymptotic statistical foundation for this decomposition procedure under very general conditions. Andersen et al. (2007a) used the HAR framework to study the roles of these two distinct components in RV prediction while Andersen et al. (2007b) documented improvements in the RV forecasting accuracy achieved by modeling these components separately. In light of these results, we estimate and remove the jump contributions from the daily Nikkei 225 RV using the Barndorff-Nielsen procedure modified by Andersen et al. (2007a), and study the continuous sample path variation. We find strong empirical evidence of conditional heteroskedasticity in the ARFIMA errors, and some evidence of predictability of the time variation in the volatility of the Nikkei 225 realized volatility.

The remainder of the paper is organized as follows. Section 2 briefly reviews the results from the theory of bipower variation and jump component extraction, Section 3 describes the data and summary statistics, Section 4 reviews the ARFIMA-GARCH model and resents the estimation results, and Section 5 reports the results of the RV and the variance of RV prediction exercises. Section 6 concludes.

2 Realized variance, realized bipower variation, and jump component extraction

The starting point of the realized volatility research is the recognition of the well-known result in the theory of continuous-time stochastic processes that the volatility of a process would be completely known if we observed a continuous record of the sample path of the process. Although in reality we do not obtain a continuous record and only observe the realized sample path of the process at discrete points in time, the sum of squared increments of the process approaches the integrated variance as the return measurement intervals shrink to zero. More precisely, if the process is a continuous semimartingale, under mild regularity conditions,

$$RV_t := \sum_{j=1}^{1/\Delta} |r_{t+j\Delta,\Delta}|^2 \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds \text{ as } \Delta \downarrow 0 \tag{1}$$

where $r_{t+j\Delta,\Delta}$ is the increment over the interval $[t+(j-1)\Delta, t+j\Delta]$ (in our context, the process is the log of the Nikkei 225 index level process so that $r_{t+j\Delta,\Delta}$ is the log return), σ_t is the diffusion coefficient (instantaneous volatility) of the process, time t has a daily unit so that RV_t is the tth day realized variance[1]. We will hereafter use the terms *realized volatility* or *realized variance* interchangeably, or their common abbreviation *RV*, to refer to RV_t defined in (1) and more loosely to other related measures such as $\sqrt{RV_t}$ or the realized bipower variation defined below. If the process is a semimartingale with finite-activity jumps, i.e., only a finite number of jumps occurring in any finite time interval, such as Poisson jumps, then the realized variance converges to the quadratic variation, which can be decomposed into the integrated variance (the continuous sample path variation) and the sum of squared jump sizes:

$$RV_t \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds + \sum_{t<s\leq t+1} \kappa_s^2 \text{ as } \Delta \downarrow 0 \tag{2}$$

where κ_s is the size of the jump occurring at time s. Barndorff-Nielsen and Shephard (2004, 2006) showed that even in the presence of jumps the *realized bipower variation*

$$BV_t := \frac{\pi}{2} \sum_{j=2}^{1/\Delta} \left| r_{t+j\Delta,\Delta} \right| \left| r_{t+(j-1)\Delta,\Delta} \right| \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds \tag{3}$$

holds under mild conditions, and proposed to use

$$RV_t - BV_t \xrightarrow{p} \sum_{t<s\leq t+1} \kappa_s^2 \tag{4}$$

or

$$J_t^* := (RV_t - BV_t)^+ \tag{5}$$

as an estimator for the sum of realized squared jumps $\sum_{t<s\leq t+1} \kappa_s^2$. J_t^* is known to take non-zero, small values very frequently due to measurement errors and due possibly to the presence of jumps of infinite-activity types. Based on the asymptotic distributional theory for these quantities developed by Barndorff-Nielsen and Shephard (2004, 2006) and Barndorff-Nielsen et al. (2006) and an extensive simulation study by Huang & Tauchen (2005), Andersen et al. (2007a) introduced what they call a shrinkage estimator for the jump contribution

$$J_t := I(Z_t > \Phi_a) \cdot (RV_t - BV_t) \tag{6}$$

where I is an indicator function, $Z_t := \frac{(RV_t - BV_t) RV_t^{-1}}{\sqrt{\left((\pi/2)^2 + \pi - 5\right) \max\left(1, TQ_t BV_t^{-2}\right) \Delta}}$ is asymptotically standard normally distributed under the null of no jumps, $\mu_1 := \sqrt{2/\pi}$, $\Phi_a := \Phi(a)$ is the standard normal distribution function, and the (standardized) *realized tripower variation*

$$TQ_t := \Delta^{-1} 4^{-1} \pi^{3/2} \Gamma\left(\frac{7}{6}\right)^{-3} \sum_{j=3}^{1/\Delta} \left| r_{t+j\Delta,\Delta} \right|^{4/3} \left| r_{t+(j-1)\Delta,\Delta} \right|^{4/3} \left| r_{t+(j-2)\Delta,\Delta} \right|^{4/3} \xrightarrow{p} \int_t^{t+1} \sigma_s^4 ds \text{ as } \Delta \downarrow 0 \tag{7}$$

[1] Time t here is loosely used in two different ways: RV_t refers to the tth day RV and $[t, t+1]$ the trading hours (or just either the morning or the afternoon session) of the tth day. Our notation here also glosses over the existence of a lunch break.

The last convergence result holds even in the presence of jumps. a is usually set to values such as .999 so that J_t picks up only "significant" jumps. With J_t, another estimator of the continuous sample path variation:

$$C_t := RV_t - J_t \tag{8}$$

may be used in place of BV_t.

In this paper, we use the microstructure-noise-robust versions of BV_t and TQ_t, due also to Andersen et al. (2007a). In these versions, the summands are respectively $(1-2\Delta)^{-1}\left|r_{t+j\Delta,\Delta}\right| \left|r_{t+(j-2)\Delta,\Delta}\right|$ and $(1-4\Delta)^{-1}\left|r_{t+j\Delta,\Delta}\right|^{4/3}\left|r_{t+(j-2)\Delta,\Delta}\right|^{4/3}\left|r_{t+(j-4)\Delta,\Delta}\right|^{4/3}$, skipping an interval of length Δ in sampling short-period returns. The definitions of J_t and C_t are modified accordingly as well.

3 Data and summary statistics

3.1 Calculation of five-minute returns from minute-by-minute Nikkei 225 data and RV measures

Nihon Keizai Shinbun, Inc. (Nikkei) computes and disseminates the Nikkei 225 index once every minute during the trading hours of the Tokyo Stock Exchange (TSE) (9:00-15:00 with a 11:00-12:30 lunch break). In this paper, we use minute-by-minute Nikkei 225 index data provided directly by Nikkei and maintained by the Center for Advanced Research in Finance at the University of Tokyo. The sample period spans March 11, 1996 through August 31, 2007. From the minute-by-minute data, we construct a series of five-minute log differences multiplied by one hundred, which we call "five-minute (percentage) returns." This choice is made to strike a balance between alleviating the microstructure-related noise and increasing the precision of volatility measurement, following the standard practice of the RV literature in handling high-frequency intraday data from highly liquid markets. Andersen et al. (2000) and the other empirical papers dealing with the RV of the Nikkei 225 cash index or futures referenced in the introduction also used five-minute returns. For further discussions of the miscrostructure-related issues, see Hansen and Lunde (2005) and references therein.

The official minutely Nikkei 225 index starts at 9:01 for the morning session, and the first five-minute return in a given day that we calculate is for the 9:05-9:10 interval as in Andersen et al. (2000). Removal of the first four observations partially alleviates possible effects of the sluggish response of the Nikkei 225 to the information accumulated overnight (or over a weekend or holiday) on volatility measurement. Given the TSE trading rules, the 9:05 index value is unlikely to have fully impounded all overnight information in it when overnight domestic and overseas events move the overall Japanese stock market up or down by a large amount relative to the previous close[2], and again our choice here is an attempt for a balance between higher precision and lower bias. For the afternoon session, the index

[2] The TSE trading rules do not allow the price of each individual stock to immediately move to a level outside of the range set according to either the last transaction or the "special quote" price. When there are buy (sell) orders remaining at the upper (lower) limit of the range, the TSE announces it as a "special quote" and wait for five minutes before shifting the range upward (downward). The special quotes of the constituent individual stocks, if there are any, are used in place of the more stale last transaction prices in the calculation of the Nikkei 225 index.

starts at 12:31. Since the effect of the TSE trading rules hampering the Nikkei 225 index from quickly reflecting the information accrued over a lunch break is likely to be much milder, the 12:31-12:35 interval is retained and used in our calculation of the first afternoon five-minute return.

For the end price of the last five-minute return calculation of each session, we use the closing price of the session. Due to such factors as delayed arrival of individual stock price data from the TSE and the real-time nature of the Nikkei 225 calculation and dissemination, the final few observations of each session are occasionally marked by time stamps up to several minutes later than 15:00 (or 11:00). For sessions with such observations as well, we use the last recorded observation of the session for closing the last five-minute interval. In total, there are 53 five-minute returns for a typical trading day, 23 from the morning session and 30 from the afternoon session. We exclude sessions from half trading days including the first and last trading days of each year from our sample, retaining 2,802 trading days.

Since there is intervening lunch break, RV_t, BV_t, TQ_t, J_t, and C_t are calculated for the morning and afternoon sessions separately and the two components (denoted with subscripts am and pm respectively) for each variable are added for the day, e.g., $RV_t = RV_{am,t}+RV_{pm,t}$. $RV_t^* := RV_t+R_{n,t}^2+R_{l,t}^2$ may be used to define the RV for the day, where $R_{n,t}$ is the previous-day-close-to-open (15:00-9:05) overnight return and $R_{l,t}$ is the lunch break (11:00-12:30) return. Since the TSE is open only 4.5 hours a day, however, it would be a stretch to treat RV_t^* as an observed realization of the volatility for the whole day. We therefore concentrate on the trading-hour RV measures in this paper, and leave the analysis of the other components and their impacts on the trading-hour RV measures for future research.

3.2 Properties of the realized volatility and related measures

Summary statistics for various returns and RV measures are presented on Tables 1. In addition to the sample skewness and kurtosis, the Jarque-Bera (JB) statistic for testing normality[3] is presented for each series. For checking temporal dependence, the first-order sample autocorrelation and the Ljung-Box statistics of orders 5, 10, and 22 (corresponding to roughly one week, two weeks, and a month) for no serial correlations up to their respective orders are shown for each series. Since the usual Bartlett's standard error, $T^{-1/2} = 0.019$, is biased under heteroskedasticity, the heteroskedasticity-adjusted standard error for the first-order autocorrelation and Ljung-Box statistics due to Diebold (1988) are also presented. Previous empirical studies have documented that daily volatility measures such as the daily return squared, the absolute daily return, and various daily RV measures of financial time series appear to have long-memory properties. For long memory processes, the influence of shocks does not last forever unlike in the case of integrated processes, but decays very slowly relative to the geometrically fast decay for short-memory processes. Formally, there are several definitions of long memory. The usual definition of long memory for a covariance stationary time series, which we adopt

[3] It is designed to jointly measure the deviations of the sample skewness and kurtosis from their respective population values, zero and three, under normality. For an $i.i.d.$ normal series, it is asymptotically distributed $\chi^2(2)$ (two degrees of freedom). The simulation work of Thomakos and Wang (2003) has shown, however, that it grossly overrejects the null of normality if the data are a sample path of a long-memory process. Hence, it presented here as an informal statistic.

in this paper, is that

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| = \infty \tag{9}$$

holds where $\gamma(k)$ is the kth order autocovariance. Another definition of long memory for a covariance stationary process is that $\gamma(k)$ decays hyperbolically, i.e.,

$$\gamma(k) \sim k^{2d-1} l(k) \tag{10}$$

as $k \to \infty$, where $l(\cdot)$ is some slowly varying function and $d \in (0, 1/2)$ is called the *long-memory parameter*. If the process is a covariance stationary one satisfying some regularity conditions, (10) implies (9). For a review of alternative definitions and their relations to each other, see, e.g., Palma (2007).

Before estimating the ARFIMA and the ARFIMA-GARCH models for the RV series, we estimate the long-memory parameter d for our various series via two popular semiparametric estimators $\widehat{d}_{GPH}$, $\widehat{d}_{Robinson}$, due respectively to Geweke-Porter-Hudak (1983) (GPH) and Robinson (1995a), with the bandwidth parameter m set at $m = T^{.7}$ where T is the sample size. The asymptotic standard errors for $\widehat{d}_{GPH}$ and $\widehat{d}_{Robinson}$ are respectively $\pi/\sqrt{24m}$ and $1/(2\sqrt{m})$, and hence the latter is asymptotically more efficient for a given m. d is equal to the fractional differencing parameter if the process is a fractionally integrated one, reviewed in the next section. The definition of the fractional differencing parameter can be extended to the nonstationary region $d \geq 1/2$. The region of the true d over which $\widehat{d}_{GPH}$ and $\widehat{d}_{Robinson}$ are consistent and asymptotically normal extends beyond $1/2$. For these properties, neither $\widehat{d}_{GPH}$ nor $\widehat{d}_{Robinson}$ requires Gaussianity, and the latter does not require conditional homoskedasticity. See Robinson (1995b), Velasco (1999), Deo and Hurvich (2001), and Robinson and Henry (1999).

Looking at the summary statistics for $RV_t = RV_{am,t} + RV_{pm,t}$, the unconditional distribution seems to be highly nonnormal with very large positive values of sample skewness and kurtosis. The LB statistics indicate that it is highly significantly serially correlated. The value of the first-order sample autocorrelation, .544, is at a medium-persistent level and well below one, but the values of $\widehat{d}_{GPH}$ (.470 with standard error .040) and $\widehat{d}_{Robinson}$ (.468 with standard error .0311) are significantly positive but below the stationary/nonstationary border of 1/2, indicating that autocorrelations decay slowly. Note, however, that $\widehat{d}_{GPH}$ and $\widehat{d}_{Robinson}$ are within two standard errors from 1/2.

Deviations from normality seem to be vastly reduced by the square-root transformation, but remain large. The log transformation brings down the sample skewness and kurtosis values for each series even further and close to zero (-.130 for RV_t) and three (3.234 for RV_t) respectively. After transformation, the first-order sample autocorrelation increases to .663 ($\sqrt{RV_t}$) and .713 ($\ln RV_t$), and $\widehat{d}_{GPH}$, $\widehat{d}_{Robinson}$ increase to values in excess of 1/2 (.557, .524 for $\sqrt{RV_t}$ and .584, .533 for $\ln RV_t$). Other than our point estimates of d being in the nonstationary region, these results are roughly in line with earlier empirical findings about RV measures constructed from high-frequency intraday exchange rate and stock returns data, which led to the popularity of the Gaussian ARFIMA model; See, e.g., Andersen et al. (2003).

Turning next to the Barndorff-Nielsen decomposition of RV_t into the contribution of squared jumps and that of continuous sample path variation, we set $a = .999$ in (6). Summary statistics of the square-root and logarithmic transformed ($\ln(1+J_t)$ for J_t) transformed series are also presented on Table 1. The sample mean of $I_t := \max(I(Z_{am,t} > \Phi_a), I(Z_{pm,t} > \Phi_a))$, which is an estimate of the unconditional jump probability over the trading hours of a day, is .150, implying a little more than one jump occurrences per week[4]. On average, the jump contribution J_t comprises about 4% $(= .045/1.087)$ of RV_t over trading hours. Similarly to the results of previous studies on the S&P 500 futures and other financial time series (e.g., Andersen et al. 2007a, Andersen et al. 2007b), J_t and its transformations seems distinctly less persistent although this does not rule out the possibility of the unobserved conditional jump probability series (as opposed to the realized jump series J_t) being more persistent. And the standard deviation of J_t (.154) is not negligible relative to that of C_t (1.03). Hence, using C_t, purged of the jump component with a different dynamic behavior, may reveal a higher-resolution picture of the dependence structure of C_t although C_t and RV_t exhibit quite similar features as summarized by the statistics on Table 1. For our analysis via the ARFIMA-GARCH model, we use C_t[5].

4 Modeling and forecasting the conditional mean and the conditional variance of the RV with the ARFIMA-GARCH model

4.1 The ARFIMA-GARCH model and its estimation

The ARFIMA model, introduced by Granger and Joyeux (1980), and Hosking (1981), is a natural extension of the ARIMA model for parsimoniously modeling time series with long memory. An ARFIMA(p, d, q) process $\{Y_t\}$ may be defined as a causal solution to

$$\phi(L)(1-L)^d (Y_t - \mu) = \psi(L)\varepsilon_t \tag{11}$$

where $\mu \in \mathbb{R}$, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, $\psi(L) = 1 + \psi_1 L + \cdots + \psi_q L^q$ are respectively the AR and MA operators, sharing no common roots, $(1-L)^d$ is the fractional differencing operator, and $\{\varepsilon_t\} \sim WN(0, \sigma^2)$, $0 < \sigma^2 < \infty$. If all roots of $\phi(z)$ and $\psi(z)$ lie outside of the unit disk and $d \in (-1, 1/2)$ holds, then there is a unique covariance stationary solution, which is invertible[6]. For $d \in (0, 1/2)$, the autocovariance function of this solution satisfies $\lim_{k\to\infty} \gamma(k)/[ck^{1-2d}] \to 1$ where c is a constant, and hence the process possesses long memory and the fractional integration parameter d corresponds to the long-memory parameter d in (10). Also of note is that an ARFIMA process with $d \geq 1/2$ is nonstationary but still mean-reverting as long as $d < 1$ (See Baillie 1996, p.21). If $d \in [1/2, 1)$, we may interpret $\{Y_t\}$ as a process, starting from some finite past and satisfying (11), which becomes a stationary ARFIMA($p, d-1, q$) process after being differenced once.

---

[4] Using a less stringent value $a = .99$, the estimated jump probability becomes .302, or once per 3.3 trading days on average.

[5] It turns out that the ARFIMA-GARCH estimation results for C_t are similar to those for RV_t. Details of the results obtained for RV_t are available upon request.

[6] Bondon and Palma (2007) recently proved invertibility for $d > -1$, relaxing the condition $d > -1/2$ often cited in the literature.

Following Ding et al. (1993) that found extremely slow decay patterns in the sample autocorrelation functions of daily volatility measures such as absolute returns, features meant to capture the long-range dependence in volatility have been incorporated into GARCH-type models (e.g., the FIGARCH model) and stochastic volatility models; see, e.g., Baillie et al. (1996a), Bollerslev and Mikkelsen (1996), Breidt et al. (1998), Deo and Hurvich (2003). Since daily RV time series including our Nikkei 225 RV are series of observed quantities which appear to have long-memory properties, the ARFIMA model is a natural modeling choice. The ARFIMA process (11) may take negative values, and hence it is strictly speaking defective as a model for $Y_t = C_t$ or $\sqrt{C_t}$ although we do not encounter negative forecasts in our empirical application. Negativity is not an issue for $\ln C_t$. Also of note is that the class of continuous-time processes discussed in Section 2 in general does not give rise to (11) exactly, let alone the ARFIMA process with $\{\varepsilon_t\} \sim i.i.d.N\left(0, \sigma^2\right)$.

Baillie et al. (1996b) extended the ARFIMA model to include a GARCH specification for conditional heteroskedasticity and used it to analyze the inflation rate time series from the G7 and three other high inflation countries, and Koopman et al. (2007) applied a model they called the periodic seasonal Reg-ARFIMA-GARCH to the daily time series of spot electricity prices. In this paper, we apply the ARFIMA-GARCH(1,1) model in which $h_t := E_{t-1}\left[\varepsilon_t^2\right]$, the conditional variance of ε_t with respect to the sigma-field $\sigma\left(Y_{t-1}, Y_{t-2}, \cdots\right)$, is given the following formulation:

$$h_t = \omega + \beta h_{t-1} + \alpha \varepsilon_{t-1}^2 \tag{12}$$

For the stationary ARFIMA model with Gaussian homoskedastic errors, Sowell's (1992) algorithm for exact maximum likelihood estimation is available. For the ARFIMA-GARCH model, however, no closed-form expression for the exact likelihood function is available. Hence, we employ the conditional sum of squares (CSS) estimator:

$$\widehat{\theta} := \underset{\theta}{\operatorname{argmax}} L\left(\theta\right) \tag{13}$$

where $\theta := \left(\mu, \theta_1'\right)'$, $\theta_1 := \left(\phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_q, d, \omega, \alpha, \beta\right)'$ and

$$L\left(\theta\right) := -\frac{1}{2}\sum_{t=1}^{T}\left(\ln h_t\left(\theta\right) + Z_t\left(\theta\right)^2\right) \tag{14}$$

$$\begin{aligned} Z_t\left(\theta\right) &:= \varepsilon_t\left(\theta\right) / \sqrt{h_t\left(\theta\right)} \\ \varepsilon_t\left(\theta\right) &:= \psi\left(L\right)^{-1}\phi\left(L\right)\left(1-L\right)^d\left(Y_t - \mu\right) \\ h_t\left(\theta\right) &= \omega + \beta h_{t-1}\left(\theta\right) + \alpha\varepsilon_{t-1}\left(\theta\right)^2 \end{aligned} \tag{15}$$

with the presample values of $Y_t - \mu$, $t = 0, -1, -2, \cdots$ set to zero and $h_1\left(\theta\right)$ set to some initial value. If $d \geq 1/2$, Y_t does not have an unconditional mean and μ cannot be interpreted as such. For this case, $\left(1-L\right)^d\left(Y_t - \mu\right) = \left(1-L\right)^{d-1}\Delta Y_t$ where $\Delta Y_t := Y_t - Y_{t-1}$, but note that $\Delta Y_1 = Y_1 - Y_0 = Y_1 - \mu$ under our choice of the presample values so that μ does not disappear. We equate $h_1\left(\theta\right)$ to the value of the sample variance as is often done in the estimation of GARCH-type models. In estimation, we do not impose $d < 1/2$ for stationarity. Neither do we impose $\alpha + \beta < 1$ for the GARCH specification. We estimate the ARFIMA with a homoskedastic specification, i.e., $\omega = \sigma^2$, $\alpha = \beta = 0$, by the CSS estimator as well for comparability of results across homoskedastic and GARCH specifications. The objective function maximized by the CSS estimator is essentially the Gaussian maximum likelihood (without the constant $-\frac{T}{2}\ln\left(2\pi\right)$) for the AR(∞) representation of the ARFIMA-GARCH model with the conditional distribution of ε_t specified as $\mathcal{N}\left(0, h_t\right)$, conditional on the presample values and h_1. Hence, the CSS estimator for the ARFIMA-GARCH model is a long-memory analogue of what we usually refer to as the quasi-maximum likelihood estimator, QMLE, in the short-memory ARMA-GARCH setting. Extending the results of Beran (2004) for the ARFIMA model, Ling and Li (1997) showed that the CSS estimator is $\sqrt{T}$-consistent and asymptotically normal for the ARFIMA (p,d,q)-GARCH(P,Q) model. Their results are also valid for the nonstationary case in which the true value of d is larger than 1/2. This is a desirable property particularly because d estimates are found to be near or greater than 1/2 in our semiparametric estimation as well as other studies of financial time series data, and is another justification for using the CSS estimator. Ling and Li (1997) derived $\sqrt{T}$-consistency for $\widehat{\theta}_1$ assuming that μ is known. To our knowledge, their results have not been rigorously verified to be applicable to $\widehat{\theta}$. However, it seems reasonable to expect that $\left(T^{1/2-d}\widehat{\mu}, T^{1/2}\widehat{\theta}_1'\right)'$ has a Gaussian limiting distribution as long as the other conditions of Ling and Li (1997) are satisfied and $d < 1/2$ [7]. Another caveat is that the asymptotic results of Ling and Li (1997) are based on the assumption $E\left[\varepsilon_t^4\right] < \infty$. If $\{\varepsilon_t\}$ follows a GARCH(1,1) process and the standardized error sequence $\left\{Z_t := \varepsilon_t h_t^{-1/2}\right\}$ is $i.i.d.$ with $\kappa := E\left[Z_t^4\right] < \infty$, this condition amounts to $\kappa\alpha^2+2\alpha\beta+\beta^2 < 1$. Our estimation results indicate that a covariance stationarity condition $\alpha+\beta < 1$, let alone the more stringent condition $E\left[\varepsilon_t^4\right] < \infty$, may not be satisfied by our untransformed series C_t. However, this is not likely to invalidate estimation and inference based on Ling and Li (1997). Lee and Hansen (1994) established (local) consistency and asymptotic normality of the Gaussian QMLE for the GARCH(1,1) model when the true values of α, β may not be in the covariance stationary region $\alpha + \beta < 1$ but are in the strict stationary region satisfying $E\left[\ln\left(\beta + \alpha Z_t^2\right)\right] < 0$. Jensen and Rahbek (2004), focusing on the estimation of (α, β) and working with a different set of assumptions on $\{Z_t\}$ proven that the consistency and asymptotic normality extend to the case of conditional variance explosion $E\left[\ln\left(\beta + \alpha Z_t^2\right)\right] > 0$[8].

Given the strong evidence to be reported shortly against conditional normality of the standardized error Z_t except when Y_t is $\ln C_t$, we also present the robust standard errors of Bollerselv and Wooldridge (1992) (BW) interpreting our estimator as a QMLE. For the models with the GARCH specification, the BW standard errors are robust to distributional misspecification of Z_t under correct specification of the conditional mean and variance and regularity conditions. We use the BIC[9] as our model selection criteria and confine our search for the best model to a total of 64 models: The full ARFIMA(2,d,2)-GARCH(1,1) model and its 63 restricted versions (at least one of the ARFIMA parameters, $\phi_1, \phi_2, \psi_1, \psi_2, d$, is fixed at zero and/or no conditional heteroskedasticity $(\alpha, \beta) = 0$). Let us denote, for example, the ARFIMA(2,d,0) model with the first-order AR coefficient restricted to be zero as the ARFIMA({2},$d, 0$). We denote the other models with second-order terms similarly.

---

[7]For the case of $d \geq 1/2$, Ling and Li (1997) considered estimating the parameters in $\phi(L)(1-L)^{d-m}((1-L)^m Y_t - \mu) = \psi(L)\varepsilon_t$ with $\mu \neq 0$ unknown and estimated separately from the other paramters, where m is the smallest positive integer such that $d - m < 1/2$. If $m = 1$, this implies that $\{Y_t\}$ has a linear time trend, which is counterintuitive in our case where Y_t is a measure of volatility.

[8]Unlike in our case, however, they both assumed $Y_t = \mu + \varepsilon_t$ with either μ a constant to be estimated jointly with the other parameters (Lee and Hansen 1994) or $\mu = 0$ known (Jensen and Rahbek 2004). To our knowledge, rigorous asymptotic theory for the ARFIMA-GARCH model with $\alpha + \beta \geq 1$ is not yet available in the literature.

[9]$BIC = -2L^*\left(\widehat{\theta}\right) + N\ln T$ where N is the number of parameters, $L^*\left(\widehat{\theta}\right)$ is the value of the maximized log likelihood function $L^*\left(\widehat{\theta}\right) = L\left(\widehat{\theta}\right) - T\ln(2\pi)$, and $L(\theta)$ is as defined in (14).

When modeling a time series of daily currency or stock returns, the conditional mean is small relative to the variance so that setting it to zero or a constant instead of fitting a more elaborate model usually does not much affect the estimation of the volatility process. Since the RV is highly persistent, it is essential to adequately model the conditional mean part of the RV even if one's primary interest is in the conditional variance of the RV. Otherwise, misspecification in the mean may masquerade as conditional heteroskedasticity. See Diebold and Nason (1990) for a study focusing on this issue.

4.2 ARFIMA-GARCH estimation results

We first use the entire sample (2802 daily observations) available to us to estimate the full ARFIMA(2,d,2)-GARCH(1,1) model and 63 restricted versions. For order selection, we employ the BIC. In applying the ARFIMA-GARCH model, we work with transformed series $\ln C_t$ and $\sqrt{C_t}$ as well as the original series C_t. It should be emphasized that, although the results for all three series are presented and discussed together and some comparisons in terms of model adequacy are made across different transformations, each series is being investigated on its own right, and that the BIC values and other specification test statistics and forecast performance measures are not directly comparable across the three series. We should also keep in mind that the square-root and log transformations alter the nature of conditional heteroskedasticity[10]. It is beyond the scope of this article to address the issue of which series should be used as the LHS variable of the ARFIMA-GARCH model in a forecast exercise with a particular loss function[11] although our results contain information relevant to this issue. Table 3 summarizes the parameter estimation results with Hessian-based and BW standard errors in parentheses. For brevity, only the results for the best short-memory model, the best homoskedastic model and its GARCH(1,1) counterpart, and the best overall model and its homoskedastic counterpart. Table 4 presents summary statistics and some specification test statistics for the residuals of these models.

4.2.1 Estimation results for the ARFIMA model with the constant error variance specification

In this subsection, we discuss our empirical results for the ARFIMA model with the constant error variance specification and various order restrictions. The results are similar to those previously reported in the literature for the RVs of the Nikkei 225 cash index and index futures and other financial time series. The order selected by the BIC is the ARFIMA(0,d,1) for all series. The estimated values of d in the BIC-selected homoskedastic ARFIMA models are 0.494 (C_t), 0.516 $\left(\sqrt{C_t}\right)$, 0.509 $(\ln C_t)$. At usual levels, all of these are significantly higher than zero, but none are significantly different from the nonstationarity boundary of 1/2.

Table 4 presents some residual diagnostic statistics. The first-order sample serial correlations (the

[10] This can be most easily seen in the related context of a continuous-time CEV diffusion process for the instantaneous variance V_t per unit time. After the square-root or log transformation, the diffusion coefficient (volatility of variance) of the original process, αV^{λ} where α, $\lambda \in \mathbb{R}_+$, becomes respectively $\frac{1}{2}\alpha V^{\lambda-1/2}$ or $\alpha V^{\lambda-1}$.

[11] Suppose, for example, that our goal is to minimize the mean squared errors in forecasting RV_t one-step-ahead. Since $\exp\left(E_{t-1}\left[\ln RV_t\right]\right) \neq E_{t-1}\left[RV_t\right]$ by Jensen's inequality, we need to specifiy more than the first two condtional moments of $\ln RV_t$ to produce an optimal forecast for RV_t if we are to work with $\{\ln RV_t\}$ instead of directly with $\{RV_t\}$.

Bartlett's standard error is $T^{-1/2} = 0.019$ and the heteroskedasticity-adjusted standard errors are given in parentheses) in the residuals are all smaller than 0.01 in absolute magnitude and insignificant at usual levels. Judging by the values of the heteroskedasticity-adjusted LB statistics due to Diebold (1988) (the first line) of orders from one through 5, 10, or 22 for the residuals, serial correlations in C_t, $\sqrt{C_t}$, and $\ln C_t$ are adequately filtered out by the best homoskedastic ARFIMA models[12].

The heteroskedasticity-adjusted LB statistics for the residuals of the BIC-selected short-memory models are only marginally significant for the raw series C_t and insignificant for the two transformed series. The best short-memory ARMA model for each series mobilizes three to four out of the four ARMA terms allowed and, in the case of $\sqrt{C_t}$ or $\ln C_t$, seem to fit the linear dependence structure as successfully as the best homoskedastic ARFIMA model[13]. However, the estimates of the long-memory parameter d, when not restricted to be zero, are significantly greater than zero and the BIC also favors the long-memory models.

The very large values of sample skewness and kurtosis of the residuals from the raw series C_t shown on Table 4 suggest that the unconditional distribution of the error term is highly nonnormal. The square-root transformation of C_t vastly reduces the values of these measures of nonnormality, but still far above those of a normal distribution. The log transformation brings the residual distributions closer to normality, but the sample kurtosis values are still nearly 4.

In sum, it appears that parsimonious ARFIMA models are reasonably successful in removing serial correlations although some evidence of remaining serial correlations is found for the residuals of the models of the raw series C_t.

The discrepancies between the values of the unadjusted LB statistics and those of the heteroskedasticity-adjusted ones already indicate the presence of conditional heteroskedasticity in the form of serial correlations in ε_t^2. Next we directly test the null of no serial correlations in ε_t^2 by calculating the Ljung-Box statistics of $\widehat{\varepsilon}_t^2$, which in this context are often called the McLeod-Li (ML) statistics[14]. The ML statistics for the squared residuals from the selected models are all very large, cleanly rejecting the null of no serial correlations of orders up to 5, 10, and 22 at usual levels. With this evidence, we now turn to modeling conditional heteroskedasticity with the ARFIMA model augmented by the GARCH(1,1) specification for the error term.

---

[12] The original LB stastistics, shown on the second line, apparently indicate that, for the raw series C_t, there is remaining serial correlations. In the presence of conditional heteroskedasticity, however, the Bartlett's standard error is not a consistent estimator of the standard deviations of the sample serial correlations. In particular, if the squared series is positively autocorrelated, the Bartlett's standard error overestimates the estimation precision of sample autocorrelations and as a result the LB statistics overreject the null. See Diebold (1988).

[13] To save space, only the results for the best models selected from among all short-memory models are shown. The best short-memory model happens to include the GARCH specification for all three series. Similar results are obtained, however, when the best models are selected from among the homoskedastic models.

[14] McLeod and Li (1983) showed that under the null of a homoskedastic ARMA model the asymptotic distribution of the ML statistic for no serial correlations in ε_t^2, constructed using sample serial correlations of k different orders in the squared residuals $\widehat{\varepsilon}_t^2$, is $\chi^2(k)$, unaffected by paramter estimation. This contrasts the case of the LB statistic applied to $\widehat{\varepsilon}_t$, which is $\chi^2(k-p-q)$, where $p+q$ is the number of the ARMA coefficients estimated.

4.2.2 Estimation results for the ARFIMA-GARCH(1,1) model

We next compare all 64 versions of the ARFIMA(2,d,2)-GARCH(1,1) model including those without the constant error variance restriction $\alpha = \beta = 0$. For each of the three series, the BIC selects a model with the GARCH(1,1) specification, more specifically, the ARFIMA($\{2\}$,d,1)-GARCH(1,1) for C_t and the ARFIMA(0,d,1)-GARCH(1,1) for $\sqrt{C_t}$ and $\ln C_t$. The estimation results for the selected model and its homoskedastic version for each series are shown on Table 3. As expected from the LB statistics for the squared residuals from the conditionally homoskedastic models, the improvement in the log likelihood value achieved by giving the GARCH(1,1) specification for the error process is substantial for each series. The addition of the GARCH specification to the homoskedastic model of the same ARFIMA order does not alter the ARFIMA parameter estimates very much, and consequently the distributions of the residuals $\widehat{\varepsilon}_t := \varepsilon_t\left(\widehat{\theta}\right)$ remain similarly nonnormal.

The point estimate of the volatility-of-the-RV persistence measure $\alpha + \beta$ in the best model for the raw series C_t is well in excess of the covariance stationarity threshold of one although not significantly so. The BW standard errors of the GARCH parameters are rather large, in fact several times larger than the Hessian-based ones and renders even $\widehat{\alpha} = 0.263$ insignificantly different from zero. The values of the sample skewness and kurtosis and the JB normality test statistic of the standardized residuals $\widehat{Z}_t := \widehat{\varepsilon}_t / \widehat{h}_t^{-1/2}$ where $\widehat{h}_t = h_t\left(\widehat{\theta}\right)$ (shown on Table 4) indicate very high degrees of conditional nonnormality of the error terms of the models for the raw series C_t[15]. This at least partially explains the large discrepancies between the BW and the Hessian-based standard errors and suggest that substantial efficiency gains may be obtained by adopting a non-Gaussian QMLE, but non-Gaussian QMLE requires additional conditions for ensuring consistency under distributional misspecification; See Newey and Steigerwald (1997). As a check of the GARCH(1,1) specification, we calculate the LB statistics for no correlations in the squared standardized residuals (Table 3). The values of the LB statistics are greatly reduced by standardization although they are still high enough to reject the null whether the degrees-of-freedom adjustment is applied or not[16].

As for the best models for $\sqrt{C_t}$ and $\ln C_t$, both α and β are estimated to be significantly above zero, but the estimates of $\alpha + \beta$ do not exceed one, although for $\ln C_t$, it is very close to one. For $\ln C_t$, the α estimate is rather small, which together with $\alpha + \beta$ estimated to be nearly one implies a slowly varying conditional mean process as shown; See Figure. Again, the values of the LB statistics for the squared residuals are substantially reduced by standardization of the residuals by $\widehat{h}_t^{1/2}$, but are still large enough to reject the null of no serial correlations. For $\sqrt{C_t}$, the degree of nonnormality in the residuals is hugely reduced but not to near normal levels. For $\ln C_t$, the effects of standardization

---

[15]Kulperger and Yu (2005) proved that the asymptotic distribution of the Jarque-Bera-type moment-based distributional test statistic based on $\widehat{Z}_t$ rather than Z_t is $\chi^2(2)$, unaffected by parameter estimation if the conditional variance is correctly specified and $\{Z_t\}$ is an i.i.d. sequence. One of their additional assumptions is that the conditional mean of the observed variable Y_t is zero, which is not satisfied by our ARFIMA-GARCH case.

[16]Unlike in the case of the McLeod-Li test statistics for no serial correlations in the squared raw residuals, parameter estimation does affect the asymptotic distribution in this case. Bollerslev and Mikkelson (1996) suggested a heuristic adjustment of reducing the degrees of freedom of the χ^2 distribution by the number of estimated parameters. (When the number of lag orders is 5, this adjustment obviously cannot be applied). Li and Mak (1994) proposed a more elaborate test that correctly accounts for the impact of parameter estimation.

are not as large, but the raw residuals are not highly nonnormal to begin with. Consistent with near normality, the discrepancies between the BW and the Hessian-based standard errors are not very large for $\ln C_t$. However, the value of the JB statistic still rejects the null of normality easily.

All in all, the addition of the GARCH(1,1) specification for the ARFIMA error term helps to capture much of the serial correlations in the squared residuals though not completely.

4.3 Further specification tests

The LB statistics are for testing the null of no remaining serial correlations in $\{Z_t\}$ and $\{Z_t^2\}$, and are not designed to detect more general forms of temporal dependence. In this paper, we only attempt to model the first two conditional moments by the ARFIMA-GARCH model. By estimating them using the Gaussian QMLE, we take a stand on neither the higher-order dependence structure nor the shape of the conditional distribution (beyond the zero-mean and unit-variance property) of $\{Z_t\}$. Nevertheless, it is of interest to run a battery of tests for detecting violations of the assumption of $\{Z_t\} \sim i.i.d.$ or $\{Z_t\} \sim i.i.d.\ \mathcal{N}(0,1)$ as a specification check[17].

The first columns of Table 4 show the values of the BDS nonlinearity test statistics due to Brock et al. (1996), which test the null of $\{Z_t\} \sim i.i.d.$ All different pairs of two consecutive segments of a fixed length k are taken from an observed time series and the number of cases in which the distance (using a particular measure) between the two segments in a pair is shorter than a preset value ϵ is counted. After normalization, this number becomes the BDS statistic with (k, ϵ), which is asymptotically distributed $\mathcal{N}(0,1)$ under the null of $\{Z_t\} \sim i.i.d.$ De Lima (1996) showed that, under some conditions (including a mixing condition and $\sqrt{T}$-consistency of the estimator), the BDS test applied to $\ln \widehat{Z}_t^2$, where $\widehat{Z}_t$ is the standardized residual of an ARCH model, is asymptotically free of the influence of parameter estimation, or "nuisance-parameter-free"[18]. Note that if $\{Z_t\} \sim i.i.d.$, then $\{\ln Z_t^2\} \sim i.i.d$ holds as well. We apply the BDS test to the log standardized residuals $\ln \widehat{Z}_t^2$. Following Chen and Kuan (2005), we set ϵ equal to 0.75 times the sample standard deviation of $\ln \widehat{Z}_t^2$, and for normalization use the bootstrap standard deviation of the raw statistic based on the 1,000 resamples from the empirical distribution of $\widehat{Z}_t$. The results for $k = 2, \cdots, 5$ are presented (the first lines). We also present the BDS statistics calculated with $\widehat{Z}_t$ (the second lines)[19]. There is strong evidence

[17] To our knowledge, for all of our diagnostic statistics, asymptotic theory available in the literature assumes some form of mixing for $\{Y_t\}$ and/or $\sqrt{T}$-consistency of the parameter estimator for establishing the invariance of the asymptotic test statistic distribution in the presence of parameter estimation or justifying the adjustment when invariance does not hold. $\sqrt{T}$-consistency of $\widehat{\mu}$ and mixing properties cannot be expected to hold for long memory processes. In fact, Gaussian long-memory ARFIMA processes are known to be non-strong-mixing (Viano et al. 1995). Strictly speaking, the diagnostic tests in this subsection as well as the more traditional ones used in the previous subsections therefore remain somewhat informal.

[18] Caporal et al. (2005) investigated by Monte Carlo simulations the finite-sample size properties of the BSD test statistics under GARCH(1,1) DGPs with various combinations of true parameter values (estimated by the Gaussian QMLE) and distributions of Z_t not necessarily satisfying De Lima's (1996) sufficiency conditions and reported that they are well-behaved for $T \geq 1000$.

[19] One of De Lima's (1996) sufficient conditions for invariance is violated in this case. The impact of parameter estimation on the variance of the statistic, however, is accounted for by the bootstrap although we still presume that the asymptotic normality of the BSD statistics extends to this case.

for nonlinear dependence for the homoskedastic models, in particular for the raw series C_t, but when we allow the GARCH specification to be chosen, evidence for nonlinear dependence becomes much weaker. In particular for the best model for $\ln C_t$, the BDS statistics are all insignificant except for the one with $k = 5$ for $\widehat{Z}_t$[20].

We next conduct a test of time reversibility due to Chen et al. (2000) (CCK). $\{Z_t\} \sim i.i.d.$ implies time reversibility, which in turn implies that the unconditional distribution of $Z_t - Z_{t-k}$ is symmetric around the origin. The CCK statistic tests this symmetry and is calculated as $\xi_k/\widehat{\sigma}_k$, where $\xi_k := (T-k)^{-1/2} \sum_{t=k+1}^{T} \xi(Z_t)$ and $\widehat{\sigma}_k$ is a consistent estimator for the standard deviation of ξ_k. There are a variety of functions that can be chosen as $\xi(\cdot)$. Following Chen and Kuan (2005), we use $\xi(Z_t) := \gamma Z_t / \left(1 + \gamma^2 Z_t^2\right)$, replace Z_t with $\widehat{Z}_t$, calculate $\widehat{\sigma}_k$ by bootstrap similarly to the case of the BDS statistics[21], and present the results for $k = 1, 2, 3$ and $\gamma = 0.5, 1$ on Table 4. A general tendency is that the values of the CCK statistics decrease after square-root transformation and become insignificantly or only marginally significant for $\ln C_t$.

For testing the correctness of the specification of the joint distribution of the series in its entirety, we may use the nuisance-parameter-free Hong-Li statistics, which formalize the popular graphical approach of Diebold et al. (1998). In our Gaussian ARFIMA-GARCH case, the test is equivalent to jointly testing the independence and standard normality of $\{Z_t\}$ (as opposed to the JB statistic that tests the normality under the maintained hypothesis of independence). The Hong-Li statistics are based on the observation that, under the null of correct model specification in its entirety (as opposed to just the first two conditional moments), the probability integral transformed series $\{U_t\}$ implied by the model is a sequence of $i.i.d.$ uniform $[0,1]$ random variables, and in particular the joint density of (U_t, U_{t-k}) should be $f(u_1, u_2) = 1$ over $[0,1] \times [0,1]$. Hong and Li (2005) showed that, under the null, a properly normalized measure of the distance (call it the Hong-Li statistic of order k, $Q_{HL}(k)$) between $f(u_1, u_2) = 1$ and $\widehat{f}(u_1, u_2)$, a nonparametric estimate of $f(u_1, u_2)$ constructed using $\left\{\widehat{U}_t\right\}$ (the probability integral transformed series implied by the estimated model), is asymptotically distributed $\mathcal{N}(0,1)$[22]. Furthermore, they showed that the asymptotic distribution of $(Q_{HL}(1), \cdots, Q_{HL}(K))$ is standard multivariate normal and hence that of a portmanteau statistic $W_{HL}(K) := \sqrt{K} \sum_{k=1}^{K} Q_{HL}(K)$ is $\mathcal{N}(0,1)$ under the null. Noting that negative values of $Q_{HL}(K)$ occur only under the null if the sample size is sufficiently large[23], they suggest using the upper-tailed

[20] As discussed in the previous subsection, the LB statistics detected serial correlations in the squared residuals of the best model for $\ln C_t$. The BDS statistics, however, keep a wider watch on various types of misspecification and may be less powerful than other specialized tests in finding specific forms of misspecification.

[21] Chen and Kuan (2005) showed that, unlike in the case of the BDS statistic, the impact of parameter estimation on the CCK statistic is of the same stochastic order as the CCK statistic calculated from the true $\{Z_t\}$ provided that the parameter estimator is $\sqrt{T}$-consistent. Hence, correcting the asymptotic variance is crucial here, and boostrap $\widehat{\sigma}_k$ serves the purpose. A caveat is that it might turn out to be the case that using the ARFIMA-GARCH standardized residuals, the effect of the error in estimating μ asymptotically dominates the other terms.

[22] For the choice of the kernel function and the bandwidth parameter involved in nonparametric density estimation, we follow Hong and Li (2005). Their expression of $Q_{HL}(K)$ has a typo, which is corrected in Egorov et al. (2006, Footnote 11).

[23] For this reason, the portmanteau statistic is designed as a scaled sum rather than a scaled sum of squares that would yield a $\chi^2(K)$ statistic.

critical values for individual $Q_{HL}(K)$'s and $W_{HL}(K)$. $Q_{HL}(1)$, $W_{HL}(5)$, $W_{HL}(10)$, $W_{HL}(22)$ are shown on Table 4. The overall pattern across the series and the models is similar to the case of the JB statistics for the standardized residuals, and as expected, the null of $i.i.d.$ normality is very strongly rejected for C_t and $\sqrt{C_t}$. However, the Hong-Li statistics values are much reduced for $\ln C_t$, and $Q_{HL}(1)$ only marginally rejects the null for $\ln C_t$ at the 5% level (Note that the upper-tailed 5% critical value is 1.645)[24].

Overall, the results of the above tests suggest that the addition of the GARCH(1,1) specification to the ARFIMA model goes a long way toward accounting for temporal dependence in the RV. In particular for $\ln C_t$, the Gaussian ARFIMA-GARCH(1,1) model appears to be a reasonably good approximation of the dynamic and distributional structures of the data generating process. However, the "separate inference" statistics, a set of nuisance-parameter-free statistics for detecting possible sources of misspecification, also proposed by Hong and Li (2005), do reveal some strong evidence of remaining higher-order dependence even for $\ln C_t$. For given m and l, we first calculate cross-correlations of all orders $j \geq 1$ (up to a truncation point, which we set to be $j = 44$) in $\widehat{U}_t^m$ and $\widehat{U}_{t-j}^j$ and take a weighted average of the squares of them, which after normalization becomes the Hong-Li separate inference statistic $M(m,l)$, asymptotically distributed $\mathcal{N}(0,1)$ under the null of correct specification of the entire joint density of the series. Table 4 shows $M(m,l)$ for $(m,l) = (1,1), (2,2), (3,3), (4,4), (1,2)$, and $(2,1)$, meant to detect autocorrelations in level, volatility, skewness, kurtosis, ARCH-in-mean, and leverage in U_t respectively. As expected, the GARCH specification substantially reduces the value of $M(2,2)$ in every case. But it is rather surprising, for example, to find the very high value of $M(4,4)$, 28.74, for the best model of $\ln C_t$ with nearly normal $\widehat{Z}_t$.

5 Forecasting the RV and the variance of the RV

5.1 RV

For one-step prediction of the RV measures, we use

$$\widehat{Y}_{t+1|t} := \widehat{\mu} + \sum_{s=1}^{t} \widehat{\pi}_s \left(Y_{t+1-s} - \widehat{\mu}\right) \tag{16}$$

where $\widehat{\mu}$ is the CSS estimate of the unconditional mean and $\widehat{\pi}_s$ are the coefficients in the AR(∞) expansion of the ARFIMA model, implied by the ARFIMA parameter estimates[25]. We also evaluate the performance of the selected model for each of the RV series in forecasting the k-days-ahead RV,

[24]As discussed earlier, the JB statistic rejects the normality of Z_t even for $\ln C_t$. However, recall that the JB statistic tests the normality under the maintained hypothesis of $\{Z_t\} \sim i.i.d.$, whereas the Hong-Li statistics in our context test the independence and standard normality jointly. If the $i.i.d.$ assumption does not hold as indicated by the LB statistics for the squared standardized residuals, the JB statistic may overreject the null.

[25]While the Durbin-Levinson algorithm may be applied to calculate the best linear one-step predictor based on the finite past, the formula (16) is more in line with our CSS estimator.

Y_{t+s}. For this, we use $\widehat{Y}_{t+s|t}$ as our k-days-ahead forecast where

$$\widehat{Y}_{t+k|t} := \widehat{\mu} + \sum_{s=1}^{k-1} \widehat{\pi}_s \left(\widehat{Y}_{t+s|t} - \widehat{\mu} \right) + \sum_{s=0}^{t-1} \widehat{\pi}_{k+s} \left(Y_{t-s} - \widehat{\mu} \right) \tag{17}$$

For evaluating predictive accuracy, we mainly look at R^2 from the Mincer-Zarnowitz regression of the realization of the target variable on the prediction.

The results of the in-sample RV prediction exercise for horizons $k = 1, 5, 10$, and 22 days in which the model parameters are estimated once using the entire sample are summarized on the left half of Table 5. In conformance with the previously reported results for the RVs of the Nikkei 225 index and other financial time series, the RV is highly predictable. For example, R^2 is nearly 60% when the target is the one-step-ahead $\ln C_t$. Although R^2 tapers off as the horizon increases, predictability of the 22-days-ahead daily RV is still substantial (nearly 30% for $\ln C_t$). For $\sqrt{C_t}$ and $\ln C_t$, the estimates of the Mincer-Zarnowitz regression intercept and coefficient are close to zero and one respectively, and the Wald test statistics do not reject the null of forecast unbiasedness, i.e., the intercept being zero and the coefficient being one, for any of the investigated horizons. Excellent performance of the ARFIMA-GARCH model can be visually confirmed for each series by the time series plot of the ARFIMA-GARCH fit; See the left panels of Figure. The results of "out-of-sample" forecasting in which the first 1,500 daily observations of the sample are used for a one-shot estimation of the parameters of the ARFIMA-GARCH model and the remaining 1,302 observations are used for forecast evaluation are on the right half of Table 5[26]. Again, the results indicate that the RV is highly predictable, and, for $\sqrt{C_t}$ and $\ln C_t$, there is little evidence of forecast bias.

5.2 Variance of the RV

We next evaluate the performance of the ARFIMA-GARCH model selected for each of the three series in predicting the volatility of the RV. For one-step-ahead forecasting, we use $\widehat{h}_{t+1}$ as the volatility-of-the-RV forecast and $\widehat{\varepsilon}^2_{t+1} = \left(Y_{t+1} - \widehat{Y}_{t+1} \right)^2$ as the proxy for the target. Our ARFIMA-GARCH estimation results indicate that the variability of $\widehat{\varepsilon}^2_{t+1}$ is much higher than that of $Var_t \ (Y_{t+1}) = E_t \left[\varepsilon^2_{t+1} \right]$, which would lead to an apparently low R^2 value of the Mincer-Zarnowitz regression even if the time-variation in $E_t \left[\varepsilon^2_{t+1} \right]$ could be well approximated by the ARFIMA-GARCH model. This parallels the resolution by Andersen and Bollerslev (1998) of a puzzle regarding the low R^2 phenomenon prevalent in earlier volatility prediction studies based on data sampled at daily or lower frequencies. The introduction of the RV measures in empirical finance has rendered financial volatility nearly observable, but here we are back in the unobservability realm. What makes the problem even more complicated in our context is that the RV is highly predictable unlike daily asset returns and that $Var \left(E_t \left[Y_{t+1} \right] \right) / Var \left(Y_{t+1} \right)$ seems to be large and hence the effect of the deviation of $\widehat{Y}_{t+1}$ from the true $E_t \left[Y_{t+1} \right]$ due to possible model misspecification, parameter estimation errors and other factors on the Mincer-Zarnowitz R^2 is potentially more serious for volatility-of-volatility prediction than for equity or currency return volatility prediction based on daily return observations. For forecasting the

[26]The model selection is based on the full sample because there is a concern that use of a poor RV forecasting model may cause spurious predictability of the volatility of the RV. Hence, our forecasting exercise is not truely out-of-sample.

multi-day volatility of the RV, we use $k^{-1}\sum_{s=1}^{k}\widehat{h}_{t+s|t}$, where $\widehat{h}_{t+s|t}$ is the s-periods-ahead single-day conditional variance of the RV implied by the ARFIMA-GARCH, as our forecast, and the daily squared errors averaged over the horizon,

$$\upsilon_t(k) := k^{-1}\sum_{s=1}^{k}\left(Y_{t+s} - \widehat{Y}_{t+s|t}\right)^2 \tag{18}$$

as a proxy for the target.

The left half of Table 6 summarizes the in-sample forecasting performance evaluation results for $k = 1, 5, 10,$ and 22. As expected, R^2's are in fact low, but not negligible. Multi-day-ahead single-day volatility-of-the-RV predictability does seem to increase as we attempt to forecast further into the future (For brevity, details are not reported on the table). In spite of this, the R^2 values are higher for multi-day average forecasts than for a one-day ahead forecast. This is not surprising because our target proxy $\upsilon_t(k)$, being squared forecast errors aggregated over forecast horizons of a week to several weeks, is a sort of realized volatility (of the RV), albeit on a coarser frequency. A noise reduction effect akin to that of the daily RV in daily volatility forecasting studies seem to start to kick in as the forecast horizon increases. For C_t, $\sqrt{C_t}$, the R^2 values are 3.55% ($k = 1$), 7.60% (5), 6.50% (10), 5.27% (22). Presumably, the effect of decreasing multiperiods-ahead single-day predictability begins to outweigh the noise reduction effect at around a weekly horizon. For $\ln C_t$, the R^2 values are 3.31% ($k = 1$), 7.29% (5), 9.30% (10), and 12.50% (22). Evidence for out-of-sample predictability appears to be much weaker, but the out-of-sample R^2 is still 9.09% for $\ln C_t$, $k = 22$ (the right half of Table 6)[27].

6 Concluding remarks

In this paper, we investigated the volatility of the daily Nikkei 225 futures realized volatility. Although much of the recent advances in volatility research has been due to the recognition that high-frequency intraday data make daily volatility essentially observable in the form of the realized volatility and related measures, we are back to the condition of unobservability when we move one order higher in terms of the moments from volatility to volatility of volatility. This makes evaluation of models such as the ARFIMA-GARCH for forecasting the volatility of the RV a difficult task. Nevertheless, the GARCH specification for the conditional variance of the RV added to the ARFIMA specification for the conditional mean of the RV seems to capture much of the persistent time variation in the conditional variance of the Nikkei 225 RV regardless of whether we use the raw, square-root- or log-transformed series. We may explore augmenting the ARFIMA model with higher-order GARCH or other GARCH-type or stochastic volatility specifications that have proven successful in improving the GARCH(1,1) fit in more traditional volatility (rather than the volatility-of-the-RV) prediction contexts. Although we focused on the trading-hour continuous sample path variation in this paper, modeling of the dynamics of the day-time jump component series and the overnight and lunch-break returns of the Nikkei 225

[27]For the unrestricted ARFIMA(2,d,2)-GARCH(1,1) model for $\ln C_t$ rather than the BIC-selected ARFIMA(0,d,1)-GARCH(1,1), the out-of-sample R^2 is as high as 17.84% (the details are not shown on Table 6).

futures, for example, along the lines of the approach developed by Andersen et al. (2007b), is also important for completing the prediction of the daily RV. We leave such attempts for future research.

References

[1] Andersen, T.G., Bollerslev, T., 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* 39, 885-905.

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

[3] Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001a, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43-75.

[4] Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001b, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42-55.

[5] Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003, Modeling and forecasting realized volatility, *Econometrica* 71, 579-625.

[6] Andersen, T.G., Bollerslev, T., Diebold, F.X., 2007a, Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility, *Review of Economics and Statistics* 89, 701-720.

[7] Andersen, T.G., Bollerslev, T., Huang, X., 2007b, A reduced form framework for modeling volatility of speculative prices based on realized variation measures, Working paper, Northwestern University and Duke University.

[8] Andersen, T.G., Bollerslev, T., Lange S., 1999, Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon, *Journal of Empirical Finance* 6, 457-477.

[9] Baillie, R.T., 1996, Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5-59.

[10] Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., 1996a, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 74, 3-30.

[11] Baillie, R.T., Chung, C., Tieslau, M.A., 1996b, Analysing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics* 11, 23-40.

[12] Barndorff-Nielsen, O.E., Graversen, S.E., Jacod J., Podolskij, M., Shephard, N., 2006, A central limit theorem for realised power and bipower variation of continuous semimartingales, in Kabanov, Y., Lipseter, R., Stoyanov, J., eds., *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*, Springer Verlag: Berlin.

[13] Barndorff-Nielsen, O.E., Shephard, N., 2002, Econometric analysis of realised volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B* 64, 253-280.

[14] Barndorff-Nielsen, O.E., Shephard, N., 2004, Power and bipower variation with stochastic volatility and jumps, *Journal of Financial Econometrics* 2, 1-37.

[15] Barndorff-Nielsen, O.E., Shephard, N., 2006, Econometrics of testing for jumps in financial economics using bipower variation, *Journal of Financial Econometrics*, 4, 217-252.

[16] Beran, J., 1994, Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *Journal of the Royal Statistical Society B* 57, 659–672.

[17] Blair, B.J., Poon, S., Taylor, S.J., 2001, Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency returns, *Journal of Econometrics* 105, 5-26.

[18] Bollerslev, T., Engle, R.F., Nelson, D.B., 1994, ARCH models, in McFadden, D., Engle, R.F., eds., *Handbook of Econometrics Vol.4*, Elsevier: Amsterdam.

[19] Bollerslev, T., Kretschmer, U., Pigorsch, C., Tauchen, G., 2005, A discrete-time model for daily S&P 500 returns and realized variations: Jumps and leverage effects, forthcoming in *Journal of Econometrics*.

[20] Bollerslev, T., Mikkelsen, H.O., 1996, Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* 73, 151-184.

[21] Bollerslev, T., Wooldridge, J., 1992, Quasi-maximum likelihood estimation and inference for dynamic models with time-varying covariances, *Econometric Reviews* 11, 143-172.

[22] Bondon, P., Palma, W., 2007, A class of antipersistent processes, *Journal of Time Series Analysis* 28, 261-273.

[23] Brock, W.A., Dechert, W.D., Scheinkman, J.A., LeBaron, B., 1987, A test for independence based on the correlation dimension, *Econometric Reviews* 15, 197-235.

[24] Caporale, G.M., Ntantamis, C., Pantelidis, T., Pittis, N., 2005, The BDS test as a test for the adequacy of a GARCH(1,1) specification: A Monte Carlo study, *Journal of Financial Econometrics* 3, 282-309.

[25] Chen, Y., Chou, R.Y., Kuan, C., 2000, Testing time reversibility without moment restrictions, *Journal of Econometrics* 95, 199-218.

[26] Chen, Y., Kuan, C., 2002, Time irreversibility and EGARCH effects in US stock index returns, *Journal of Applied Econometrics* 17, 565-578.

[27] Corsi, F., 1994, A simple long memory model of realized volatility, Unpublished manuscript, University of Southern Switzerland.

[28] Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008, The volatility of realized volatility, *Econometric Reviews* 27, 46-78.

[29] Breidt, F.J., Crato, N., de Lima, P., 1998, On detection and estimation in stochastic volatility, *Journal of Econometrics* 83, 325-348.

[30] De Lima, P.J.F., 1996, Nuisance parameter free properties of correlation integral based statistics, *Econometric Reviews* 15, 237-259.

[31] Deo, R.S., Hurvich, C.M., 2001, On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models, *Econometric Theory* 17, 686-710.

[32] Deo, R.S., Hurvich, C.M., 2003,

[33] Diebold, F.X., 1988, *Empirical modeling of exchange rate dynamics*, Springer.

[34] Diebold, F.X., Gunther, T., Tay, A., 1998, Evaluating density forecasts with applications to financial risk management, *International Economic Review* 39, 863-883.

[35] Diebold, F.X., Nason, J.A., 1990, Nonparametric exchange rate prediction? *Journal of International Economics* 28, 315-332.

[36] Ding, Z., Granger, C.W.J., Engle, R.F., 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83-106.

[37] Ebens, H., 1999, Realized stock volatility, Working paper, John Hopkins University.

[38] Egorov, A., Hong, Y., Li, H., 2006, Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk?, *Journal of Econometrics* 135, 255-284.

[39] Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* 50, 987-1008.

[40] Giot, P., Laurent, S., Modelling daily value-at-risk using realized volatility and ARCH type models, *Journal of Empirical Finance* 11, 379-398.

[41] Geweke, J., Porter-Hudak, S., 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221-238.

[42] Granger, C.W.L., Joyeux, R., 1980, An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-29.

[43] Hansen, P.R., Lunde, A., 2005, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* 24, 127-161.

[44] Hansen, P.R., Lunde, A., 2006, Consistent ranking of volatility models, *Journal of Econometrics* 131, 97-121.

[45] Hosking, J.R., 1981, Fractional differencing, *Biometrika* 68, 165-176.

[46] Hong, Y., Li, H., 2005, Nonparametric specification testing for continuous-time models with applications to term structure of interest rates, *Review of Financial Studies* 18, 37-84.

[47] Huang, X., Tauchen, G., 2005, The relative contribution of jumps to total price variation, *Journal of Financial Econometrics* 3, 456-499.

[48] Jensen, S.T., Rahbek, A., 2004, Asymptotic inference for nonstationary GARCH, *Econometric Theory*, 20, 1203-1126.

[49] Koopman, S.J., Jungbacker, B., Hol, E., 2005, Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements, *Journal of Empirical Finance* 12, 445-475.

[50] Koopman, S.J., Ooms, M., Carnero, M.A., 2007, Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices, *Journal of the American Statistical Association* 102, 16-27.

[51] Kulperger, R., Yu, H., 2005, High moment partial sum processes of residuals in GARCH models and their applications, *Annals of Statistics* 33, 2395-2422.

[52] Lee, S.-W. Hansen, B.E., 1994, Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* 10, 29-52.

[53] Li, Mak, 1994, On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity, *Journal of Time Series Analysis* 15, 627-636.

[54] Ling, S., Li, W.K., 1997, On fractionally integrated autoregressive moving-average time series models with conditional heteroskedasticity, *Journal of the American Statistical Association* 92, 1184-1194.

[55] Martens, M., 2001, Forecasting daily exchange rate volatility using intraday returns, *Journal of International Money and Finance* 20, 1-23.

[56] McLeod, A.I., Li, W.K., 1983, Diagnostic checking ARMA time series using squared-residual autocorrelations, *Journal of Time Series Analysis* 4, 269-273.

[57] Newey, W., Steigerwald, D., 1997, Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, *Econometrica* 65, 587-599.

[58] Oomen, R.C.A., 2001, Using high frequency stock market index data to calculate, model & forecast realized return variance, Working paper, European University Institute.

[59] Ooms, M., Doornik, J, 1999, Inference and forecasting for fractional autoregressive integrated moving average models, with an application to US and UK inflation, Econometric Institute Report 9947/A, Erasmus University.

[60] Palma, W., 2007, *Long-memory Time Series: Theory and Methods*, Wiley: Hoboken, New Jersey.

[61] Patton, A., 2006, Volatility forecast comparison using imperfect volatility proxies, Working paper, London School of Economics.

[62] Robinson, P.M., 1995a, Gaussian semiparametric estimation of long range dependence, *Annals of Statistics*, 23, 1630-1661.

[63] Robinson, P.M., 1995b, Log-periodogram regression of time series with long range dependence, *Annals of Statistics*, 23, 1048-1072.

[64] Robinson, P.M., Henry, M., 1999, Long and short memory conditional heteroskedasticity in estimating the memory parameter of levels, *Econometric Theory*, 15, 299-336.

[65] Shephard, N., 2005, *Stochastic volatility*: *Selected Readings*, Oxford University Press: Oxford.

[66] Shibata, M., 2004, Realized volatility wo mochiita GARCH model no yosokuryoku hikaku, Chapter 2, Ph.D. Dissertation, Tokyo Metropolitan Universality (in Japanese).

[67] Shibata, M., 2008, Kohindo data ni yoru volatility no suitei: Realized volatility no survey to nihon no kabuka shisu oyobi shisu sakimono no jissho bunseki, *Kin-yu Kenkyu* 27 (1)**,** 1-54 (in Japanese).

[68] Shibata, M., Watanabe, T., 2004, Realized volatility wo mochiita Nikkei 225 sakimono kakaku no volatility yoshoku, Osaka Securities Exchange Sakimono Option Report 16-12 (in Japanese).

[69] Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics* 53, 165-188.

[70] Thomakos, D., Tang, T., 2003, Realized volatility in the futures markets, *Journal of Empirical Finance* 10, 321-353.

[71] Velasco, C., 1999, Non-stationary log-periodogram regression, *Journal of Econometrics* 91, 325-371.

[72] Viano, M.C., Deniau, Cl., Oppenheim, C., Long-range dependence and mixing for discrete time fractional processes, *Journal of Time Series Analysis* 16, 323-338.

[73] Watanabe, T., 2005, Nikkei 225 sakimono kakaku no realized volatility, Osaka Securities Exchange Sakimono Option Report 17-5 (in Japanese).

[74] Watanabe, T., Sasaki, K., 2006, ARCH-gata model to "realized volatility" ni yoru volatility yosoku to value-at-risk, *Kin-yu Kenkyu*, 25 (2)**,** 39-79 (in Japanese).

[75] Watanabe, T., Yamaguchi, K., 2006, Nikkei 225 no "realized volatility" to implied volatility, Osaka Securities Exchange Sakimono Option Report 18-12 (in Japanese).

Table 1: Summary statistics: Nikkei 225 daily returns, realized variance and related measures

Series	*Mean*	*Std. Dev.*	*Skew.*	*Kurt.*	*Min.*	*Max.*	*Jarque-Bera*	ρ_1	S.E.(ρ_1)	*LB* (5)	*DLB* (5)	*LB* (10)	*DLB* (10)	*LB* (22)	*DLB* (22)	Long-memory d *GPH*	Long-memory d *Robinson*
$\boldsymbol{R}$	-0.007	1.423	-0.064	4.962	-7.234	7.660	451.13	-0.041	(0.022)	**19.207**	13.251	**23.805**	16.598	**37.588**	26.732	-0.018	-0.011
$\boldsymbol{R_n}$	0.039	0.631	0.009	3.480	-3.186	3.849	26.89	0.036	(0.020)	**12.112**	10.265	**14.545**	12.334	**33.345**	29.016	0.054	0.044
$\boldsymbol{R_{am}}$	-0.025	0.802	-0.242	7.700	-7.394	5.078	2,606.19	-0.033	(0.022)	**9.678**	6.342	**11.876**	7.938	**23.167**	15.856	-0.048	-0.031
$\boldsymbol{R_l}$	-0.026	0.164	0.015	4.905	-0.736	0.786	423.62	0.026	(0.024)	**16.083**	10.712	**22.007**	15.017	**57.635**	40.153	0.071	0.113
$\boldsymbol{R_{pm}}$	0.005	0.774	0.103	5.818	-3.675	4.881	932.38	-0.099	(0.025)	**53.107**	33.121	**60.194**	37.744	**80.960**	53.633	-0.107	-0.060
$\boldsymbol{RV_{am}}$	0.573	0.622	5.970	78.350	0.011	12.748	679,502.99	0.429	(0.060)	**2,253.238**	223.283	**3,635.044**	442.419	**5,801.812**	809.331	0.460	0.454
$\boldsymbol{RV_{pm}}$	0.514	0.595	12.096	335.024	0.015	18.941	12,938,877.08	0.355	(0.059)	**1,298.273**	168.714	**1,981.101**	390.580	**2,873.448**	822.065	0.364	0.376
$\boldsymbol{RV} = RV_{am} + RV_{pm}$	1.087	1.030	5.336	67.840	0.065	20.493	504,141.92	0.540	(0.071)	**3,237.586**	250.406	**4,970.638**	529.470	**7,586.328**	1,004.412	0.470	0.468
$RV^{1/2}$	0.968	0.388	1.459	8.520	0.255	4.527	4,551.03	0.663	(0.097)	**4,997.214**	211.504	**8,342.521**	402.196	**13,887.091**	793.872	0.557	0.524
$\ln RV$	-0.214	0.778	-0.130	3.234	-2.733	3.020	14.25	0.713	(0.030)	**5,927.105**	2,427.793	**10,322.146**	4,474.032	**18,365.815**	8,625.484	0.584	0.533
$\boldsymbol{R_n^2}$	0.399	0.628	8.937	168.634	0.000	14.812	3,240,305.54	0.062	(0.017)	**88.876**	81.982	**159.721**	133.651	**241.634**	221.775	0.197	0.202
$\boldsymbol{R_l^2}$	0.028	0.054	4.178	26.776	0.000	0.618	74,150.20	0.145	(0.033)	**241.331**	93.607	**396.055**	161.320	**793.223**	308.027	0.278	0.255
$\boldsymbol{RV+R_n^2+R_l^2}$	1.514	1.321	4.040	37.141	0.082	21.151	143,704.07	0.461	(0.059)	**2,475.549**	257.636	**3,928.348**	521.401	**6,076.282**	980.450	0.451	0.458
$(RV+R_n^2+R_l^2)^{1/2}$	1.149	0.440	1.283	6.990	0.286	4.599	2,626.47	0.563	(0.090)	**3701.681**	179.179	**6248.175**	337.833	**10273.433**	651.265	0.542	0.502
$\ln (RV+R_n^2+R_l^2)$	0.139	0.752	-0.174	3.327	-2.506	3.052	26.58	0.603	(0.026)	**4291.539**	2415.713	**7482.022**	4,479.544	**12918.003**	8,479.105	0.571	0.505
$\boldsymbol{I_{am}}$	0.054	0.227	3.936	16.492	0.000	1.000	28,485.68	0.082	(0.030)	**40.003**	18.276	**66.501**	30.686	**171.483**	77.806	0.102	0.148
$\boldsymbol{I_{pm}}$	0.105	0.306	2.585	7.680	0.000	1.000	5,676.54	0.063	(0.024)	**66.480**	37.339	**113.491**	65.899	**191.848**	114.649	0.137	0.189
$I = \max (I_{am}, I_{pm})$	0.150	0.357	1.965	4.863	0.000	1.000	2,209.40	0.086	(0.024)	**122.378**	67.857	**201.622**	117.207	**382.053**	229.039	0.198	0.215
$\boldsymbol{J_{am}}$	0.018	0.111	12.378	254.404	0.000	3.126	7,450,609.97	0.024	(0.022)	**23.634**	4.842	**24.062**	5.409	**35.036**	45.902	0.115	0.094
$\boldsymbol{J_{pm}}$	0.027	0.108	6.936	72.084	0.000	1.819	579,666.63	-0.009	(0.008)	**22.952**	11.622	**34.800**	17.749	**78.481**	37.759	0.108	0.137
$\boldsymbol{J = J_{am} + J_{pm}}$	0.045	0.154	6.750	82.923	0.000	3.126	767,042.11	0.099	(0.087)	**82.202**	15.450	**95.866**	22.705	**150.284**	53.468	0.152	0.156
$J^{1/2}$	0.076	0.198	2.834	11.464	0.000	1.768	12,114.74	0.063	(0.030)	**80.413**	36.245	**114.470**	59.428	**226.672**	124.609	0.177	0.184
$\ln (1+J)$	0.037	0.112	4.232	26.939	0.000	1.417	75,267.00	0.067	(0.046)	**71.425**	21.879	**90.850**	33.287	**172.347**	75.446	0.165	0.168
$\boldsymbol{C_{am}}$	0.554	0.617	6.094	81.196	0.011	12.748	731,228.51	0.429	(0.060)	**2,210.926**	215.305	**3,526.428**	431.947	**5,639.629**	789.466	0.426	0.446
$\boldsymbol{C_{pm}}$	0.488	0.592	12.327	343.571	0.012	18.941	13,612,657.54	0.349	(0.059)	**1,247.790**	188.381	**1,891.862**	415.115	**2,714.858**	825.181	0.368	0.373
$\boldsymbol{C = C_{am} + C_{pm}}$	1.042	1.021	5.426	70.160	0.056	20.493	540,350.80	0.537	(0.070)	**3,177.375**	250.504	**4,859.086**	527.560	**7,371.936**	995.301	0.477	0.464
$C^{1/2}$	0.943	0.391	1.449	8.415	0.237	4.527	4,403.37	0.658	(0.094)	**4,880.797**	218.977	**8,129.868**	416.252	**13,491.335**	821.386	0.549	0.518
ln C	-0.278	0.811	-0.161	3.183	-2.879	3.020	16.02	0.703	(0.031)	**5,698.077**	2,265.783	**9,909.682**	4,165.279	**17,598.452**	7,995.639	0.577	0.523

The sample period is from March 11, 1996, through August 31, 2007 (2802 observations). ρ_1 and S.E.(ρ_1) in parentheses are respectively the first-order sample autocorrelations and heteroskedasticity-robust standard errors (the usual Bartlett's standard errors are $T^{-1/2} = 0.018$). *LB* (k) are the Ljeung-Box statistics of orders up to k and *DLB* (k) are their heteroskedasticity adjusted versions. The 5% critical values for $\chi^2(k)$ are 5.991 (k=2), 11.070 (5), 18.307 (10), 33.924 (22). The standard errors of the GPH and Robinson estimators for the long-memory parameter d are 0.034 and 0.031 respectively.

Table 2: ARFIMA-GARCH estimation results

Var.	Model	ARFIMA parameters						GARCH paramters			LL
		μ	d	φ_1	φ_2	ψ_1	ψ_2	σ^2, ω	β	α	BIC
	Best short-memory	**0.7155**		**0.3404**	**0.6116**		**-0.4749**	**0.0134**	**0.7857**	**0.2595**	**-2,365.42**
	memory	(0.0932)		(0.0245)	(0.0267)		(0.0296)	(0.0023)	(0.0223)	(0.0320)	**4,786.40**
	(2,0,{2})-G	(0.1343)		(0.0361)	(0.0425)		(0.0396)	(0.0110)	(0.0932)	(0.1325)	
	Best homoskedastic	**0.7893**	**0.4941**			**-0.2216**		**0.6329**			**-3,335.09**
	(0,d,1)	(0.3326)	(0.0322)			(0.0394)		(0.0169)			**6,701.93**
		(0.2169)	(0.0869)			(0.1086)		(0.1367)			
	GARCH ver. of	**0.4635**	**0.5661**			**-0.2375**		**0.0120**	**0.7993**	**0.2423**	**-2,366.79**
C	best homosked.	(0.2665)	(0.0500)			(0.0626)		(0.0018)	(0.0184)	(0.0274)	**4,781.21**
	(0,d,1)-G	(0.2163)	(0.1022)			(0.1301)		(0.0084)	(0.0730)	(0.1083)	
	Homoskedastic ver.	**0.7704**	**0.5163**		**-0.0321**	**-0.2371**		**0.6324**			**-3,333.97**
	of best overall model	(0.7140)	(0.0342)		(0.0211)	(0.0383)		(0.0168)			**6,707.64**
	({2},d,1)	(0.8760)	(0.0978)		(0.0504)	(0.1010)		(0.1344)			
	Best overall	**0.5185**	**0.6251**		**-0.0966**	**-0.2753**		**0.0120**	**0.7966**	**0.2457**	**-2,360.72**
	({2},d,1)-G	(0.3330)	(0.0467)		(0.0267)	(0.0527)		(0.0019)	(0.0191)	(0.0284)	**4,777.00**
		(0.2393)	(0.0843)		(0.0450)	(0.0879)		(0.0084)	(0.0732)	(0.1099)	
	Best short-memory	**0.6985**		**1.8988**	**-0.8990**	**-1.5908**	**0.5965**	**0.0029**	**0.8630**	**0.0982**	**-47.29**
	memory	(0.0872)		(0.0175)	(0.0174)	(0.0298)	(0.0287)	(0.0007)	(0.0198)	(0.0138)	**158.08**
	(2,0,2)-G	(0.0661)		(0.0188)	(0.0188)	(0.0354)	(0.0342)	(0.0015)	(0.0415)	(0.0271)	
	Best homoskedastic	**0.8151**	**0.5160**			**-0.1847**		**0.0723**			**-295.10**
	(0,d,1)	(0.1152)	(0.0305)			(0.0393)		(0.0019)			**621.94**
		(0.0811)	(0.0515)			(0.0681)		(0.0050)			
	GARCH ver. of	**0.7145**	**0.4990**			**-0.1633**		**0.0027**	**0.8714**	**0.0926**	**-50.53**
$C^{1/2}$	best homosked.	(0.1177)	(0.0297)			(0.0373)		(0.0006)	(0.0179)	(0.0126)	**148.68**
	(0,d,1)-G	(0.1107)	(0.0306)			(0.0397)		(0.0013)	(0.0370)	(0.0248)	
	Homoskedastic ver.	**0.8151**	**0.5160**			**-0.1847**		**0.0723**			**-295.10**
	of best overall model	(0.1152)	(0.0305)			(0.0393)		(0.0019)			**621.94**
	(0,d,1)	(0.0811)	(0.0515)			(0.0681)		(0.0050)			
	Best overall	**0.7145**	**0.4990**			**-0.1633**		**0.0027**	**0.8714**	**0.0926**	**-50.53**
	(0,d,1)-G	(0.1177)	(0.0297)			(0.0373)		(0.0006)	(0.0179)	(0.0126)	**148.68**
		(0.1107)	(0.0306)			(0.0397)		(0.0013)	(0.0370)	(0.0248)	
	Best short-memory	**-0.2637**		**1.1508**	**-0.1685**	**-0.7792**		**0.0002**	**0.9894**	**0.0102**	**-2,100.31**
	memory	(0.1042)		(0.0322)	(0.0298)	(0.0240)		(0.0002)	(0.0032)	(0.0028)	**4,256.19**
	(2,0,1)-G	(0.0945)		(0.0368)	(0.0337)	(0.0289)		(0.0003)	(0.0040)	(0.0036)	
	Best homoskedastic	**-0.5282**	**0.5089**			**-0.1641**		**0.2731**			**-2,157.44**
	(0,d,1)	(0.3443)	(0.0283)			(0.0374)		(0.0073)			**4,346.63**
		(0.4589)	(0.0330)			(0.0465)		(0.0088)			
	GARCH ver. of	**-0.5293**	**0.5030**			**-0.1452**		**0.0002**	**0.9889**	**0.0105**	**-2,095.63**
ln$C$	best homosked.	(0.2520)	(0.0273)			(0.0362)		(0.0002)	(0.0034)	(0.0030)	**4,238.89**
	(0,d,1)-G	(0.1958)	(0.0286)			(0.0398)		(0.0003)	(0.0046)	(0.0039)	
	Homoskedastic ver.	**-0.5282**	**0.5089**			**-0.1641**		**0.2731**			**-2,157.44**
	of best overall model	(0.3443)	(0.0283)			(0.0374)		(0.0073)			**4,346.63**
	(0,d,1)	(0.4589)	(0.0330)			(0.0465)		(0.0088)			
	Best overall	**-0.5293**	**0.5030**			**-0.1452**		**0.0002**	**0.9889**	**0.0105**	**-2,095.63**
	(0,d,1)-G	(0.2520)	(0.0273)			(0.0362)		(0.0002)	(0.0034)	(0.0030)	**4,238.89**
		(0.1958)	(0.0286)			(0.0398)		(0.0003)	(0.0046)	(0.0039)	

The sample period is from March 11, 1996, through August 31, 2007 (2802 observations). For each of the three series, the parameter estimates for the best short-memory model (top), the best homoskedastic model and its GARCH version with the same ARFIMA order, the homoskedastic version of the best model, and the best model (bottom), selected by the BIC from the 64 restricted versions of the ARFIMA(2,d,2)-GARCH(1,1) model, are shown with the Hessian-based (on the first line beneath the parameter estimates) and the Bollerslev-Wooldridge (on the second line) standard errors in parentheses, the log likelihood (LL) and BIC values. The ARFIMA(2,d,1)-GARCH(1,1) model with the first-order AR coefficient restricted to be zero, for example, is denoted as ({2},d,1)-G. For $C^{1/2}$ and ln$C$, the best homooskedastic model and the best model have a common ARFIMA order, resulting in duplications on Table 2.

Table 3: Residual Diagnostic Statistics

Var.	Model	Residuals ε_t									Squared residuals ε_t^2				Standardized residuals Z_t					Squared standardizd residuals Z_t^2			
		mean	*std.*	*skew.*	*kurt.*	*JB*	ρ_1	*LB* (5)	*LB* (10)	*LB* (22)	ρ_1	*LB* (5)	*LB* (10)	*LB* (22)	*mean*	*std.*	*skew.*	*kurt.*	*JB*	ρ_1	*LB* (5)	*LB* (10)	*LB* (22)
C	Short-mem.	0.030	0.801	6.472	122.420	1,684,532.807	**-0.061**	**32.707**	**61.594**	**109.406**	0.112	83.031	85.197	98.084	0.071	0.997	3.295	25.344	63,358.309	-0.002	31.719	33.070	36.102
	(2,0,{2})-*G*						(0.072)	5.340	17.063	30.983													
	Homosked.	0.003	0.796	7.167	131.645	1,956,131.830	**0.003**	**13.625**	**22.272**	**62.813**	0.070	52.801	54.189	64.745	0.004	1.000	7.167	131.645	1,956,131.830				
	(0,*d*,1)						(0.060)	2.237	6.465	19.133													
	+ GARCH	0.007	0.798	6.740	127.021	1,816,978.170	**-0.049**	**23.326**	**34.366**	**77.775**	0.101	70.518	72.028	83.001	0.026	0.999	3.403	27.339	74,569.180	-0.002	32.466	33.766	36.753
	(0,*d*,1)-*G*						(0.070)	3.848	9.357	23.242													
	Homosk..ver.of best	0.003	0.795	7.098	130.267	1,914,529.203	**-0.003**	**9.912**	**20.051**	**61.243**	0.074	55.928	57.387	68.599	0.004	1.000	7.098	130.267	1,914,529.203				
	({2},*d*,1)						(0.061)	1.619	6.422	19.014													
	Best overall	0.004	0.799	6.483	122.088	1,675,364.830	**-0.071**	**24.334**	**40.853**	**88.084**	0.117	86.038	87.770	101.463	0.027	0.999	3.360	26.619	70,401.910	-0.002	32.707	34.055	36.937
	({2},*d*,1)-*G*						(0.073)	3.563	11.070	25.013													
$C^{1/2}$	Short-mem.	0.009	0.269	1.726	14.521	16,888.778	**0.028**	**6.392**	**9.984**	**31.365**	0.135	182.378	217.176	273.101	0.028	1.000	1.310	7.685	3,363.966	0.013	28.953	32.971	37.029
	(2,0,2)-*G*						(0.032)	3.167	5.574	18.568													
	Homosked.	0.001	0.269	1.712	14.590	17,050.511	**0.005**	**7.183**	**11.215**	**28.938**	0.137	179.276	214.392	268.211	0.004	1.000	1.712	14.590	17,050.511				
	(0,*d*,1)						(0.032)	3.859	6.519	17.750													
	Best overall	0.004	0.269	1.730	14.660	17,270.139	**0.001**	**7.538**	**12.025**	**30.246**	0.138	177.706	211.891	265.939	0.009	1.000	1.332	7.801	3,518.869	0.011	34.676	38.197	42.318
	(0,*d*,1)-*G*						(0.032)	3.882	6.803	18.233													
$\ln C$	Short-mem.	-0.001	0.524	0.196	3.897	111.742	**-0.016**	**5.457**	**10.085**	**19.664**	0.099	68.566	133.160	171.932	-0.003	0.987	0.250	3.636	76.444	0.033	24.887	30.485	38.184
	(2,0,1)-*G*						(0.021)	4.502	8.199	17.053													
	Homosked.	0.002	0.523	0.219	3.908	118.640	**0.006**	**5.708**	**8.300**	**16.803**	0.112	78.223	140.330	175.775	0.004	1.000	0.219	3.908	118.640				
	(0,*d*,1)						(0.022)	5.056	7.264	15.227													
	Best overall	0.002	0.523	0.221	3.908	119.116	**-0.007**	**6.063**	**8.751**	**17.454**	0.112	77.339	140.100	176.198	0.000	0.988	0.250	3.666	80.957	0.041	26.919	33.700	41.156
	(0,*d*,1)-*G*						(0.022)	5.367	7.664	15.808													

JB, ρ_1, and *LB* (k) stand respectively for the Jarque-Bera statistic for nonnormality, first-order sample autocorrelation, and the Ljung-Box statistic for no serial correlations of orders up to k (*LB* for the squared residuals ε_t^2 are also called the McLeod-Li statistics). For the residuals ε_t, the heteroskedasticity-consistent standard errors for ρ_1 are given in parentheses (the usual Bartlett's standard error is $T^{-1/2}$ = 0.018) and both the usual *LB* (upper lines) and the heteroskedasticity-adjusted *LB* (lower lines) are shown. The 5% critical values for $\chi^2(k)$ are 5.991 (k=2), 7.815 (3), 9.488 (4), 11.070 (5), 12.592 (6), 14.067 (7), 15.507 (8), 16.919 (9), 18.307 (10), 19.675 (11), 21.026 (12), 22.262 (13), 23.685 (14), 24.996 (15), 26.296 (16), 27.587 (17), 28.869 (18), 30.144 (19), 31.410 (20), 32.671 (21), 33.924 (22).

Table 4: Further specification tests

Var.	Model	*BDS* Line 1: $\ln Z_t^2$ Line 2: Z_t $k=2$	3	4	5	*CCK* $\gamma = 0.5$ $k=1$	2	3	$\gamma = 1.0$ $k=1$	2	3	*Hong-Li* $Q_{HL}(1)$	$W_{HL}(5)$	$W_{HL}(10)$	$W_{HL}(20)$	*Hong-Li separate inference* (m,l) = (1,1)	(1,2) ARCH-*m*	(2,1) Leverage	(2,2)	(3,3)	(4,4)
C	Short-mem.	**1.296**	**1.085**	**0.881**	**-1.896**	4.547	5.571	1.925	2.155	3.095	3.478	135.709	296.375	412.433	602.941	-0.188	0.690	0.117	3.569	3.368	0.896
	(2,0,{2})-*G*	1.511	2.004	3.174	4.439																
	Homosked.	**12.201**	**13.217**	**14.048**	**-2.592**	5.179	4.923	4.328	4.281	4.953	4.424	342.874	740.090	1,011.522	1,441.198	4.639	13.205	22.458	8.789	-2.623	-3.358
	(0,*d*,1)	18.880	22.370	25.656	29.452																
	+ GARCH	**1.947**	**2.557**	**3.709**	**5.024**	4.290	5.353	3.055	3.098	2.163	2.664	137.079	297.039	414.158	605.267	-0.619	0.854	0.138	3.689	3.008	0.419
	(0,*d*,1)-*G*	0.913	1.159	1.304	-1.792																
	Homosk.. ver. of best	**12.651**	**13.742**	**14.481**	**-2.846**	5.242	5.071	3.835	3.833	5.168	4.649	341.469	735.879	1,006.099	1,433.161	4.110	14.516	22.997	9.529	-2.598	-3.352
	({2},*d*,1)	18.891	22.310	25.595	29.404																
	Best overall	**1.079**	**1.352**	**1.333**	**-2.360**	4.895	5.854	1.373	1.621	2.758	3.228	134.695	288.186	402.269	589.806	-1.831	0.251	0.233	3.754	3.202	0.737
	({2},*d*,1)-*G*	1.861	2.245	3.274	4.589																
$C^{1/2}$	Short-mem.	**0.550**	**0.259**	**0.026**	**-1.182**	1.694	2.362	0.894	0.500	1.425	1.725	30.906	68.276	95.601	139.663	-0.358	-0.240	1.551	3.768	7.181	8.302
	(2,0,2)-*G*	0.818	0.912	1.846	2.922																
	Homosked.	**2.712**	**2.868**	**2.915**	**-1.253**	2.498	2.800	1.340	1.038	2.215	2.206	52.669	116.046	158.849	226.107	0.624	1.522	23.293	45.588	5.783	0.825
	(0,*d*,1)	9.168	11.606	14.520	17.888																
	Best overall	**0.351**	**0.288**	**0.094**	**-0.889**	2.097	2.663	0.828	0.334	1.102	1.423	33.284	73.196	102.143	149.058	-1.010	0.334	1.985	4.840	10.672	13.526
	(0,*d*,1)-*G*	0.967	1.171	2.278	3.456																
$\ln C$	Short-mem.	**-0.249**	**-0.338**	**-0.515**	**-0.154**	-1.229	-0.422	-2.242	-2.538	-0.341	-0.084	0.489	2.636	3.312	4.350	-0.494	2.941	-0.128	3.731	8.468	27.987
	(2,0,1)-*G*	0.802	0.874	1.373	2.197																
	Homosked.	**0.248**	**0.342**	**0.382**	**-0.153**	-1.531	-0.737	-2.347	-2.481	-0.202	-0.094	3.677	8.651	11.500	15.658	-0.755	5.292	1.415	28.678	6.514	21.892
	(0,*d*,1)	3.291	4.015	5.097	6.344																
	Best overall	**0.376**	**0.391**	**0.451**	**-0.184**	-1.536	-0.754	-2.099	-2.407	-0.430	-0.215	1.725	3.017	3.681	5.178	-0.846	2.767	-0.144	4.459	10.056	28.736
	(0,*d*,1)-*G*	1.155	1.192	1.675	2.476																

Each test statistic is asymptotically distributed as $N(0,1)$ under its respective null. The null is $Z_t \sim i.i.d.$ for the BDS and CCK statistics and $Z_t \sim i.i.d. N(0,1)$ for the Hong-Li statistics and Hong-Li separate inference sta

Table 5: RV forecast performance evaluation

Variable / Selected Model	Horizon : Day (s)	In-sample RMSE	In-sample Mincer-Zarnowits Regression Int.	Coef.	R^2	*Wald*	Out-of-sample RMSE	Out-of-sample Mincer-Zarnowits Regression Int.	Coef.	R^2	*Wald*
	1	0.7986	0.1063	0.9014	0.3933	7.074	0.4417	-0.0629	1.0662	0.4691	8.382
			(0.0433)	(0.0484)		(0.029)		(0.0235)	(0.0356)		(0.015)
C	5	0.9006	0.2224	0.7930	0.2397	17.408	0.4991	-0.0459	1.0343	0.3217	1.974
({2},d,1)-G			(0.0534)	(0.0565)		(0.000)		(0.0388)	(0.0503)		(0.373)
	10	0.9450	0.2934	0.7273	0.1701	31.865	0.5170	-0.0575	1.0428	0.2751	1.504
			(0.0523)	(0.0534)		(0.000)		(0.0576)	(0.0700)		(0.471)
	22	1.0052	0.4571	0.5727	0.0835	36.492	0.5376	-0.0685	1.0485	0.2222	0.948
			(0.0816)	(0.0715)		(0.000)		(0.0951)	(0.1097)		(0.623)
	1	0.2689	0.0086	0.9951	0.5269	1.042	0.2038	-0.0352	1.0314	0.5770	6.209
			(0.0198)	(0.0230)		(0.594)		(0.0188)	(0.0240)		(0.045)
$C^{1/2}$	5	0.3090	0.0271	0.9800	0.3764	1.997	0.2359	-0.0328	1.0182	0.4349	3.935
(0,d,1)-G			(0.0319)	(0.0364)		(0.369)		(0.0340)	(0.0397)		(0.140)
	10	0.3294	0.0592	0.9490	0.2932	3.004	0.2486	-0.0434	1.0232	0.3770	3.831
			(0.0437)	(0.0489)		(0.223)		(0.0518)	(0.0584)		(0.147)
	22	0.3548	0.1491	0.8573	0.1866	4.487	0.2645	-0.0466	1.0156	0.3034	3.507
			(0.0723)	(0.0750)		(0.106)		(0.0859)	(0.0938)		(0.173)
	1	0.5226	0.0015	0.9979	0.5843	0.063	0.4676	-0.0098	1.0202	0.6425	2.995
			(0.0106)	(0.0163)		(0.969)		(0.0162)	(0.0209)		(0.224)
ln$C$	5	0.5996	0.0070	1.0080	0.4534	0.170	0.5482	-0.0317	1.0203	0.5110	3.701
(0,d,1)-G			(0.0175)	(0.0311)		(0.918)		(0.0247)	(0.0408)		(0.157)
	10	0.6386	0.0072	1.0031	0.3808	0.094	0.5867	-0.0488	1.0207	0.4437	4.106
			(0.0240)	(0.0462)		(0.954)		(0.0324)	(0.0618)		(0.128)
	22	0.6885	-0.0011	0.9656	0.2813	0.224	0.6384	-0.0860	0.9997	0.3505	4.585
			(0.0319)	(0.0750)		(0.894)		(0.0459)	(0.0971)		(0.101)

The target variable is C, $C^{1/2}$, or ln$C$ *(one-day-ahead, or multi-days-ahead single-day).* The Newey-West standard errors (allowing for serial correlations of orders up to 5, 10, 20, and 44 respectively for horizons 1, 5, 10, and 22 days) for the OLS estimates of the regression intercept and the coefficient are given in parentheses. The Wald test statistic with p-value in parentheses is for testing the joint hypothesis of the intercept being zero and the coefficient being one (distributed as $\chi^2(2)$ under the null).

Table 6: Variance-of-RV forecast performance evaluation

Variable / Selected Model	Horizon : Day (s)	In-sample RMSE	In-sample MZ Int.	In-sample MZ Coef.	In-sample MZ R^2	In-sample MZ *Wald*	Out-of-sample RMSE	Out-of-sample MZ Int.	Out-of-sample MZ Coef.	Out-of-sample MZ R^2	Out-of-sample MZ *Wald*
	1	7.3971	0.4192	0.2631	0.0159	156.044	1.0938	0.1520	0.1220	0.0031	184.469
			(0.1091)	(0.0676)		(0.000)		(0.0398)	(0.0712)		(0.000)
C	5	4.3643	0.5073	0.2544	0.0547	277.402	0.7122	0.1653	0.1386	0.0125	194.109
({2},d,1)-G			(0.0976)	(0.0488)		(0.000)		(0.0477)	(0.0823)		(0.000)
	10	4.0076	0.6141	0.2062	0.0517	590.623	0.6767	0.1879	0.1001	0.0095	478.750
			(0.1195)	(0.0329)		(0.000)		(0.0494)	(0.0600)		(0.000)
	22	3.6497	0.7351	0.1543	0.0503	856.786	0.7624	0.1660	0.1120	0.0190	862.298
			(0.1407)	(0.0277)		(0.000)		(0.0577)	(0.0494)		(0.000)
	1	0.2641	0.0197	0.7190	0.0355	4.728	0.1123	0.0195	0.3499	0.0068	72.171
			(0.0093)	(0.1630)		(0.094)		(0.0080)	(0.1211)		(0.000)
$C^{1/2}$	5	0.1706	0.0306	0.7422	0.0760	17.726	0.0724	0.0226	0.4023	0.0168	45.994
(0,d,1)-G			(0.0081)	(0.1272)		(0.000)		(0.0112)	(0.1597)		(0.000)
	10	0.1634	0.0425	0.7081	0.0650	20.623	0.0626	0.0289	0.3575	0.0129	50.369
			(0.0096)	(0.1148)		(0.000)		(0.0121)	(0.1527)		(0.000)
	22	0.1498	0.0561	0.6952	0.0527	18.422	0.0487	0.0095	0.6281	0.0347	30.278
			(0.0149)	(0.1611)		(0.000)		(0.0148)	(0.1841)		(0.000)
	1	0.4585	0.0176	0.9062	0.0331	1.083	0.3512	0.0831	0.5885	0.0048	4.576
			(0.0305)	(0.1204)		(0.582)		(0.0468)	(0.2077)		(0.101)
$\ln C$	5	0.3142	0.0584	0.9345	0.0729	21.464	0.2433	0.0618	0.8841	0.0220	12.812
(0,d,1)-G			(0.0315)	(0.1198)		(0.000)		(0.0493)	(0.2149)		(0.002)
	10	0.2987	0.0761	0.9912	0.0930	43.871	0.2284	0.0368	1.1279	0.0418	32.289
			(0.0369)	(0.1354)		(0.000)		(0.0557)	(0.2372)		(0.000)
	22	0.3006	0.0897	1.1124	0.1250	69.770	0.2226	-0.0170	1.5349	0.0909	54.646
			(0.0468)	(0.1730)		(0.000)		(0.0692)	(0.2983)		(0.000)

The proxy for the target is the squared prediction errors (one-day-ahead or multi-day average) from the selected model for each of the six series. The Newey-West standard errors (allowing for serial correlations of orders up to 5, 10, 20, and 44 respectively for horizons 1, 5, 10, and 22 days) for the OLS estimates of the regression intercept and the coefficient are given in parentheses. The Wald test statistic with p-value in parentheses is for testing the joint hypothesis of the intercept being zero and the coefficient being one (distributed as $\chi^2(2)$ under the null).

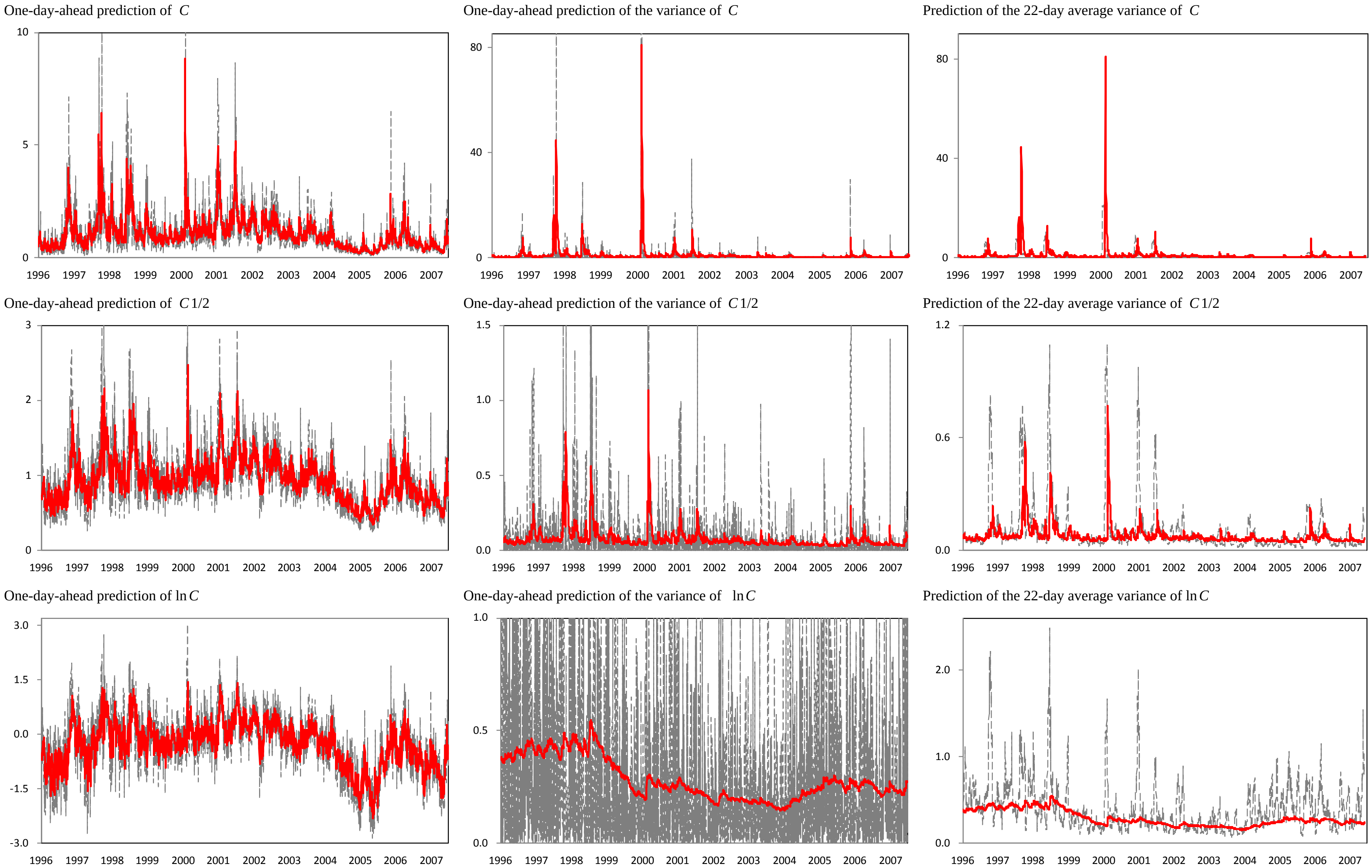

Figure: The left panels plot the daily continuous sample path variation time series and its square-root and log transformations of the Nikkei 225 index (dotted lines) along with the corresponding in-sample one-day-ahead ARFIMA-GARCH forecasts (solid lines). The center panels plot the squared residuals (dotted lines) along with the corresponding in-sample one-day-ahead conditional variance estimate (solid lines) from the ARFIMA-GARCH model. The right panels plot squared residuals (dotted lines) and 22-day-average conditional variance estimates (solid lines) from the ARFIMA-GARCH model.