HITOTSUBASHI UNIVERSITY

Global COE Hi-Stat Discussion Paper Series 033

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

An Optimal Weight for Realized Variance Based on Intermittent High-Frequency Data

Hiroki Masuda
Takayuki Morimoto

February 2009

Hi-Stat Discussion Paper

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

An optimal weight for realized variance based on intermittent high-frequency data*

Hiroki Masuda
Graduate School of Mathematics, Kyushu University,
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Takayuki Morimoto†
Graduate School of Economics, Hitotsubashi University,
2-1 Naka, Kunitachi, Tokyo 186-8601, Japan

Last Revised: November 28, 2008

Abstract

In Japanese stock markets, there are two kinds of breaks, i.e., night-time and lunch break, where we have no trading, entailing inevitable increase of variance in estimating daily volatility via naive realized variance (RV). In order to perform a much more stabilized estimation, we are concerned here with a modification of the weighting technique of Hansen and Lunde (2005). As an empirical study, we estimate optimal weights in a certain sense for Japanese stock data listed on the Tokyo Stock Exchange. We found that, in most stocks appropriate use of the optimally weighted RV can lead to remarkably smaller estimation variance compared with naive RV, hence substantially to more accurate forecasting of daily volatility.

JEL classification: C19, C22, C51
Keywords: high-frequency data, market microstructure noise, realized volatility, Japanese stock markets, variance of realized variance

*We thank Peter Hansen and many others for their helpful comments. We wish to acknowledge the financial support from Grant-in-Aid for Young Scientists (B) by Japan Society for Promotion of Science (Grant No. 19730159).

†Corresponding author: Tel.: +81-42-580-8524; email: morimot@econ.hit-u.ac.jp

1 Introduction

Recently, it has been well recognized that diurnal activity affects the intraday phenomenon, namely, when detailed intraday information is stockpiled, it has a big impact on the market. * The notion of realized variance (RV) has been introduced to deal with this phenomenon, and it has come under intense investigation. For example, see Andersen and Bollerslev (1998a, b), Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen et al. (2001, 2003), Barndorff-Nielsen and Shephard (2002, 2004), as well as references therein. Then, the RV has become one of the critical notions in analyzing market microstructure, as it captures market information more precisely than daily returns, through intraday (high-frequency) data.

Theoretically, RV can be viewed as a proxy variable of Integrated Variance (IV) calculated from intraday full high-frequency log returns, when adopting the semimartingale-model setup having a continuous-martingale part for the underlying log-price process, nowadays widely accepted. Thus we need to employ full high-frequency data for 24 hours in estimation of RV as a measure of daily volatility in actual analysis. We can always observe "full" high-frequency data in case of, e.g., an exchange rate: then we could follow the same line of thought as Andersen et al. (2003) argued in forecasting volatilities in future periods. However, in some stock markets the market activities are restricted, e.g., to 4-5 hours a day in Japanese stock markets. In such a situation, we can only observe *intermittent high-frequency data*, and then variance of computing naive RV over whole day may be much larger compared with the full high-frequency case, due to possible larger fluctuations over longer time-intervals. †

In order to tackle this problem, Hansen and Lunde (2005) have regarded it as a smoothing problem to the period when data is not observed, and estimated an optimal weight to the volatility of each period as a constrained optimization problem. Taking into account only the stock markets in the U.S., they have assumed that markets have only one inactive period within a day, which is, they only consider close-to-open period. We will adopt their approach in order to construct an optimal weight applicable to the Japanese stock markets having two breaks a day, that is, nighttime and

---

*There has been a lot of literature focusing on intraday activity in financial markets investigated by using tick-by-tick data referred to as high-frequency data. For example, please see Dacorogna et al. (2001) for further details.

†There have been some studies which investigates the impact of the overnight return on daily volatility. For example, Gallo (2001) reports the one in New York stock exchange (NYSE) by using GARCH models.

lunch break. As an empirical study, we will estimate optimal weights for Japanese stock data listed on the Tokyo Stock Exchange (First Section) for 3 years, from January 4, 2004 to November 28, 2006. These data are TOPIX (index) and TOPIX core 30 (individual stocks). We found that, in most stocks appropriate use of the optimally weighted RV can lead to remarkably smaller estimation variance compared with naive RV, hence substantially to more accurate forecasting of daily volatility.

The remainder of this article is organized as follows. Section 2 presents the construction of an optimally weighted RV, following the technique of Hansen and Lunde (2005). Section 3 provides some empirical analyses concerning the optimally weighted RV based on the intermittent high-frequency data of the Tokyo Stock Exchange. Section 4 reports the comparison of the forecast performance between optimally weighted and no-weighted RV by using a time series model. Section 5 concludes.

2 An optimal weight for RV under conditional proportionality

Japanese market opens at 9:00 and closes at 15:00 (at each business day) with lunch break from 11:00 to 12:30. Let $T > 0$ represent 24-hours length expediently. Put $I = [0, T] = [(\text{yesterday's closing time}), (\text{today's closing time})]$. Then I can be split into four subperiods:

$$I = \bigcup_{i=1}^{4} I_i,$$

where I_i are regarded as follows:

$$\begin{aligned} I_1 &: \text{ nighttime}, \\ I_2 &: \text{ morning trading hours}, \\ I_3 &: \text{ lunch break}, \\ I_4 &: \text{ afternoon trading hours}. \end{aligned}$$

For convenience, let us put $I_i = [T_{i-1}, T_i]$, so that

$$0 = T_0 < T_1 < T_2 < T_3 < T_4 = T.$$

We can get high-frequency data only over the active periods I_2 and I_4. Based on intermittent high-frequency data over I, we want to estimate the integrated volatility over I, say V. If the underlying log-price process is described by a Brownian semimartingale $X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dw_s$, then the integrated volatility over the period $[u, v]$ is formally defined to be $\int_u^v \sigma_s^2 ds$.

Let V_i stand for the integrated volatility over I_i. Then, in view of the additive character of the integrated volatility, we have $V = \sum_{i=1}^4 V_i$. Denote by $X = (X_t)_{t \in \mathbb{R}}$ the underlying log-price process. A common estimator of V is the *naive RV* given by

$$RV = \sum_{i=1}^{4} \hat{V}_i,$$

where

$$\begin{aligned}
\hat{V}_1 &:= (X_{T_1} - X_{T_0})^2 = (\text{squared return over nighttime}),\\
\hat{V}_2 &:= (\text{RV over } I_2),\\
\hat{V}_3 &:= (X_{T_3} - X_{T_2})^2 = (\text{squared return over lunch break}),\\
\hat{V}_4 &:= (\text{RV over } I_4).
\end{aligned}$$

It may be expected that estimation and prediction of (V_1, V_3) is more unstable compared with that of (V_2, V_4), due to the lack of high-frequency data therein. At the same time, we should not simply preclude fluctuations over each I_1 and I_3 in general, as they may exhibit non-negligible impact for the target variable V.

Instead of the naive RV, we are concerned here with a *weighted RV* of the form

$$RV(\lambda) := \sum_{i=1}^{4} \lambda_i \hat{V}_i$$

for some constant $\lambda = (\lambda_i)_{i \leq 4}$. A natural optimal weight, say $\lambda^* = (\lambda_i^*)_{i \leq 4}$, is then given by the minimizer of the mean square error

$$\lambda \mapsto \text{MSE}(\lambda) := E[|RV(\lambda) - V|^2].$$

In general it is impossible to get an empirical variant of λ^* as V cannot be observed. Following the approach taken in Hansen and Lunde (2005, Section 2), we can provide a closed-form solution to this optimization problem under a kind of conditional proportionality assumption, which entails that $RV(\lambda)$ is V_k-conditionally unbiased.

Write $\mu_0 = E[V]$, $\mu_i = E[\hat{V}_i]$, $\eta_{ij} = cov[\hat{V}_i, \hat{V}_j]$, and $\gamma_{ij} = \eta_{ij}/(\mu_i \mu_j)$ for $1 \leq i, j \leq 4$. Further, put $d_{ij} = \mu_i \mu_j (\gamma_{44} + \gamma_{ij} - \gamma_{i4} - \gamma_{j4})$ and $b_i =$

$\mu_0\mu_i(\gamma_{44} - \gamma_{i4})$ for $1 \leq i, j \leq 3$, and then

$$D_4 = \begin{pmatrix} d_{11} & d_{12} & d_{13} & 0 \\ d_{21} & d_{22} & d_{23} & 0 \\ d_{31} & d_{32} & d_{33} & 0 \\ \mu_1 & \mu_2 & \mu_3 & \mu_4 \end{pmatrix}, \qquad b_4 = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \mu_0 \end{pmatrix}.$$

By means of Lemma A with $m = 4$ and $\mathcal{G} = \{\phi, \Omega\}$, we have

Lemma. *Suppose that $\mu_i > 0$ a.s., and that for each $i \leq 4$ there exists a constant ρ_i such that*

$$E_V[\hat{V}_i] = \rho_i V, \tag{1}$$

Then $\lambda \mapsto \mathrm{MSE}(\lambda)$ defined on

$$\Lambda := \left\{ \lambda = (\lambda_i)_{i=1}^4 \in \mathbb{R}_+^4 \,\middle|\, \sum_{i=1}^4 \lambda_i \mu_i = \mu_0 \right\}$$

is minimized by $\lambda^* = \mathrm{argmin}_{\lambda \in \Lambda} var[\hat{V}(\lambda)]$, which is explicitly given by $\lambda^* = D_4^{-1} b_4$ as soon as D_4 is invertible.

This lemma is a multi-intermittence variant of Hansen and Lunde (2005, Sections 2.2 and 2.3), which corresponds to the case where $m = 2$ and $\mathcal{G} = \{\phi, \Omega\}$ in Lemma A. The assumption (1), which leads to the unbiasedness of $\hat{V}(\lambda)$ for every $\lambda \in \Lambda$, cannot be suppressed in general for computing the λ^* without involving the latent variable V.

Our task toward empirical analysis is to evaluate constants $(\mu_i)_{i=0}^4$ and $[\eta_{ij}]_{i,j=1}^4$, and of course this in principle requires specification of underlying model structure and forms of V_i as well as their relation to V. As in Hansen and Lunde (2005), in the empirical study given in the next section we will simply use the empirical quantities for evaluations of $(\mu_i)_{i=0}^4$ and $[\eta_{ij}]_{i,j=1}^4$.

3 Empirical study

In this section we apply our optimal weight for intermittent high-frequency data to Japanese stock data. We use Japanese stock data listed on the Tokyo Stock Exchange (First Section) for 3 years, from January 4, 2004 to November 28, 2006. These are TOPIX (index) and TOPIX core 30 (individual stocks). However, we deselect four stocks, Seven & I Holdings, Mitsubishi UFJ Financial Group, Sumitomo Mitsui Financial Group, and Mizuho Financial Group. The Seven & I Holdings is done for the reason that it was formed on September 1, 2005, and the other three banking holding companies is done for the reason that we cannot optimize the weights for these data fluctuating irregularly after Japan's financial big bang. As a result, we use one index and 27 individual stocks. In sum, we perform our empirical analysis using 27 data series. These are listed in Table 1 along with the number of observations N.

As mentioned before, the Japanese stock market is divided into two sessions by a lunch break, i.e., the morning session from 9:00 to 11:00 and the afternoon session from 12:30 to 15:00.[1] [‡] Taking into consideration the minimum observation interval of the Japanese stock market, we take 1 minute as a sampling frequency. Thus, the sample size of zenba and goba are 120 and 150, respectively. Now let $(Y_{k,2,i})_{i=1}^{120}$ and $(Y_{k,4,i})_{i=1}^{150}$ denote the kth-day intraday returns over zenba and goba, respectively, and then define the kth-day naive realized variance by

$$\begin{aligned} RV_k &:= Y_{k,1}^2 + RV_{k,2} + Y_{k,3}^2 + RV_{k,4}, \\ &= Y_{k,1}^2 + \sum_{i=1}^{120} Y_{k,2,i}^2 + Y_{k,3}^2 + \sum_{j=1}^{150} Y_{k,4,j}^2. \end{aligned}$$

where $Y_{k,1}^2$, $RV_{k,2}$, $Y_{k,3}^2$, and $RV_{k,4}$ denote the square of close-to-open return, RV in morning session, the square of lunch break return, and RV in afternoon session on kth day, respectively.

As in the case of U.S.-stock market handled in Hansen and Lunde (2005), unrestricted estimates are found to be strongly influenced by the most extreme values. So we filter the raw data for outliers. We classify 1% of the observations $Y_{.,1}$, $Y_{.,2}$, $Y_{.,3}$, and $Y_{.,4}$ as outliers and omitted from the estimation.[2] [§]

The literature says that the data are contaminated with market microstructure noise if sampling frequency is too high, and that it leads to a biased estimate. Then, in order to mitigate the influence of the noise, we use Newey-West type modified realized variance (RV_{NW}) in our analysis following Hansen and Lunde (2005). The RV_{NW} estimators over the k th lunch break and the k th nighttime, say $RV_{NW,k,2}$ and $RV_{NW,k,4}$, respectively, are

[‡]These two sessions are respectively called "zenba" and "goba".

[§]As for JAPAN TOBACCO, we take 0.1% data as outliers.

defined based on the Bartlett kernel:

$$RV_{NW,k,2} := \sum_{i=1}^{120} Y_{k,2,i}^2 + 2\sum_{h=1}^{q}\left(1 - \frac{h}{q+1}\right)\sum_{j=1}^{120-h} Y_{k,2,j}Y_{k,2,j+h},$$
$$RV_{NW,k,4} := \sum_{i=1}^{150} Y_{k,4,i}^2 + 2\sum_{h=1}^{q}\left(1 - \frac{h}{q+1}\right)\sum_{j=1}^{150-h} Y_{k,4,j}Y_{k,4,j+h},$$

where q is the number of autocovariances in our empirical study,[3] we will utilize the $RV_{NW,k,i}$ for $RV_{k,i}$, $i = 2, 4$. [¶] This estimator has the advantage that it is guaranteed to be nonnegative; see Newey and West (1987). We show how the bias occurs in too high-frequency sampling and how the RV_{NW} can correct it by plotting the volatility signature plot introduced by Anderson et al. (2000). See Figure 1. The upper panel is for the TOPIX and the lower for the JAPAN TOBACCO. In these figures, the horizontal axis is the sampling interval ranging from 1 to 20 minutes. The vertical axis is the averaged RV over all sampling periods.

From these figures we can clearly see that RV_{NW}s are relatively stable at every sampling frequency, while RVs estimated in usual way are widely ranged depending on sampling frequency. Furthermore, the plot of the TOPIX has upward bias; conversely, the others including the JAPAN TOBACCO have downward bias.

Hereafter we will omit the subscript $_{NW}$ in $RV_{NW,k,2}$ and $RV_{NW,k,4}$.

3.1 Estimation of optimal weight

Here, we estimate the optimal weight λ^* obtained in Section **??** for the volatilities in each intraday period with real data. The λ^* can be obtained by some optimal measures μ_i and $\eta_{i,j}$ (simply, $\eta_i := \eta_{i,i}$), which are estimated as expected values and variances. Let $\hat{V}_{k,1} = Y_{k,1}^2$, $\hat{V}_{k,2} = RV_{k,2}$, $\hat{V}_{k,3} = Y_{k,3}^2$,

[¶] We take $q = 10$ which spans a 10-minute period.

and $\hat{V}_{k,4} = RV_{k,4}$, then

$$\hat{\mu}_0 = \frac{1}{n}\sum_{t=1}^{n}(\hat{V}_{t,1} + \hat{V}_{t,2} + \hat{V}_{t,3} + \hat{V}_{t,4}),$$

$$\hat{\mu}_i = \frac{1}{n}\sum_{t=1}^{n}\hat{V}_{t,i}, \quad i = 1, 2, 3, 4,$$

$$\hat{\eta}_i = \frac{1}{n}\sum_{t=1}^{n}(\hat{V}_{t,i} - \hat{\mu}_i)^2, \quad i = 1, 2, 3, 4,$$

$$\hat{\eta}_{i,j} = \frac{1}{n}\sum_{t=1}^{n}(\hat{V}_{t,i} - \hat{\mu}_i)(\hat{V}_{t,j} - \hat{\mu}_j), \quad i, j = 1, 2, 3, 4,$$

where n is the number of daily observations over the sample period.

Tables 1-4 show the estimates of these optimal measure and optimal weight for each data. From these tables we have several interesting observations as follows.

- Table 1 shows that each volatility of index or TOPIX is very low compared with the individual stocks. Moreover, the volatilities of $\hat{\mu}_3$, i.e., volatilities in lunch time are remarkably low compared with others.
- Table 2 indicates variance estimates of each volatility. The values of $\hat{\eta}_1$ are quite larger than others through all stocks. This implies the need for obtaining "optimal weight" in empirical analysis.
- Table 3 has correlation estimates between volatilities. This has a noticeable consequence that the estimates between $\hat{\eta}_1$ and $\hat{\eta}_3$, i.e., close-to-open and lunch break in several stocks have negative correlations. As expected, the estimates in all stocks have very high correlation between $\hat{\eta}_2$ and $\hat{\eta}_4$, i.e., morning session and afternoon session volatilities.
- Finally Table 4 gives estimates $\hat{\lambda}^* = (\hat{\lambda}^*_i)_{i\le4}$ of the optimal weight λ^*. These estimates are large in the order of $\hat{\lambda}^*_1$, $\hat{\lambda}^*_3$, $\hat{\lambda}^*_2$, and $\hat{\lambda}^*_4$ on average. However, it is also interesting that $\hat{\lambda}^*_4$s are larger than $\hat{\lambda}^*_2$s in some stocks.[4] ‖

3.2 Result and discussion

In this subsection, we investigate whether variances of RVs are reduced well by using the estimates obtained above. For the purpose, we compare RV

‖When the optimal weight $\hat{\lambda}$ has a negative component, we there set zero conveniently.

calculated by usual way and weighted RV. These two RVs are obtained from

$$RV_k = Y_{k,1}^2 + RV_{k,2} + Y_{k,3}^2 + RV_{k,4},$$
$$RV_k(\hat{\lambda}^*) = \hat{\lambda}_1^* Y_{k,1}^2 + \hat{\lambda}_2^* RV_{k,2} + \hat{\lambda}_3^* Y_{k,3}^2 + \hat{\lambda}_4^* RV_{k,4}.$$

The sample period for estimation of optimal weights is ranged from 2004 to 2006, which means that we perform in-sample estimation. Table 5 shows the result. By definition, there is no change in these averages. However, these variances are significantly reduced in all stocks. Additionally, we plot these RVs in Figure 2. The upper panel is for the TOYOTA and the lower for the Nomura Holdings. In this figure, crosses indicate conventional RVs and open circles indicate weighted RVs. We recognize at a glance that the variances of RVs are reduced over estimation periods. In Figure 3, we plot $\hat{V}_{k,i}$ or $\lambda_i \hat{V}_{k,i}$ in each time period, separately. The upper panel is for the $\hat{V}_{k,i}$ of TOYOTA and the lower for the $\lambda_i \hat{V}_{k,i}$. It can be recognized from this figure that the overnight variance notably gets smaller and the variances in active periods get larger by optimally weighting the data. In view of the stylized fact that there is a positive correlation between volume and volatility (for example, see the extensive survey of Karpoff (1987)), it is quite natural that the optimal weight λ_i in inactive periods such as overnight and lunchtime one is relatively small. After all, we can conclude that the optimal weight may significantly reduce the “variance of RV” for more accurate forecasting of volatility based on intermittent high-frequency data.

Furthermore, we analyze two aditional cases of intermittent high-frequency data. First, we set the number of lambdas to be estimated to 2 by merging lunchtime squared return $Y_{k,3}^2$ into overnigtht one $Y_{k,1}^2$ and morning realized volatility $RV_{k,2}$ into afteroon one $RV_{k,4}$, respectively.

$$RV_{w2,k}(\hat{\lambda}^*) = \hat{\lambda}_1^* (Y_{k,1}^2 + Y_{k,3}^2) + \hat{\lambda}_2^* (RV_{k,2} + RV_{k,4}).$$

This case is essentially identical to Hansen and Lunde (2005). Secondly, we set it to 3 by uniting morning realized volatility $Y_{k,1}^2$ and $Y_{k,2}^3$.

$$RV_{w3,k}(\hat{\lambda}^*) = \hat{\lambda}_1^* (Y_{k,1}^2 + Y_{k,3}^2) + \hat{\lambda}_2^* RV_{k,2} + \hat{\lambda}_3^* RV_{k,4}.$$

Table 6 and 7 show the result.This indicates that the optimal weight for $RV_{k,2}$ or $RV_{k,4}$ is heavier than the one of $(Y_{k,1}^2 + Y_{k,3}^2)$, which is consistent with the $RV_{k,4}$ case.

4 Ccomparison of the forecast performance

Finally, we compare the forecast performance of weighted and non-weighted RVs by using a time series model. Many literatures have reported that the specification of RV with the following ARFIMA (autoregressive fractionally integrated moving average) model provides better accurate forecast performance than any other time series models since realized volatility follows a long-memory process, e.g. Andersen et al. (2003) or Watanabe and Yamaguchi (2007) for Japanese stock market, and so on.

$$\phi(L)(1-L)^d RV_k = \theta(L)u_k, \quad u_k \sim NID(0,\sigma^2),$$

where $NID(0,\sigma^2)$ denotes normally and independently distributed with zero mean and variance σ^2, L denotes the lag operator and $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ are the p-th and q-th order lag polynomials. So we now estimate ARFIMA(p,d,q) model for four RV series obtained above in order to compare the forecast performance. More specifically, we estimate the memory parameter d in the model by using Reisen (1994) estimator[**] and optimal lag orders p and q are chosen by using the minimum SIC criterion.[††] Table 8 and 9 show these estimates.[‡‡]

After estimating parameters of ARFIMA model for each RV, we compare the forecast performance by using two loss functions such as RMSE (root mean squared error), MAE (mean absolute error):

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(RV_t - \hat{\sigma}^2_{t|t-1}\right)^2}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|RV_t - \hat{\sigma}^2_{t|t-1}\right|$$

where N is the number of trading days in the sample period such as from January 4, 2004 to November 28, 2006 and $\hat{\sigma}_{t|t-1}$ denotes the in-sample one-step-ahead volatility forecast regarding the realized volatility as a proxy for the true volatility. Table 10-13 show the values of loss Functions and the ratios of these values of three weighted RVs against ones of no-weighted

[**] It is based on the regression equation using the smoothed peridogram function as an estimate of the spectral density. See Reisen (1994) in detail.

[††] We also use AIC criterion but the selection is almost the same as the SIC's.

[‡‡] If $d = 0$, ARFIMA model collapses to stationary ARMA model and if $d = 1$, it becomes ARIMA model. If $0 < d < 0.5$, RV_k follows a stationary long-memory process and if $0.5 \leq d < 1$, RV_k follows a nonstationary long-memory process.

RV. From these tables, we can see that weighted RVs virtually overcome no-weighted RV in both of RMSE and MAE but there is no noticeable difference among three weighted RVs.

Anyway, these results here imply that modeling RV with optimal weights can significantly improve the forecast performance of daily volatility.

5 Concluding remarks

In this article, in order to perform estimation of the integrated volatility with variance being less than conventional RV, we first formulated an optimal closed-form random weighting procedure under the conditional proportionality of the computable "basis" variable $(V_j)_{j \leq m}$. Then we have obtained the preferable empirical evidence that applying this weighting procedure can reduce the variances of estimating integrated volatility for most stocks. Our empirical analysis substantially implies that, as soon as we are concerned with intermittent high-frequency data, the optimally weighted RV can lead to more accurate forecasting of daily volatility than the common naive RV.

Appendix.

Here we will compute the explicit form of λ^* given in Section 2 within a more formal setup.

Let $(\Omega, \mathcal{F}, P)$ be an underlying probability space. Given any natural number $m \geq 2$ (say $m = m' + m''$, where, in the main context, m' corresponds to the number of inactive periods of tradings, and m'' to that of active periods where we can get reasonably high-frequency data). Let V and $\hat{V}_i$, $i \leq m$, be nonnegative random variables. Fix a sub σ-field $\mathcal{G} \subset \mathcal{F}$ and write $\mathcal{H} = \mathcal{G} \vee \sigma(V)$, so that $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$. Now V is the target (latent) variable to be estimated based on all available information, and we want to find the optimal $\mathcal{G}$-measurable random weight $\lambda^* = (\lambda_i^*)_{i \leq m}$, which a.s. minimizes the $\mathcal{G}$-conditional mean square error given by

$$\lambda = (\lambda_i)_{i \leq m} \mapsto \mathrm{MSE}_{\mathcal{G}}(\lambda) := E_{\mathcal{G}}[|\hat{V}(\lambda) - V|^2],$$

where $E_{\mathcal{G}}$ stands for the $\mathcal{G}$-conditional expectation operator, and the estimator $\hat{V}(\lambda)$ of V is supposed to take the form

$$\hat{V}(\lambda) = \sum_{i=1}^{m} \lambda_i \hat{V}_i. \tag{2}$$

As in Hansen and Lunde (2005), we here focus on $\lambda = (\lambda_j)_{j \leq m} \in \Lambda_{\mathcal{G}}$ with the random index set $\Lambda_{\mathcal{G}}$ being

$$\Lambda_{\mathcal{G}} = \left\{ \lambda = (\lambda_i)_{i=1}^m \in \mathbb{R}_+^m \,\middle|\, \sum_{i=1}^m \lambda_i \mu_i = \mu_0 \right\},$$

where

$$\mu_0 = E_{\mathcal{G}}[V] \quad \text{and} \quad \mu_i = E_{\mathcal{G}}[\hat{V}_i].$$

Here we implicitly suppose $\mu_i > 0$ a.s. Write

$$\eta_{ij} = cov_{\mathcal{G}}[\hat{V}_i, \hat{V}_j] \quad \text{and} \quad \gamma_{ij} = \frac{\eta_{ij}}{\mu_i \mu_j}.$$

With these notation, we are going to derive the explicit form of $\lambda^* \in \Lambda_{\mathcal{G}}$ under an additional assumption of a kind of $\mathcal{H}$-conditional proportionality of $\hat{V}_i$ to V, in a similar manner to Hansen and Lunde (2005, Theorem 5), which corresponds to the case of $m = 2$ and $\mathcal{G} = \{\phi, \Omega\}$. In the sequel we will suppress the term "a.s." for brevity in equations involving random variables and/or conditional expectations.

Suppose that for each $i \leq m$ there exists an $\mathcal{G}$-measurable random variable ρ_i such that

$$E_{\mathcal{H}}[V_i] = \rho_i V.$$

Then, by taking the conditional expectation $E_{\mathcal{G}}$ in (2) we have

$$E_{\mathcal{H}}[\hat{V}(\lambda)] = \sum_{i=1}^{m} \lambda_i \rho_i V, \tag{3}$$

hence taking $E_{\mathcal{G}}$ and using the fact $\mathcal{G} \subset \mathcal{H}$ yield

$$E_{\mathcal{G}}[\hat{V}(\lambda)] = \mu_0 \sum_{i=1}^{m} \lambda_i \rho_i. \tag{4}$$

On the other hand, taking $E_{\mathcal{G}}$ in (2) yields that

$$E_{\mathcal{G}}[\hat{V}(\lambda)] = \sum_{i=1}^{m} \lambda_i \mu_i = \mu_0 \tag{5}$$

for $\lambda \in \Lambda_{\mathcal{G}}$. Equating the right-hand sides of (4) and (5) yields $\sum_{i=1}^{m} \lambda_i \rho_i = 1$ for $\lambda \in \Lambda_{\mathcal{G}}$. Therefore, from (3) we get for $\lambda \in \Lambda_{\mathcal{G}}$

$$E_{\mathcal{G}}[\hat{V}(\lambda)] = V. \tag{6}$$

(hence $E_{\mathcal{G}}[\hat{V}(\lambda)] = \mu_0$) According to (6) and simple conditioning argument we get $E_{\mathcal{G}}[|\hat{V}(\lambda) - V|^2] = var_{\mathcal{G}}[\hat{V}(\lambda)] - 2E_{\mathcal{G}}[\{\hat{V}(\lambda) - V\}(V - \mu_0)] - var_{\mathcal{G}}[V] = var_{\mathcal{G}}[\hat{V}(\lambda)] - var_{\mathcal{G}}[V]$ for $\lambda \in \Lambda_{\mathcal{G}}$, thereby we arrive at

$$\lambda^* = \text{argmin}_{\lambda \in \Lambda} var_{\mathcal{G}}[\hat{V}(\lambda)],$$

which serves as the optimal $\mathcal{G}$-measurable random weight within $\Lambda_{\mathcal{G}}$ for $L^2(P|_{\mathcal{G}})$-projection of V onto the linear space spanned by $\{V_1, V_2, \ldots, V_m\}$, where $P|_{\mathcal{G}}$ denotes the restriction of P to $\mathcal{G}$.

For any $\lambda = (\lambda_i)_{i \leq m} \in \Lambda_{\mathcal{G}}$ we may set

$$\lambda_m = \frac{1}{\mu_m}\left(\mu_0 - \sum_{i=1}^{m-1} \lambda_i \mu_i\right).$$

Then observe that

$$var_{\mathcal{G}}[\hat{V}(\lambda)] = \sum_{i=1}^{m} \lambda_i^2 \eta_{ii}^2 + 2 \sum_{1 \leq i < j \leq m} \lambda_i \lambda_j \eta_{ij}^2 =: \zeta(\lambda_1, \ldots, \lambda_{m-1}).$$

For each $i \in \{1, \ldots, m-1\}$ we have

$$\partial_{\lambda_i} \zeta(\lambda_1, \ldots, \lambda_{m-1}) = 2\left(d_{ii}\lambda_i + \sum_{1 \le j \le m-1, j \ne i} \lambda_j d_{ij} - b_i \right),$$

where

$$\begin{aligned} d_{ij} &= \mu_i \mu_j (\gamma_{mm} + \gamma_{ij} - \gamma_{im} - \gamma_{jm}), \\ b_i &= \mu_0 \mu_i (\gamma_{mm} - \gamma_{im}) \end{aligned}$$

for $1 \le i, j \le m-1$. In view of the first-order condition $\nabla_{(\lambda_1, \ldots, \lambda_{m-1})} \zeta(\lambda_1, \ldots, \lambda_{m-1}) = 0$ and the definition of $\Lambda_{\mathcal{G}}$, we see that for $\lambda \in \Lambda_{\mathcal{G}}$ the optimal $\mathcal{G}$-measurable weight $\lambda^* = (\lambda_i^*)_{i=1}^m$ fulfils $D\lambda^* = b$, where $D \in \mathbb{R}^m \otimes \mathbb{R}^m$ and $b \in \mathbb{R}^m$ are given by

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1,m-1} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ d_{m-1,1} & \cdots & d_{m-1,m-1} & 0 \\ \mu_1 & \cdots & \mu_{m-1} & \mu_m \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_{m-1} \\ \mu_0 \end{pmatrix}.$$

Summarizing the above yields the following assertion.

Lemma A. *Suppose that $\mu_i > 0$ a.s., and that for each $i \le m$ there exists a $\mathcal{G}$-measurable random variable ρ_i such that*

$$E_{\mathcal{H}}[V_i] = \rho_i V, \quad \text{a.s.} \tag{7}$$

Then, the $\mathcal{G}$-measurable function $\lambda \mapsto E_{\mathcal{G}}[|\hat{V}(\lambda) - V|^2]$ defined on $\Lambda_{\mathcal{G}}$ is a.s. minimized by $\lambda^* = \operatorname{argmin}_{\lambda \in \Lambda} var_{\mathcal{G}}[\hat{V}(\lambda)]$, which is in turn explicitly given by a solution of $D\lambda = b$. Therefore $\lambda^* = D^{-1}b$ as soon as D is invertible.

References

- Andersen, T. G. and T. Bollerslev (1998a), "Answering the skeptics: yes, standard volatility models do provide accurate forecasts," *International Economic Review* **39**, 885–905.

- Andersen, T. G. and T. Bollerslev (1998b), "Deutsche Mark-Dollar volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*, **53**, 219–265.

- Andersen, T. G., T. Bollerslev, F. X. Diebold and H. Ebens (2001), "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, **61**, 43–76.

- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000), "Great Realizations," *Risk*, **13**, 105-108.

- Andersen, T. G., T. Bollerslev, F.X. Diebold and P. Labys (2001), "The distribution of realized exchange rate volatility," *J. Amer. Statist. Assoc.* **96**, 42–55.

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

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

- Barndorff-Nielsen, O. E. and N. Shephard (2004), "Econometric analysis of realized covariation: high frequency covariance, regression and correlation in financial economics," *Econometrica*, **72**, 885–925.

- Dacorogna, M., U. Muller, R. Gencay, R. Olsen and O. Pictet (2001), "*An Introduction to High-Frequency Finance*," Academic Press, San Diego, CA.

- Gallo, G.M. (2001), "Modelling the Impact of Overnight Surprises on Intra-daily Volatility," *Australian Economic Papers*, **40**, 567–580.

- Hansen, P. R. and A. Lunde (2005), "A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data," *Journal of Financial Econometrics*, **3**, 525–554.

- Karpoff, J. M. (1987), “The relation between price changes and trading volume: a survey,” *Journal of Financial and Quantitative Analysis*, **22**, 109–126.

- Newey, W., and K. West (1987), “A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica* **55**, 703–708.

- Reisen, V. A. (1994). “Estimation of the fractional difference parameter in the ARFIMA(p,d,q) model using the smoothed periodogram,” Journal Time Series Analysis, 15, pp. 335–350.

- Watanabe, T. and K. Yamaguchi (2007). “Measuring, Modeling and Forecasting Realized Volatility in the Japanese Stock Market,” In Proceedings of the FCS Conference in Honor of the Retirement of Professor Hajime Wago, pp. 6–16.

Asset	N	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
TOPIX	700	0.632	0.229	0.220	0.009	0.174
JAPAN TOBACCO	727	5.208	1.239	2.005	0.135	1.829
Shin-Etsu Chemical	699	2.646	0.796	0.934	0.052	0.865
Takeda Pharm.	699	1.613	0.442	0.584	0.027	0.559
Astellas Pharma Inc.	699	2.872	0.926	1.003	0.053	0.890
FUJIFILM Holdings	699	2.557	0.728	0.893	0.056	0.879
NIPPON STEEL	699	3.613	0.855	1.324	0.062	1.371
JFE Holdings,Inc.	699	3.357	0.992	1.199	0.051	1.115
Hitachi,Ltd.	699	2.351	0.946	0.734	0.032	0.639
Matsushita	699	2.121	0.892	0.630	0.033	0.566
SONY	699	2.855	1.073	0.900	0.036	0.845
NISSAN MOTOR	699	2.048	0.959	0.562	0.025	0.503
TOYOTA	699	2.077	0.663	0.683	0.030	0.700
HONDA MOTOR	699	2.548	0.932	0.803	0.039	0.774
CANON INC.	699	2.168	0.855	0.649	0.031	0.632
Nintendo Co.,Ltd.	699	2.810	1.253	0.889	0.051	0.617
Mitsubishi Corp.	699	2.980	1.101	1.003	0.041	0.835
ORIX	698	4.208	1.714	1.375	0.076	1.042
Nomura Holdings	699	3.090	1.331	0.919	0.043	0.796
Millea Holdings	695	5.842	1.052	2.263	0.143	2.383
Mitsubishi Estate	699	3.896	1.347	1.418	0.055	1.075
East Japan Railway	699	1.574	0.397	0.617	0.035	0.525
NTT	699	2.715	0.876	0.944	0.040	0.855
KDDI	699	2.727	0.858	0.964	0.051	0.854
NTT DoCoMo,Inc.	699	5.669	1.050	2.059	0.148	2.412
Tokyo Electric Power	699	1.312	0.236	0.513	0.029	0.534
SOFTBANK CORP.	699	7.374	1.918	2.902	0.082	2.472

Table 1 Empirical estimates $\hat{\mu}$

Asset	$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$	$\hat{\eta}_4$
TOPIX	0.089	0.032	0.000	0.021
JAPAN TOBACCO	5.947	3.342	0.074	2.483
Shin-Etsu Chemical	1.438	0.317	0.007	0.261
Takeda Pharm.	0.575	0.099	0.002	0.097
Astellas Pharma Inc.	2.135	0.337	0.007	0.222
FUJIFILM Holdings	1.216	0.200	0.006	0.163
NIPPON STEEL	1.690	0.465	0.009	0.548
JFE Holdings,Inc.	2.616	0.554	0.007	0.624
Hitachi,Ltd.	2.200	0.285	0.002	0.211
Matsushita	2.419	0.236	0.004	0.187
SONY	2.666	0.227	0.003	0.162
NISSAN MOTOR	2.113	0.158	0.002	0.138
TOYOTA	0.889	0.118	0.002	0.113
HONDA MOTOR	2.065	0.199	0.004	0.177
CANON INC.	1.738	0.130	0.002	0.110
Nintendo Co.,Ltd.	3.645	0.803	0.017	0.550
Mitsubishi Corp.	3.213	0.597	0.004	0.432
ORIX	8.499	1.551	0.028	0.866
Nomura Holdings	4.341	0.471	0.007	0.398
Millea Holdings	2.663	1.511	0.035	1.353
Mitsubishi Estate	5.060	1.534	0.010	0.849
East Japan Railway	0.479	0.144	0.003	0.098
NTT	2.431	0.403	0.003	0.280
KDDI	2.171	0.372	0.007	0.333
NTT DoCoMo,Inc.	3.104	0.283	0.028	0.348
Tokyo Electric Power	0.149	0.112	0.002	0.099
SOFTBANK CORP.	11.671	7.381	0.023	5.385

Table 2 Empirical estimates $\hat{\eta}$

Asset	$\frac{\hat{\eta}_{12}}{\sqrt{\hat{\eta}_1}\sqrt{\hat{\eta}_2}}$	$\frac{\hat{\eta}_{13}}{\sqrt{\hat{\eta}_1}\sqrt{\hat{\eta}_3}}$	$\frac{\hat{\eta}_{14}}{\sqrt{\hat{\eta}_1}\sqrt{\hat{\eta}_4}}$	$\frac{\hat{\eta}_{23}}{\sqrt{\hat{\eta}_2}\sqrt{\hat{\eta}_3}}$	$\frac{\hat{\eta}_{24}}{\sqrt{\hat{\eta}_2}\sqrt{\hat{\eta}_4}}$	$\frac{\hat{\eta}_{34}}{\sqrt{\hat{\eta}_3}\sqrt{\hat{\eta}_4}}$
TOPIX	0.204	−0.028	0.182	0.272	0.514	0.233
JAPAN TOBACCO	0.220	0.134	0.148	0.277	0.622	0.360
Shin-Etsu Chemical	0.220	0.013	0.181	0.171	0.473	0.130
Takeda Pharm.	0.197	−0.011	0.099	0.094	0.388	0.137
Astellas Pharma Inc.	0.133	0.078	0.127	0.106	0.367	0.238
FUJIFILM Holdings	0.109	0.028	0.099	0.104	0.308	0.169
NIPPON STEEL	0.156	0.052	0.145	0.226	0.540	0.266
JFE Holdings,Inc.	0.117	−0.030	0.118	0.163	0.406	0.106
Hitachi,Ltd.	0.213	0.043	0.091	0.169	0.462	0.143
Matsushita	0.288	−0.022	0.195	0.190	0.444	0.145
SONY	0.175	−0.028	0.165	0.141	0.405	0.128
NISSAN MOTOR	0.229	0.052	0.146	0.223	0.484	0.198
TOYOTA	0.108	0.028	0.217	0.155	0.479	0.136
HONDA MOTOR	0.262	0.079	0.191	0.255	0.474	0.185
CANON INC.	0.163	−0.040	0.174	0.102	0.449	0.051
Nintendo Co.,Ltd.	0.136	0.140	0.123	0.120	0.293	0.088
Mitsubishi Corp.	0.250	−0.003	0.163	0.245	0.507	0.228
ORIX	0.159	0.047	0.181	0.277	0.520	0.283
Nomura Holdings	0.174	0.123	0.214	0.225	0.485	0.244
Millea Holdings	0.178	−0.057	0.054	0.120	0.485	0.238
Mitsubishi Estate	0.105	0.028	0.177	0.325	0.507	0.243
East Japan Railway	0.230	0.101	0.211	0.134	0.392	0.142
NTT	0.318	0.082	0.304	0.136	0.483	0.098
KDDI	0.252	0.139	0.165	0.173	0.429	0.136
NTT DoCoMo,Inc.	0.072	−0.007	0.143	0.053	0.199	−0.041
Tokyo Electric Power	0.362	0.104	0.295	0.200	0.705	0.177
SOFTBANK CORP.	0.213	0.145	0.270	0.322	0.606	0.317

Table 3 Empirical estimates of correlation

Asset	$\hat{\lambda}_1^*$	$\hat{\lambda}_2^*$	$\hat{\lambda}_3^*$	$\hat{\lambda}_4^*$
TOPIX	0.175	1.047	0.182	2.069
JAPAN TOBACCO	0.083	1.545	0.026	1.096
Shin-Etsu Chemical	0.039	1.427	0.140	1.476
Takeda Pharm.	0.025	1.762	0.152	1.017
Astellas Pharma Inc.	0.037	1.267	0.072	1.756
FUJIFILM Holdings	0.032	1.451	0.081	1.402
NIPPON STEEL	0.041	2.223	0.018	0.462
JFE Holdings,Inc.	0.067	1.786	0.176	1.023
Hitachi,Ltd.	0.081	0.959	0.259	2.444
Matsushita	0.041	1.033	0.165	2.524
SONY	0.023	1.015	0.120	2.263
NISSAN MOTOR	0.079	0.997	0.132	2.805
TOYOTA	0.031	1.345	0.113	1.619
HONDA MOTOR	0.014	1.256	0.040	1.971
CANON INC.	0.033	1.006	0.220	2.342
Nintendo Co.,Ltd.	0.193	1.063	0.149	2.619
Mitsubishi Corp.	0.079	1.149	0.227	2.074
ORIX	0.119	1.044	0.047	2.461
Nomura Holdings	0.084	1.182	0.072	2.372
Millea Holdings	0.048	1.958	0.114	0.564
Mitsubishi Estate	0.122	1.198	0.101	1.885
East Japan Railway	0.005	1.773	0.131	0.902
NTT	−0.019	1.173	0.281	1.885
KDDI	0.020	1.642	0.124	1.312
NTT DoCoMo,Inc.	−0.003	2.408	0.077	0.291
Tokyo Electric Power	−0.001	1.566	0.226	0.940
SOFTBANK CORP.	0.096	1.720	0.106	0.886
Min.	0.000	0.959	0.018	0.291
Max.	0.193	2.408	0.281	2.805
Average	0.058	1.407	0.132	1.647

Table 4 Empirical estimates $\hat{\lambda}^*$

	RV		RV_{w4}		Diff.
Code	Mean	Var.	Mean	Var.	Var.
TOPIX	0.632	0.209	0.632	0.195	0.014
JAPAN TOBACCO	5.208	19.292	5.208	17.443	1.849
Shin-Etsu Chemical	2.646	2.844	2.646	1.824	1.021
Takeda Pharm.	1.613	0.995	1.613	0.551	0.444
Astellas Pharma Inc.	2.872	3.348	2.872	1.699	1.649
FUJIFILM Holdings	2.557	1.916	2.557	0.980	0.935
NIPPON STEEL	3.613	3.896	3.613	3.011	0.885
JFE Holdings,Inc.	3.357	4.888	3.357	3.369	1.519
Hitachi,Ltd.	2.351	3.409	2.351	2.128	1.281
Matsushita	2.121	3.745	2.121	1.983	1.762
SONY	2.855	3.709	2.855	1.440	2.269
NISSAN MOTOR	2.048	2.996	2.048	1.713	1.284
TOYOTA	2.077	1.450	2.077	0.761	0.689
HONDA MOTOR	2.548	3.228	2.548	1.455	1.773
CANON INC.	2.168	2.394	2.168	1.007	1.387
Nintendo Co.,Ltd.	2.810	6.335	2.810	6.187	0.148
Mitsubishi Corp.	2.980	5.880	2.980	4.033	1.846
ORIX	4.208	14.533	4.208	10.600	3.933
Nomura Holdings	3.090	6.792	3.090	4.278	2.514
Millea Holdings	5.842	7.997	5.842	7.850	0.147
Mitsubishi Estate	3.896	10.070	3.896	8.180	1.890
East Japan Railway	1.574	1.047	1.574	0.684	0.363
NTT	2.715	4.602	2.715	2.244	2.358
KDDI	2.727	3.986	2.727	2.258	1.728
NTT DoCoMo,Inc.	5.669	4.316	5.669	1.758	2.558
Tokyo Electric Power	1.312	0.688	1.312	0.584	0.103
SOFTBANK CORP.	7.374	40.986	7.374	38.900	2.086

Table 5 Mean and variance of RVs

Asset	λ_1^*	λ_2^*	λ_3^*	λ_4^*
TOPIX	0.481	1.314	–	–
JAPAN TOBACCO	0.342	1.236	–	–
Shin-Etsu Chemical	0.136	1.407	–	–
Takeda Pharm.	0.096	1.371	–	–
Astellas Pharma Inc.	0.126	1.452	–	–
FUJIFILM Holdings	0.121	1.389	–	–
NIPPON STEEL	0.175	1.281	–	–
JFE Holdings,Inc.	0.238	1.344	–	–
Hitachi,Ltd.	0.212	1.561	–	–
Matsushita	0.090	1.705	–	–
SONY	0.061	1.597	–	–
NISSAN MOTOR	0.172	1.766	–	–
TOYOTA	0.109	1.447	–	–
HONDA MOTOR	0.038	1.592	–	–
CANON INC.	0.084	1.634	–	–
Nintendo Co.,Ltd.	0.469	1.460	–	–
Mitsubishi Corp.	0.208	1.492	–	–
ORIX	0.298	1.519	–	–
Nomura Holdings	0.199	1.642	–	–
Millea Holdings	0.260	1.190	–	–
Mitsubishi Estate	0.367	1.356	–	–
East Japan Railway	0.047	1.361	–	–
NTT	−0.043	1.530	–	–
KDDI	0.088	1.456	–	–
NTT DoCoMo,Inc.	−0.006	1.269	–	–
Tokyo Electric Power	0.033	1.244	–	–
SOFTBANK CORP.	0.388	1.228	–	–

Table 6 Empirical Estimates λ^* for $RV_{w2,k}$

Asset	λ_1^*	λ_2^*	λ_3^*	λ_4^*
TOPIX	0.181	1.096	1.998	–
JAPAN TOBACCO	0.092	1.393	1.251	–
Shin-Etsu Chemical	0.044	1.421	1.483	–
Takeda Pharm.	0.028	1.733	1.051	–
Astellas Pharma Inc.	0.042	1.206	1.821	–
FUJIFILM Holdings	0.037	1.370	1.483	–
NIPPON STEEL	0.044	2.076	0.600	–
JFE Holdings,Inc.	0.073	1.827	0.979	–
Hitachi,Ltd.	0.091	1.017	2.371	–
Matsushita	0.041	1.063	2.498	–
SONY	0.023	1.028	2.253	–
NISSAN MOTOR	0.083	1.010	2.786	–
TOYOTA	0.035	1.341	1.623	–
HONDA MOTOR	0.016	1.226	2.002	–
CANON INC.	0.034	1.064	2.290	–
Nintendo Co.,Ltd.	0.217	1.054	2.576	–
Mitsubishi Corp.	0.081	1.213	2.001	–
ORIX	0.125	1.010	2.490	–
Nomura Holdings	0.089	1.162	2.385	–
Millea Holdings	0.057	1.745	0.766	–
Mitsubishi Estate	0.129	1.209	1.861	–
East Japan Railway	0.013	1.700	0.989	–
NTT	−0.014	1.266	1.792	–
KDDI	0.028	1.614	1.341	–
NTT DoCoMo,Inc.	−0.001	2.206	0.467	–
Tokyo Electric Power	0.012	1.562	0.951	–
SOFTBANK CORP.	0.104	1.714	0.886	–

Table 7 Empirical Estimates λ^* for $RV_{w3,k}$

Asset	RV	RV_{w2}	RV_{w3}	RV_{w4}
	d	d	d	d
TOPIX	0.499	0.520	0.423	0.520
JAPAN TOBACCO	0.601	0.763	0.745	0.786
Shin-Etsu Chemical	0.508	0.508	0.564	0.503
Takeda Pharm.	0.337	0.455	0.290	0.490
Astellas Pharma Inc.	0.338	0.542	0.413	0.559
FUJIFILM Holdings	0.232	0.372	0.603	0.374
NIPPON STEEL	0.624	0.705	0.311	0.660
JFE Holdings,Inc.	0.431	0.520	0.199	0.505
Hitachi,Ltd.	0.450	0.448	0.505	0.461
Matsushita	0.425	0.676	0.522	0.694
SONY	0.403	0.447	0.319	0.441
NISSAN MOTOR	0.446	0.561	0.361	0.495
TOYOTA	0.477	0.557	0.417	0.550
HONDA MOTOR	0.532	0.532	0.291	0.516
CANON INC.	0.345	0.397	0.434	0.389
Nintendo Co.,Ltd.	0.403	0.454	0.225	0.397
Mitsubishi Corp.	0.492	0.547	0.659	0.571
ORIX	0.591	0.659	0.465	0.674
Nomura Holdings	0.605	0.679	0.251	0.691
Millea Holdings	0.500	0.534	0.629	0.504
Mitsubishi Estate	0.537	0.668	0.453	0.688
East Japan Railway	0.345	0.338	0.367	0.353
NTT	0.133	0.287	0.426	0.299
KDDI	0.450	0.373	0.489	0.358
NTT DoCoMo,Inc.	0.119	0.246	0.001	0.184
Tokyo Electric Power	0.605	0.649	0.401	0.649
SOFTBANK CORP.	0.480	0.524	0.574	0.528

Table 8 Estimates of d in ARFIMA (p, d, q) model

RV		RV_{w2}		RV_{w3}		RV_{w4}	
p	q	p	q	p	q	p	q
0	1	0	1	0	1	0	1
1	1	1	1	0	2	1	1
1	1	0	2	0	1	0	2
0	1	0	2	0	1	0	2
0	1	1	1	0	1	1	1
1	0	1	0	0	1	0	1
0	1	1	1	1	0	1	1
0	1	0	2	0	1	0	2
1	1	0	1	1	1	2	0
0	1	1	1	0	1	1	1
0	1	1	1	0	1	0	1
0	1	0	1	0	1	0	1
1	1	1	1	0	1	1	1
1	1	0	1	1	0	0	1
1	1	0	1	2	0	0	1
0	1	0	1	0	1	0	1
0	2	0	1	0	2	0	1
1	1	1	1	0	1	1	1
0	1	0	1	0	1	1	1
0	1	0	1	0	1	0	1
1	1	1	1	0	1	1	1
0	1	1	0	0	1	1	0
1	1	1	0	0	1	0	1
0	1	1	0	0	1	1	0
1	0	0	1	1	0	1	0
0	1	0	1	0	1	0	1
0	1	1	1	1	1	1	1

Table 9 Estimates of p and q

Asset	RV	RV_{w2}	RV_{w3}	RV_{w4}
TOPIX	0.387	0.328	0.332	0.433
JAPAN TOBACCO	3.330	2.695	2.746	2.868
Shin-Etsu Chemical	1.505	1.121	1.131	1.131
Takeda Pharm.	0.931	0.619	0.626	0.627
Astellas Pharma Inc.	1.708	1.144	1.175	1.145
FUJIFILM Holdings	1.341	0.893	0.896	0.922
NIPPON STEEL	1.667	1.268	1.396	1.427
JFE Holdings,Inc.	2.021	1.579	1.562	1.558
Hitachi,Ltd.	1.656	1.106	1.133	1.143
Matsushita	1.721	1.046	1.093	1.096
SONY	1.805	0.996	1.021	1.021
NISSAN MOTOR	1.593	0.996	1.106	1.099
TOYOTA	1.120	0.700	0.701	0.701
HONDA MOTOR	1.635	1.001	1.007	1.074
CANON INC.	1.465	0.833	0.854	0.857
Nintendo Co.,Ltd.	2.374	2.060	2.247	2.266
Mitsubishi Corp.	2.046	1.465	1.506	1.514
ORIX	3.216	2.473	2.584	2.582
Nomura Holdings	2.279	1.388	1.424	1.423
Millea Holdings	2.303	1.942	2.142	2.261
Mitsubishi Estate	2.633	2.051	2.092	2.097
East Japan Railway	0.932	0.694	0.716	0.724
NTT	1.978	1.301	1.286	1.288
KDDI	1.810	1.260	1.272	1.273
NTT DoCoMo,Inc.	2.037	1.035	1.207	1.281
Tokyo Electric Power	0.500	0.525	0.387	0.388
SOFTBANK CORP.	4.849	4.390	4.770	4.782

Table 10 RMSE for ARFIMA Model

Asset	RV_{w2}/RV	RV_{w3}/RV	RV_{w4}/RV
TOPIX	0.849	0.858	1.118
JAPAN TOBACCO	0.809	0.825	0.861
Shin-Etsu Chemical	0.745	0.752	0.751
Takeda Pharm.	0.665	0.672	0.673
Astellas Pharma Inc.	0.670	0.688	0.670
FUJIFILM Holdings	0.666	0.668	0.687
NIPPON STEEL	0.760	0.838	0.856
JFE Holdings,Inc.	0.781	0.773	0.771
Hitachi,Ltd.	0.668	0.684	0.690
Matsushita	0.608	0.635	0.637
SONY	0.552	0.565	0.566
NISSAN MOTOR	0.625	0.694	0.690
TOYOTA	0.626	0.626	0.626
HONDA MOTOR	0.612	0.615	0.656
CANON INC.	0.568	0.583	0.585
Nintendo Co.,Ltd.	0.868	0.947	0.955
Mitsubishi Corp.	0.716	0.736	0.740
ORIX	0.769	0.804	0.803
Nomura Holdings	0.609	0.625	0.624
Millea Holdings	0.843	0.930	0.982
Mitsubishi Estate	0.779	0.794	0.796
East Japan Railway	0.745	0.768	0.777
NTT	0.658	0.650	0.651
KDDI	0.696	0.703	0.704
NTT DoCoMo,Inc.	0.508	0.593	0.629
Tokyo Electric Power	1.050	0.774	0.776
SOFTBANK CORP.	0.905	0.984	0.986

Table 11 RMSE Ratios between RV and RV_{w}.

Asset	RV	RV_{w2}	RV_{w3}	RV_{w4}
TOPIX	0.283	0.229	0.224	0.239
JAPAN TOBACCO	2.083	1.729	1.716	1.766
Shin-Etsu Chemical	1.064	0.805	0.813	0.813
Takeda Pharm.	0.637	0.457	0.463	0.463
Astellas Pharma Inc.	1.201	0.877	0.890	0.876
FUJIFILM Holdings	0.974	0.696	0.700	0.710
NIPPON STEEL	1.168	0.906	0.982	1.000
JFE Holdings,Inc.	1.399	1.142	1.186	1.181
Hitachi,Ltd.	1.117	0.800	0.812	0.817
Matsushita	1.060	0.726	0.769	0.771
SONY	1.218	0.723	0.737	0.737
NISSAN MOTOR	1.081	0.722	0.793	0.789
TOYOTA	0.799	0.502	0.502	0.502
HONDA MOTOR	1.100	0.745	0.754	0.767
CANON INC.	1.006	0.630	0.647	0.651
Nintendo Co.,Ltd.	1.630	1.496	1.563	1.573
Mitsubishi Corp.	1.333	1.022	1.021	1.022
ORIX	2.159	1.707	1.788	1.787
Nomura Holdings	1.505	1.028	1.056	1.056
Millea Holdings	1.713	1.465	1.627	1.720
Mitsubishi Estate	1.808	1.435	1.433	1.436
East Japan Railway	0.658	0.531	0.542	0.546
NTT	1.273	0.934	0.920	0.919
KDDI	1.211	0.935	0.946	0.947
NTT DoCoMo,Inc.	1.354	0.801	0.922	0.975
Tokyo Electric Power	0.340	0.290	0.269	0.270
SOFTBANK CORP.	3.162	2.875	3.044	3.049

Table 12 MAE for ARFIMA Model

Asset	RV_{w2}/RV	RV_{w3}/RV	RV_{w4}/RV
TOPIX	0.809	0.792	0.846
JAPAN TOBACCO	0.830	0.824	0.848
Shin-Etsu Chemical	0.757	0.765	0.764
Takeda Pharm.	0.717	0.727	0.728
Astellas Pharma Inc.	0.730	0.741	0.729
FUJIFILM Holdings	0.714	0.718	0.729
NIPPON STEEL	0.776	0.841	0.856
JFE Holdings,Inc.	0.817	0.848	0.845
Hitachi,Ltd.	0.717	0.727	0.731
Matsushita	0.685	0.725	0.727
SONY	0.594	0.605	0.605
NISSAN MOTOR	0.668	0.734	0.730
TOYOTA	0.629	0.628	0.628
HONDA MOTOR	0.677	0.685	0.697
CANON INC.	0.626	0.644	0.647
Nintendo Co.,Ltd.	0.918	0.959	0.965
Mitsubishi Corp.	0.767	0.766	0.767
ORIX	0.791	0.828	0.828
Nomura Holdings	0.683	0.702	0.701
Millea Holdings	0.855	0.950	1.004
Mitsubishi Estate	0.793	0.792	0.794
East Japan Railway	0.807	0.824	0.830
NTT	0.734	0.723	0.722
KDDI	0.772	0.781	0.782
NTT DoCoMo,Inc.	0.592	0.681	0.721
Tokyo Electric Power	0.853	0.792	0.793
SOFTBANK CORP.	0.909	0.963	0.964

Table 13 MAE Ratios between RV and RV_{w}.

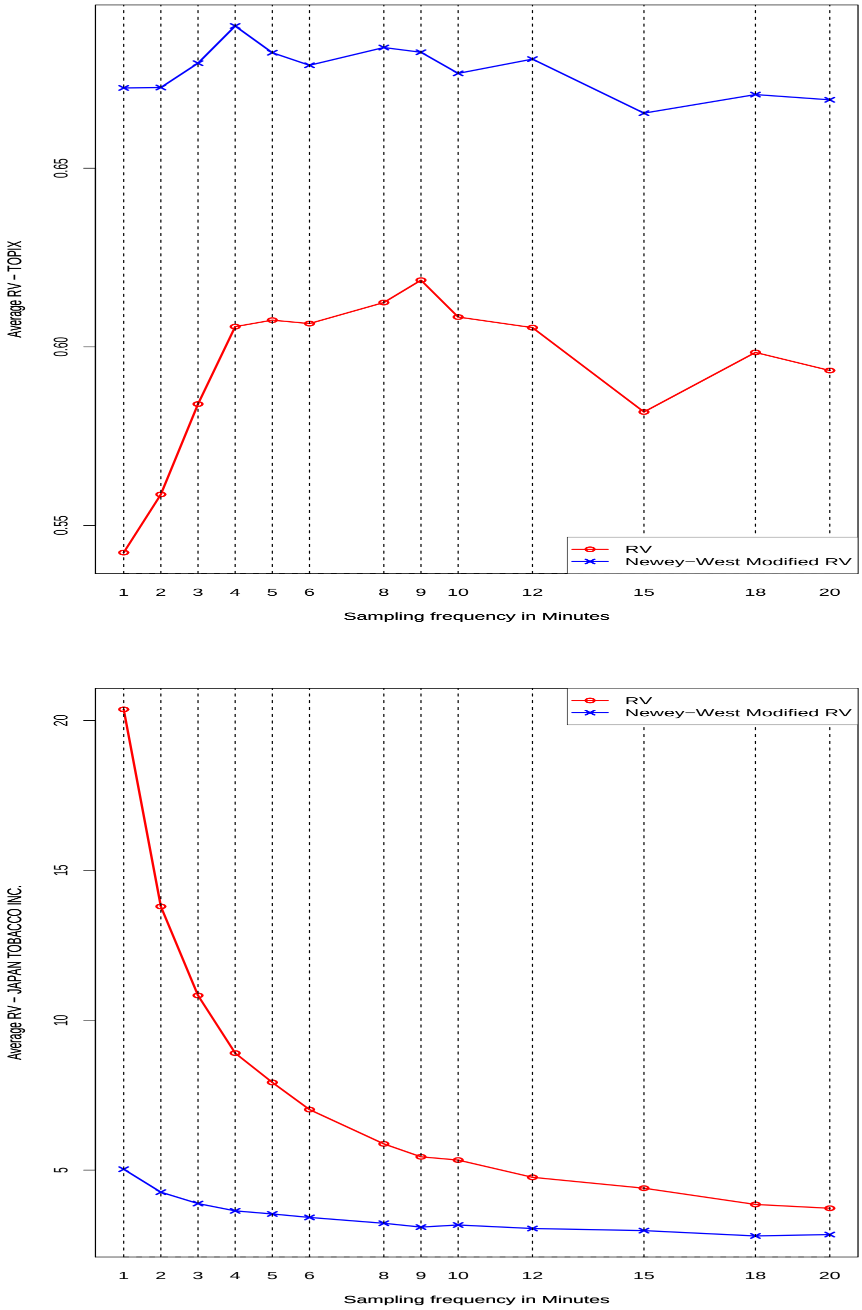


Figure 1 Volatility signature plot (TOPIX and JAPAN TOBACCO)

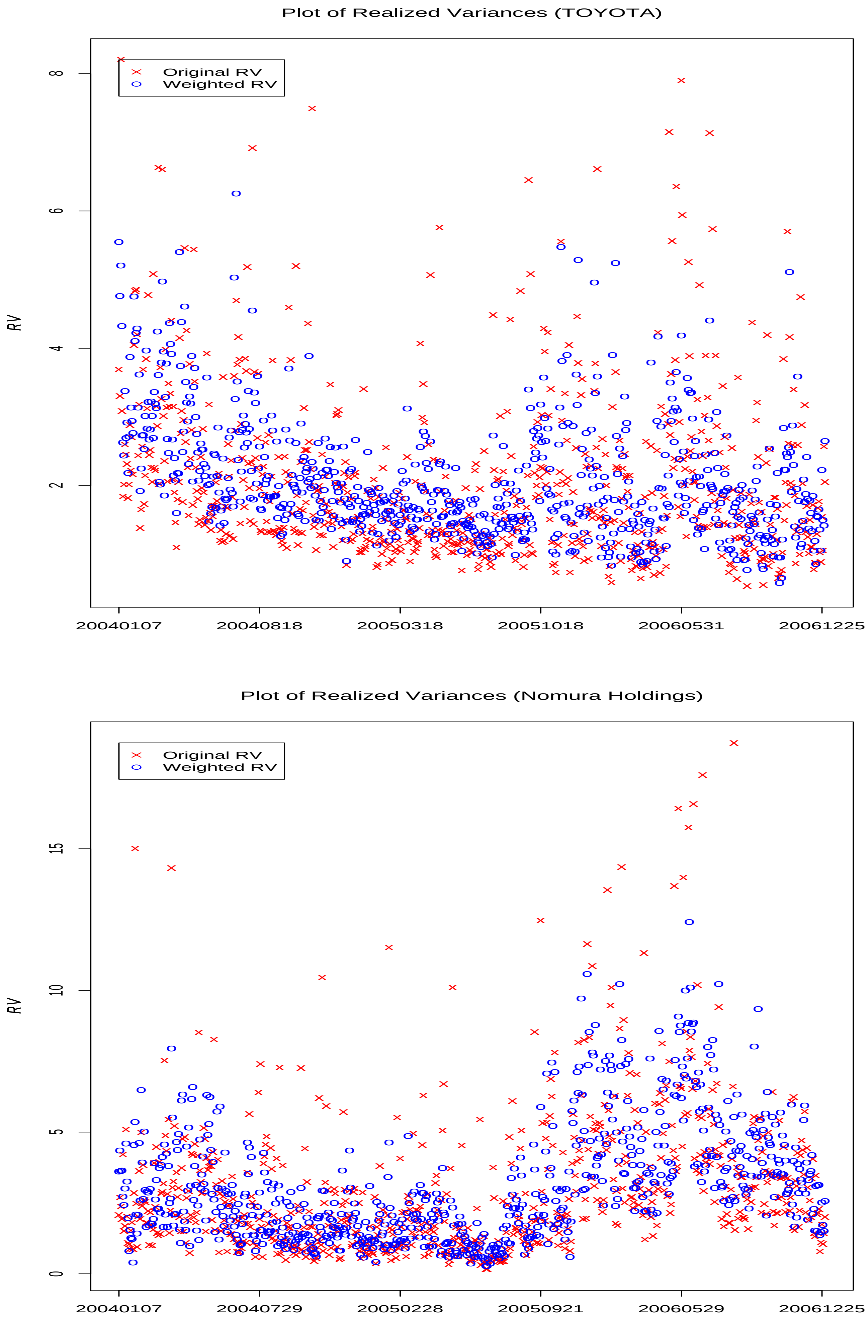
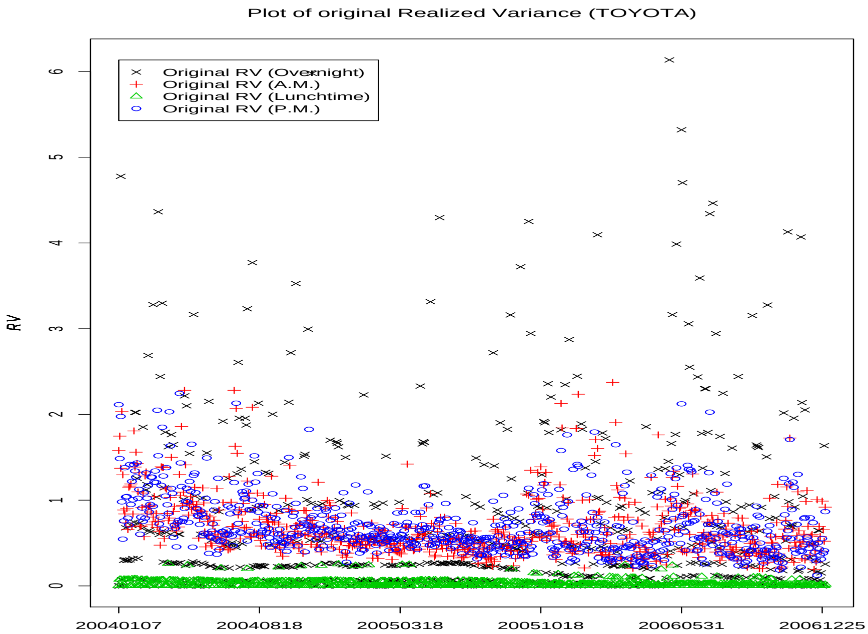


Figure 2 Realized variance (TOYOTA and Nomura Holdings)



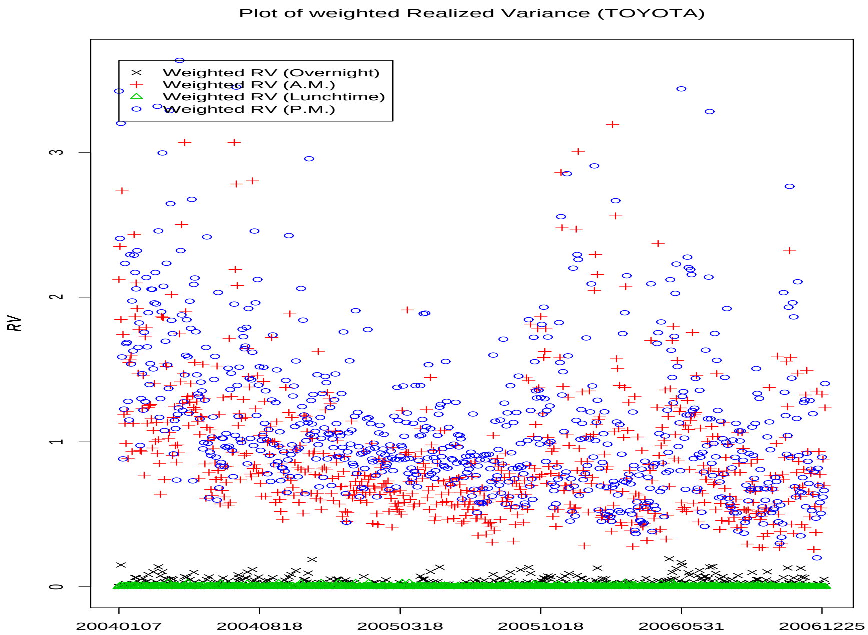


Figure 3 Original and weighted RV (TOYOTA)