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: October 19, 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 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 concludes.

2 An optimal weighting procedure

In this section we will present a conditional variant of the argument in Hansen and Lunde (2005, Section 2), so that we can perform empirical studies of cases where there are more than two inactive periods. Under a formal setup, which is model-free about an underlying stock-price process, we will construct an optimal weight under a kind of conditional-proportionality assumption.

2.1 Objective

First let us recall the structure of Japanese market more precisely. The market opens at 9:00 and closes at 15:00 (at each weekday) with lunch break 11:00∼12:30.

Let $T > 0$ represent 24-hours length expediently, and let kT stand for the closing time 15:00 on kth day. Let $\mathcal{G}$ denote "all available information up to time $(k-1)T$", such as all observed price data up to time $(k-1)T$. Then we want to estimate the kth-day integrated (accumulated) volatility V_k, i.e., the volatility over the kth-day period $I_k := [(k-1)T, kT]$, based on intermittent high-frequency data over I_k together with the available information $\mathcal{G}$. Each I_k can be split into four sub periods:

$$I_k = \bigcup_{i=1}^{4} I_{k,i}, \quad \text{where} \quad I_{k,i} := [T_{k,i-1}, T_{k,i}].$$

Here $(k-1)T := T_{k,0} < T_{k,1} < T_{k,2} < T_{k,3} < T_{k,4} := kT(= T_{k+1,0})$, where $I_{k,1}$ denotes nighttime period, $I_{k,2}$ morning trading time, $I_{k,3}$ lunch break, and $I_{k,4}$ afternoon trading time.

If $V_{k,i}$ stands for the integrated volatility over $I_{k,i} := [T_{k,i}, T_{k,i-1}]$, then, in view of the additive character of the integrated volatility we have $V_k = \sum_{i=1}^{4} V_{k,i}$. Denote by $X = (X_t)_{t\in\mathbb{R}}$ the underlying log-price process. Then, a common estimator of V_k is given by the *naive RV*

$$RV_k := \sum_{i=1}^{4} \hat{V}_{k,i},$$

where

$$\begin{aligned}
\hat{V}_{k,1} &:= (X_{T_{k,1}} - X_{T_{k,0}})^2, \\
\hat{V}_{k,2} &:= (\text{realized volatility over } I_{k,2}), \\
\hat{V}_{k,3} &:= (X_{T_{k,3}} - X_{T_{k,2}})^2, \\
\hat{V}_{k,4} &:= (\text{realized volatility over } I_{k,4}),
\end{aligned}$$

all of which are direct to compute, of course. When estimating V_k using past daily time series of each $\hat{V}_{k,i}$, it is clear that $\hat{V}_{k,1}$ and $\hat{V}_{k,3}$ exhibit much larger variances compared with $\hat{V}_{k,2}$ and $\hat{V}_{k,4}$, due to the lack of high-frequency data therein. Note that, at the same time, we should not simply ignore the fluctuations over each $I_{k,1}$ and $I_{k,3}$ in general, as they often have non-negligible impact for the target variable V_k.

Instead of RV_k, we will consider a (randomly) weighted version of the form

$$RV_k(\lambda) := \sum_{i=1}^{4} \lambda_i \hat{V}_{k,i}$$

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

$$\lambda \mapsto \mathrm{MSE}(\lambda) := P_{\mathcal{G}}[|RV_k(\lambda) - V_k|^2],$$

where $P_{\mathcal{G}}[v]$ stands for the conditional expectation of v given $\mathcal{G}$. Unfortunately, it is impossible to get an empirical variant of the optimal λ as V_k cannot be observed. Following the approach taken in Hansen and Lunde (2005), we will impose a kind of proportionality assumption to obtain a closed-form solution to this optimization problem.

2.2 Derivation under conditional proportionality

Building on the argument in the previous subsection, we will derive the explicit form of λ^* given above within a more formal setup.

Fix a complete probability space $(\Omega, \mathcal{F}, P)$. Given any natural number $m \geq 2$ (say $m = m' + m''$, where m' is the number of inactive periods of tradings, and m'' is that of active periods where we can get reasonably high-frequency data). Let V and 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 variable to be estimated based on all available information, and we want to find the optimal weight $\lambda^* = (\lambda_i^*)_{i \leq m}$, which a.s. minimizes the $\mathcal{G}$-conditional mean square error given by

$$\lambda \mapsto \mathrm{MSE}_{\mathcal{G}}(\lambda) := P_{\mathcal{G}}[|\hat{V}(\lambda) - V|^2].$$

Here, as before, we will focus on $\hat{V}(\lambda)$ of the form

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

with $\lambda = (\lambda_j)_{j \leq m} \in \Lambda$, where the random index set Λ is defined by

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

with $\mathcal{G}$-measurable random variables

$$\mu_0 = P_{\mathcal{G}}[V], \quad \mu_i = P_{\mathcal{G}}[V_i], \quad \text{and} \quad \eta_{ij}^2 = cov_{\mathcal{G}}[V_i, V_j].$$

Supposing $\mu_i > 0$ a.s. we set

$$\gamma_{i,j} = \frac{\eta_{i,j}^2}{\mu_i \mu_j}.$$

With the setup above, we are going to derive the explicit form of $\lambda^* \in \Lambda$ under an additional assumption of a kind of $\mathcal{H}$-conditional proportionality of 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 "almost surely (under P)" 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

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

Then, by taking the conditional expectation $P_{\mathcal{H}}$ in (1) we have

$$P_{\mathcal{H}}[\hat{V}(\lambda)] = \sum_{i=1}^{m} \lambda_i \rho_i V, \tag{2}$$

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

$$P_{\mathcal{G}}[\hat{V}(\lambda)] = \mu_0 \sum_{i=1}^{m} \lambda_i \rho_i. \tag{3}$$

On the other hand, taking $P_{\mathcal{G}}$ in (1) yields that

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

for $\lambda \in \Lambda$. Equating the right-hand sides of (3) and (4) yields $\sum_{i=1}^{m} \lambda_i \rho_i = 1$ for $\lambda \in \Lambda$. Therefore, from (2) we get

$$P_{\mathcal{H}}[\hat{V}(\lambda)] = V \tag{5}$$

(hence, in particular $P_{\mathcal{G}}[\hat{V}(\lambda)] = \mu_0$) for $\lambda \in \Lambda$. According to (5) and simple conditioning argument we get

$$\begin{aligned} P_{\mathcal{G}}[|\hat{V}(\lambda) - V|^2] &= var_{\mathcal{G}}[\hat{V}(\lambda)] - 2P_{\mathcal{G}}[\{\hat{V}(\lambda) - V\}(V - \mu_0)] - var_{\mathcal{G}}[V] \\ &= var_{\mathcal{G}}[\hat{V}(\lambda)] - var_{\mathcal{G}}[V] \end{aligned}$$

for $\lambda \in \Lambda$, so that we arrive at

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

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

For any $\lambda = (\lambda_i)_{i \le m} \in \Lambda$ we can set

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

Observe that

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

For each $i \in \{1, \ldots, m-1\}$, elementary computations lead to

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

where

$$d_{i,j} := \mu_i \mu_j (\gamma_{m,m} + \gamma_{i,j} - \gamma_{i,m} - \gamma_{j,m}),$$
$$b_i := \mu_0 \mu_i (\gamma_{m,m} - \gamma_{i,m})$$

for $1 \le i, j \le m-1$. In view of the first-order condition $\nabla_{(\lambda_1,\dots,\lambda_{m-1})}\zeta(\lambda_1,\dots,\lambda_{m-1}) = 0$ and the definition of Λ, we see that for $\lambda \in \Lambda$ 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

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

Summarizing the above now yields the following assertion.

Theorem. *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*

$$P_{\mathcal{H}}[V_i] = \rho_i V, \quad \text{a.s.} \tag{6}$$

Then, the $\mathcal{G}$-measurable function $\lambda \mapsto P_{\mathcal{G}}[|\hat{V}(\lambda) - V|^2]$ defined on Λ is a.s. minimized by $\lambda^* = \mathrm{argmin}_{\lambda\in\Lambda} var_{\mathcal{G}}[\hat{V}(\lambda)]$, which is in turn explicitly given by a solution of $D\lambda = b$. In particular, $\lambda^* = D^{-1}b$ if further D is invertible.

This theorem formulates a conditional and multi-intermittence version of Hansen and Lunde (2–5, Sections 2.2 and 2.3); again let us note that their result corresponds to the case where $m = 2$ and $\mathcal{G} = \{\phi, \Omega\}$ in our framework. Apart from the ad-hoc assumption (6), which cannot be suppressed for computing the λ^* without involving the latent variable V, our task toward positive analysis is to evaluate $\mathcal{G}$-measurable random variables $(\mu_i)_{i=0}^m$ and $[\eta_{i,j}]_{i,j=1}^m$, and of course this in principle requires specification of underlying model structure and forms of V_i as well as their relation to V. In the empirical study given in the next section, where $m = 4$, we will simply utilize the sample-mean type estimators for evaluations of the quantities $(\mu_i)_{i=0}^4$ and $[\eta_{i,j}]_{i,j=1}^4$ as in Hansen and Lunde (2005).

Remark. *The condition (6) is in principle impossible to verify without knowledge of the underlying price process X; see Hansen and Lunde (2005, Section 3.2) for an ad-hoc verification procedure of the proportionality condition. In practice, instead mere increments of X, one may use those of*

X centered by the sample mean in each period. In our empirical study, we will simply adopt this, however, we also observed that the gap between the cases of centered and non-centered versions was negligible, rendering that the condition (6) is not so irrelevant.

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 ex-

*These two sessions are respectively called "zenba" and "goba".

treme values. So we filter the raw data for outliers. We classify 1% of the observations $Y_{\cdot,1}$, $Y_{\cdot,2}$, $Y_{\cdot,3}$, and $Y_{\cdot,4}$ 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 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 2.2 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 for JAPAN TOBACCO, we take 0.1% data as outliers.

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

as expected values and variances. Let $\hat{V}_{k,1} = Y_{k,1}^2$, $\hat{V}_{k,2} = RV_{k,2}$, $\hat{V}_{k,3} = Y_{k,3}^2$, and $\hat{V}_{k,4} = RV_{k,4}$, then

$$
\begin{aligned}
\hat{\mu}_0 &= \frac{1}{n}\sum_{t=1}^{n}(\hat{V}_{t,1} + \hat{V}_{t,2} + \hat{V}_{t,3} + \hat{V}_{t,4}),\\
\hat{\mu}_i &= \frac{1}{n}\sum_{t=1}^{n}\hat{V}_{t,i}, \quad i = 1,2,3,4,\\
\hat{\eta}_i &= \frac{1}{n}\sum_{t=1}^{n}(\hat{V}_{t,i} - \hat{\mu}_i)^2, \quad i = 1,2,3,4,\\
\hat{\eta}_{i,j} &= \frac{1}{n}\sum_{t=1}^{n}(\hat{V}_{t,i} - \hat{\mu}_i)(\hat{V}_{t,j} - \hat{\mu}_j), \quad i,j = 1,2,3,4,
\end{aligned}
$$

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 $\hat{\lambda}^* = (\hat{\lambda}_i^*)_{i \le 4}$ 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] [§]

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

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 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.

4 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 $(V_j)_{j \le m}$. 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.

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.

- 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.

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 PHARMACEUTICAL	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 CORPORATION	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 PHARMACEUTICAL	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 CORPORATION	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 PHARMACEUTICAL	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 CORPORATION	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 PHARMACEUTICAL	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 CORPORATION	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	0.000	1.173	0.281	1.885
KDDI	0.020	1.642	0.124	1.312
NTT DOCOMO,INC.	0.000	2.408	0.077	0.291
TOKYO ELECTRIC POWER	0.000	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 $\hat{\lambda}^*$

Code	RV		RV_{weighted}		Diff.	Diff.
	Mean	Var.	Mean	Var.	Mean	Var.
TOPIX	0.632	0.209	0.632	0.195	0.000	0.014
JAPAN TOBACCO	5.208	19.292	5.208	17.443	0.000	1.849
SHIN-ETSU CHEMICAL	2.646	2.844	2.646	1.824	0.000	1.021
TAKEDA PHARMACEUTICAL	1.613	0.995	1.613	0.551	0.000	0.444
ASTELLAS PHARMA INC.	2.872	3.348	2.872	1.699	0.000	1.649
FUJIFILM HOLDINGS	2.557	1.916	2.557	0.980	0.000	0.935
NIPPON STEEL	3.613	3.896	3.613	3.011	0.000	0.885
JFE HOLDINGS,INC.	3.357	4.888	3.357	3.369	0.000	1.519
HITACHI,LTD.	2.351	3.409	2.351	2.128	0.000	1.281
MATSUSHITA	2.121	3.745	2.121	1.983	0.000	1.762
SONY	2.855	3.709	2.855	1.440	0.000	2.269
NISSAN MOTOR	2.048	2.996	2.048	1.713	0.000	1.284
TOYOTA	2.077	1.450	2.077	0.761	0.000	0.689
HONDA MOTOR	2.548	3.228	2.548	1.455	0.000	1.773
CANON INC.	2.168	2.394	2.168	1.007	0.000	1.387
NINTENDO CO.,LTD.	2.810	6.335	2.810	6.187	0.000	0.148
MITSUBISHI CORPORATION	2.980	5.880	2.980	4.033	0.000	1.846
ORIX	4.208	14.533	4.208	10.600	0.000	3.933
NOMURA HOLDINGS	3.090	6.792	3.090	4.278	0.000	2.514
MILLEA HOLDINGS	5.842	7.997	5.842	7.850	0.000	0.147
MITSUBISHI ESTATE	3.896	10.070	3.896	8.180	0.000	1.890
EAST JAPAN RAILWAY	1.574	1.047	1.574	0.684	0.000	0.363
NTT	2.715	4.602	2.715	2.244	0.000	2.358
KDDI	2.727	3.986	2.727	2.258	0.000	1.728
NTT DOCOMO,INC.	5.669	4.316	5.669	1.758	0.000	2.558
TOKYO ELECTRIC POWER	1.312	0.688	1.312	0.584	0.000	0.103
SOFTBANK CORP.	7.374	40.986	7.374	38.900	0.000	2.086

Table 5 Mean and variance of RVs

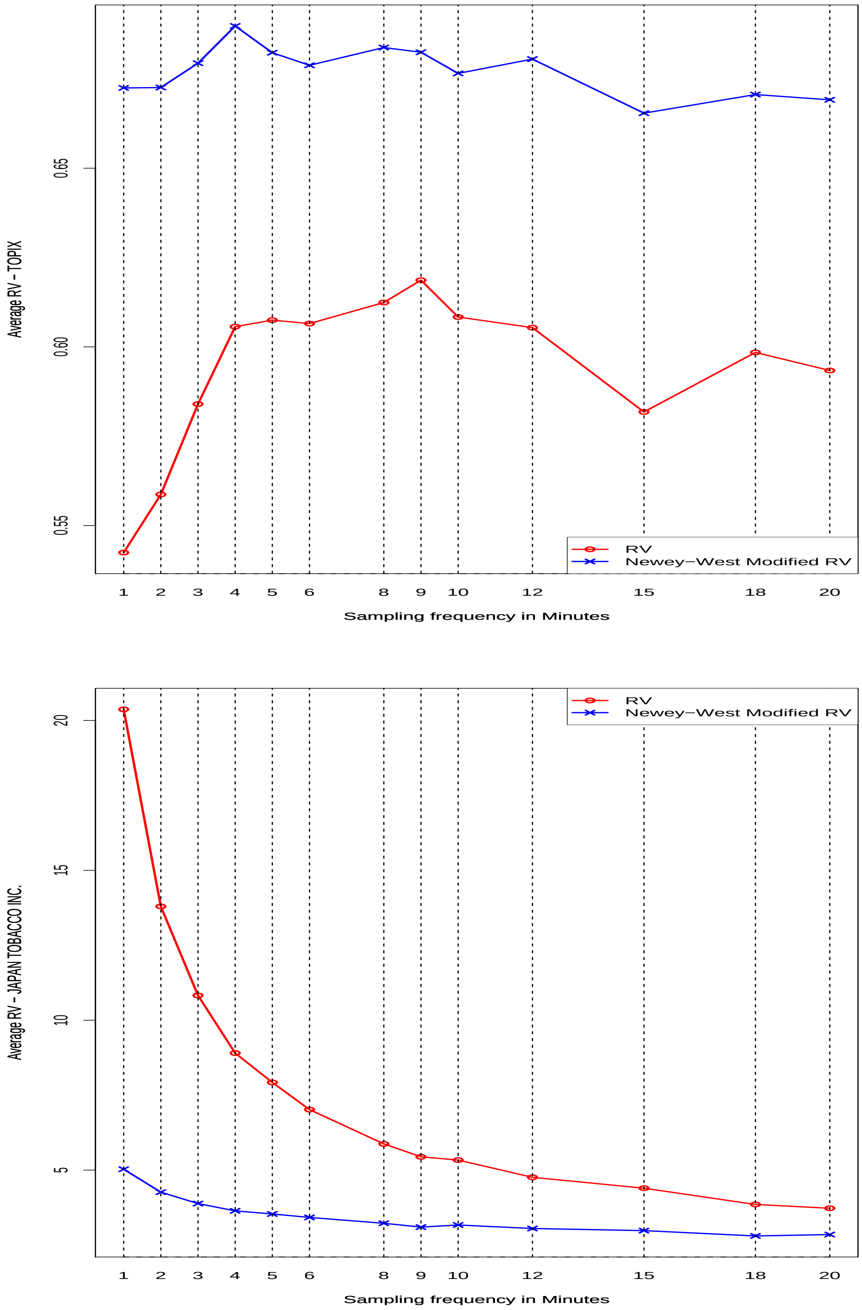

Figure 1 Volatility signature plot (TOPIX and JAPAN TOBACCO)

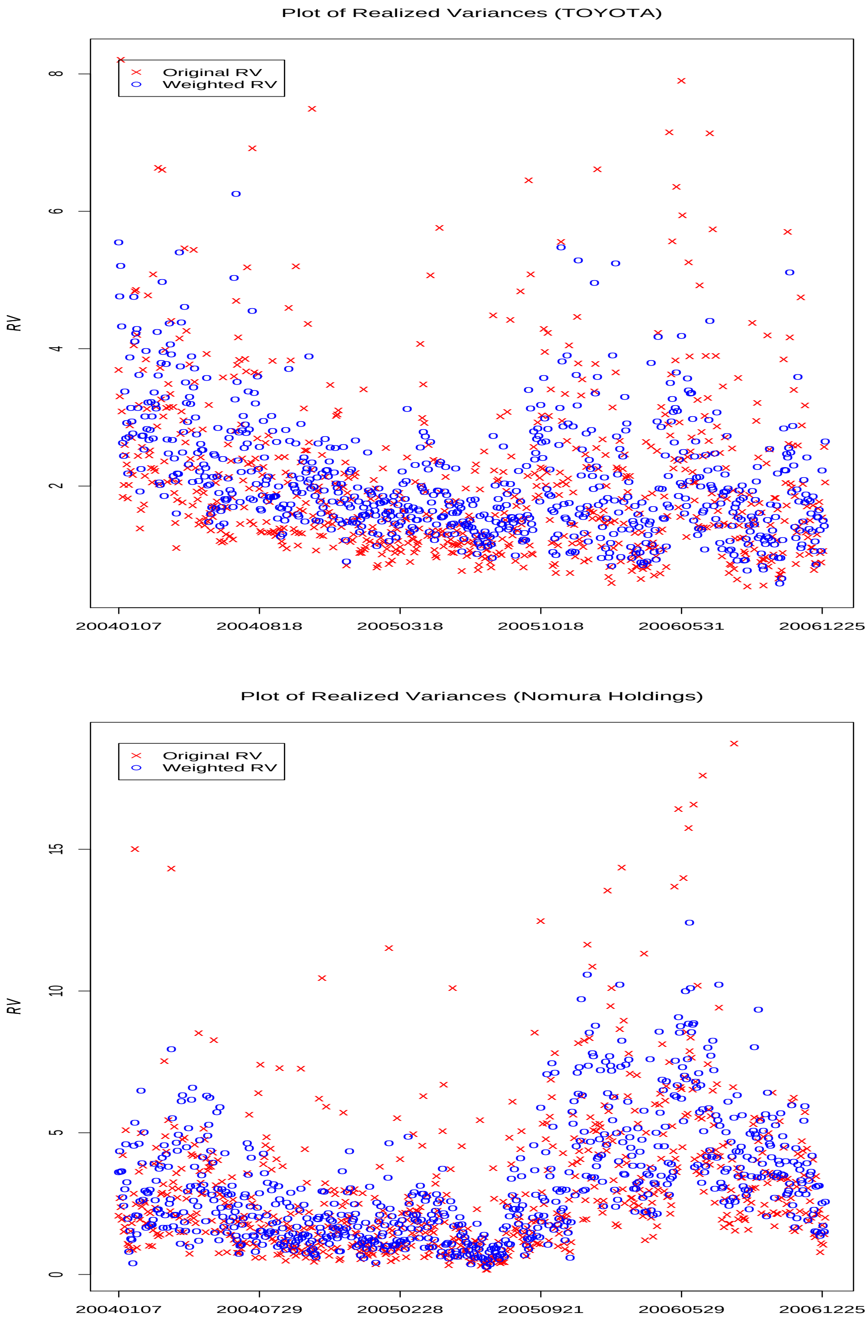

Figure 2 Realized variance (TOYOTA and Nomura Holdings)

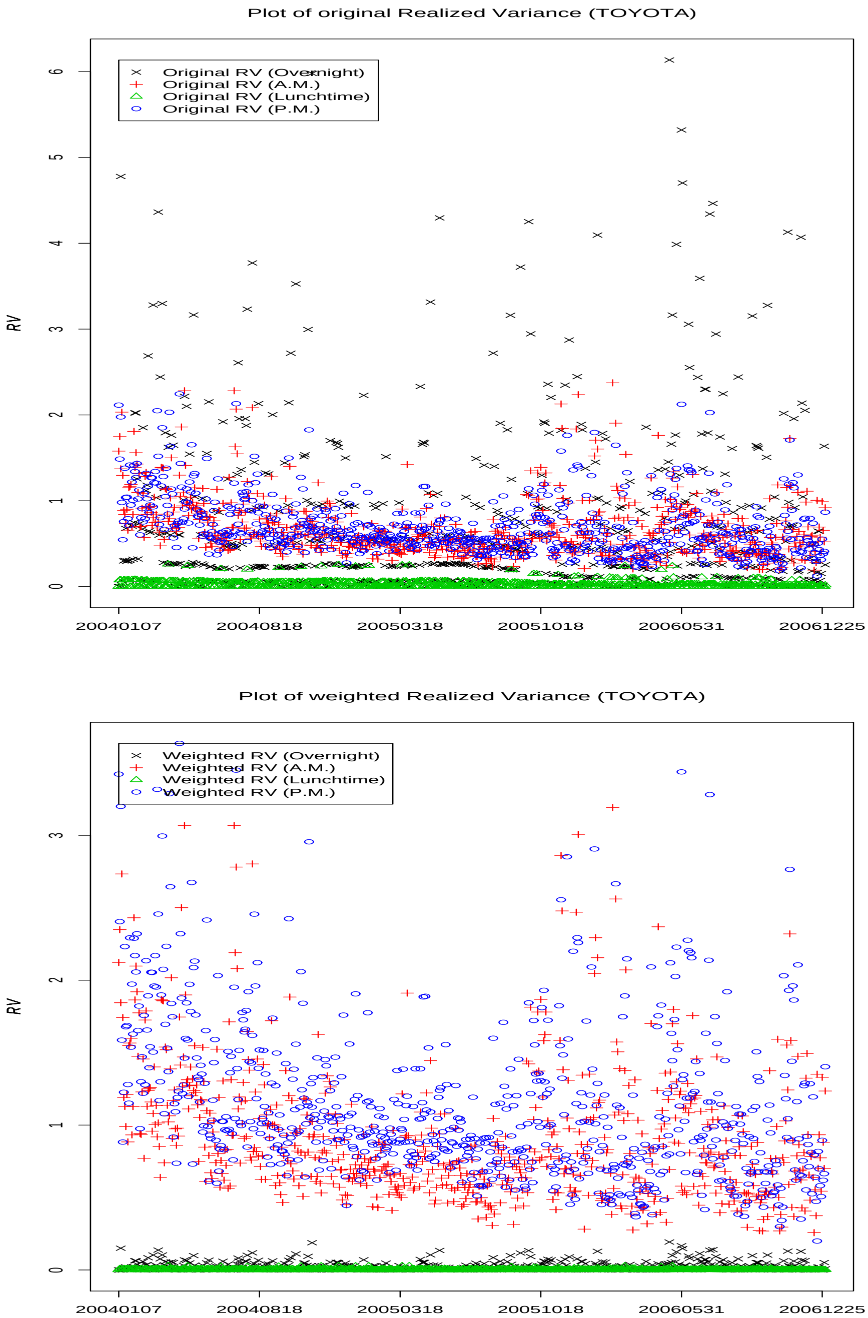

Figure 3 Original and weighted RV (TOYOTA)