UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FACULTY OF BUSINESS ADMINISTRATION AND ECONOMICS

Working Paper Series

Working Paper No. 2018-02

A comparison study of realized kernels using different sampling frequencies

Chen Zhou

April 2018

A comparison study of realized kernels using different sampling frequencies

Chen Zhou

University of Paderborn

Abstract

A crucial problem by applying realized kernels (RK) is the selection of the bandwidth. We improve the iterative plug-in (IPI) algorithms for selecting bandwidth of RK under independent (Feng and Zhou, 2015b) and dependent microstructure (MS) noise assumptions (Wang, 2014). The realized estimators calculated by both algorithms are consistent to the daily integrated volatility. The nice practical performance of these algorithms are illustrated by the application to 9 years of data on 10 European firms. Moreover, using these two algorithms we calculate RK based on different sampling frequencies and compare them with several other realized estimators. The comparison of these realized estimators is carried out by assessing their performances in the computation of Value-at-Risk (VaR) based on the Semi-FI-Log-ACD model. The RK estimators based on the tick-by-tick returns calculated by both IPI algorithms mentioned above have good performances and are hence recommended using as the estimators of volatility in practice.

Keywords: Realized volatility, realized kernels, microstructure noise, bandwidth selection, iterative plug-in, sampling frequency, Value-at-Risk

JEL Codes: C14, C51

1 Introduction

In recent years, there has been a large and rapidly expanding literature on estimation of the daily integrated volatility (IV). With the improvement of availability of high-frequency financial data Andersen et al. (2001a, b) proposed a model-free volatility approach, called realized volatility (RV_0), which exploits the information in high frequency returns and is constructed as the sums of intraday squared returns. However, observed prices are contaminated by market MS noise, which leads to the bias problem at high sampling frequencies (Hanse and Lunde, 2006a). To solve this bias problem, different approaches are introduced into the literature. Bandi and Rusell (2006, 2008, 2011) proposed a method of selecting the optimal sampling frequency based on a trade-off between the variance and bias. Zhang et al. (2005) proposed a subsampling method, called two time scales, to deal with the bias problem under i.i.d. MS noise. Their proposal was developed to the case under dependent MS noise by Zhang (2006) and Aït-Sahalia et al. (2008). This is also the first consistent estimator of IV. Furthermore, a pre-filter was used to weaken the effects of MS noise (e.g. Bollen and Inder, 2002). Large (2007) introduced an alternative estimator to control the MS effects. In addition, Zhou (1996) proposed to include the cross-products between two consequent observations. The author showed that his estimator, called RV_Z is unbiased under i.i.d. MS noise. This is also the first kernel method to deal with the problem of MS noise. Hansen and Lunde (2004, 2006b) studied extensively this estimator and proposed a simple kernel-based estimator. Barndorff-Nielsen et al. (2008, 2009, 2011) proposed realized kernels, which are a generalization of RV_Z and also are consistent estimators of the IV under given conditions.

A crucial problem by applying RK is the selection of the bandwidth. This problem is investigated first by Barndorf-Nielsen et al. (2009) for non-negative RK with an asymptotically optimal bandwidth of the order $O(n^{3/5})$ and an optimal rate of convergence of the order $O(n^{-1/5})$, where n is the number of observations on a trading day. Feng and Zhou (2015b) introduced a fully automatic iterative plug-in (IPI) algorithm to select bandwidth for RK by adapting the idea of Gasser et al. (1991). However, their algorithm is only applicable for the case under the assumption of i.i.d. MS noise. And the algorithm does not work, if three special cases occur. In this paper we improve this algorithm slightly so that it works for all cases. This improved algorithm is called the IN algorithm in the context of this paper. Ikeda (2015) proposed two-scale RK, which is a convex combination of two realized kernels with different bandwidths. He showed that his estimator converges to the IV in the presence of the dependent MS noise. Based on Ikeda (2015)'s conclusion, Wang (2014) proposed two IPI algorithms for RK under dependent noise assumption. In his work algorithm B is a fully automatic data-driven algorithm and after slight adjustment we call this algorithm in this paper the DN algorithm. End effects in the computation of the RK are considered. The non-flat-top Parzen kernel is utilized, because it can guarantee the nonnegativity of RK. Theoretically, in the fourth iteration the two bandwidths G and H have achieved the optimal rate of convergence. After several iterations, the resulted RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$. The processes of the IN and DN algorithms for RK based on different sampling frequency starting with different bandwidths are investigated. It is shown, that no matter the starting bandwidths, the selected bandwidths as of the second iteration are very close and that the final selected bandwidths are indeed the same. Most recently, Liu et al. (2015) has studied the accuracy of a wide variety of volatility measures constructed from high-frequency data. They concluded that when RV_0 calculated by 5-minute returns is taken as the benchmark measure, it is very hard to be beaten by any measure. Furthermore, if no benchmark is specified, the best estimators appear to e.g. RV_0 based on 1-minute data and RK. Barndorff-Nielsen et al. (2009) compared tick-by-tick and 1-minute RK with RV_0 computed by several sampling frequencies. Their conclusions were obtained by measuring the disagreement between these estimates based on transaction prices and mid-quote prices over 123 days. It was found that both RK estimates are better than any of the RV_0 and the statistical results of both RK estimates are similar. Motivated by all of these we study a comparison of RK using different sampling frequencies (tick-by-tick, 1-minute, 5-minute and 15-minute) calculated by the IN algorithm as well as RK using tick-by-tick returns calculated by the DN algorithm within a long period over 2000 trading days. Meanwhile, RV_0 computed from the different frequency returns are compared with these RK estimators mentioned above. In total, we have 4 sampling frequencies, 2 types of realized measures and two algorithms of a given transaction price series. The detailed comparison of these realized estimators are investigated by comparing their performances in the computation of Value-at-Risk based on the Semi-FI-Log-ACD model (also called the exponential SEMIFAR model, Feng and Zhou, 2015a). Value-at-Risk is considered as an extremely important measure in financial risk management to determine the amount of assets needed to cover possible losses. It is found that the performances of two RK estimators based on the tick-by-tick returns calculated by the IN and DN algorithms are better than any other realized estimator and hence recommended using in practice.

The rest of this paper is organized as follows. Different sampling schemes and several volatility estimators used in this paper are introduced in section 2. The improved bandwidth selectors under independent and dependent MS noise assumptions are proposed in section 3. The implementation of algorithms as well as the comparison of volatility estimators based on different frequencies of data are reported in section 4. Final remarks in section 5 conclude.

2 Realized measures

2.1 Different sampling schemes

Prices are practically observed at discrete and irregularly time intervals. Sampling schemes are rules of data recording. Different sampling schemes can be used for calculating realized measures. Let $p^*_{t,i}$, $i = 1, \ldots, n$, be the logarithmic efficient asset prices at time points $0 = \tau_0 < \tau_1 < \ldots \tau_n < \tau_{n+1} = T$ on trading day t, where n is the total number of observations at day t. Support that, $p^*_{t,i}$ follow a continuous time diffusion process $dp^*_t = \sigma_t dW_t$, where W is a standard Brownian motion and σ_t is the spot volatility process. The daily integrated volatility is given by

$$\mathrm{IV}_t = \int_0^T \sigma_t^2 dt. \tag{1}$$

We divide the interval $[0, T]$ in n subintervals. The length of the ith subinterval is defined by $\delta_{i,n} = \tau_i - \tau_{i-1}$. The integrated volatility for each of the subintervals is

$$\mathrm{IV}_{i,t} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt.$$

The choice of sampling schemes and sampling frequencies can have a strong influence on the estimation of the IV. Sampling schemes are mainly classified according to the concept of time.

Calendar Time Sampling (CTS). CTS is sampled by regularly spaced calendar time and is defined by $\delta_{i,n} = \frac{1}{n}$ for all i (see e.g. French et al. (1987); Hsieh (1991); Andersen and Bollerslev (1998); Andersen et al. (2001a)). Because of irregularity of the intraday data, calendar time sampled data must be builded artificially (refer to Wasserfallen and Zimmermann (1985); Andersen and Bollerslev (1997); and Dacorogna et al. (2001)). In this paper we use this sampling scheme. The original tick-by-tick price is sampled by different sampling frequencies, namely 1-minute, 5-minute or 15-minute. We utilize the data for the period between 9:00 am and 17:30 pm and use the previous tick method that takes the first observation as the sampled price (Hansen and Lunde, 2006a). For instance, when the price process is sampled at 1-minute interval, this will yield 511 price observations on one day.

Tick Time Sampling (TkTS) and Transaction Time Sampling (TrTS). Prices are recorded at every price change in TkTS (see e.g. Corsi et al., 2001; Zhou, 1996). Prices are sampled every kth transaction in TrTS (see e.g. Hansen and Lunde, 2006a). Grifin and Oomen (2008) investigated the difference between transaction time and tick time sampling. They found that the MSE of RV in tick time is lower than that in transaction time, especially when the level of noise, number of ticks, or the arrival frequency of efficient price moves is low.

Business Time Sampling (BTS). The sampling time are chosen such that $\mathrm{IV}_{i,t} = \frac{\mathrm{IV}_t}{n}$. The observation times for BTS are unobserved. The BTS transactions are often sampled in order to ensure approximately equal volatility of the returns over each interval. Peters and Vilder (2006), Andersen et al. (2007) and Andersen et al. (2010) selected time points to sample BTS returns with a target volatility.

The CTS, TrTS and TkTS schemes are constructed based on explicit criteria such as the regular calendar-time length or the number of ticks/price change. In addition, the observation times are observed. In contrast, the BTS scheme depends on the unobserved latent volatility. As a result, the most widely used sampling scheme is CTS. In many cases, the BTS scheme has the minimal MSE of realized variance among all sampling schemes, however, it is used less frequently in practice.

2.2 The effect of microstructure noise

Let $p_{t,i}$ be the ith logarithmic observed prices during day t. Assuming that the prices are observed with noise, we have

$$p_{t,i} = p^*_{t,i} + u_{t,i}, \tag{2}$$

where $u_{t,i}$ represents a stationary logarithmic microstructure noise process with mean zero. It is assumed that $u_{t,i}$ is independent of $p^*_{t,i}$. The noise process itself can be a white noise or a dependent process. The corresponding noise contaminated observed returns are obtained by

$$r_{t,i} = p_{t,i} - p_{t,i-1} = r^*_{t,i} + e_{t,i}, \tag{3}$$

where $e_{t,i} = u_{t,i} - u_{t,i-1}$ is the noise in the observed returns. In the presence of MS noise, realized variance $\mathrm{RV}_0 = \sum_{i=1}^{n} r_{t,i}$ is no more consistent of the integrated variance, and can be rewritten as

$$\mathrm{RV}_0 = \sum_{i=1}^{n} (r^*_{t,i})^2 + 2\sum_{i=1}^{n} r^*_{t,i} e_{t,i} + \sum_{i=1}^{n} e^2_{t,i}.$$

Under independent noise assumptions, the variance of noise is $var(u_{t,i}) = \omega^2$. The bias of RV_0 is $\mathrm{B}(\mathrm{RV}_0) = 2n\omega^2$ and the asymptotic variance of RV_0 is $var(\mathrm{RV}_0) \approx 4nE(u_i^4)$, as $n \to \infty$. For the dependent noise structure, Zhang (2006) and Aït-Sahalia et al. (2008) showed that the bias is still $2n\omega^2$ and the variance tents to $4n\Theta$, where $\Theta = V[(u_{t,1} - u^2_{t,0})] + 2\sum_{i=1}^{\infty} Cov[(u_{t,1} - u^2_{t,0}), (u_{t,i+1} - u^2_{t,i})]$. However, for large values of n, the bias and variance of RV_0 are infinite. Different approaches have been introduced into the literature to improve the performance of RV_0 in the presence of MS noise. Andersen et al. (2001a, 2003) proposed to select arbitrarily lower frequencies returns to treat the MS bias, such as every 5 or 15 minutes, instead of at every tick. RV_0 using different sampling frequencies is defined by

$$\mathrm{RV}^s_0 = \sum_{i=0}^{n_s} r^2_{t,i}, \tag{4}$$

where n_s denotes the number of observations. It is showed that the bias due to noise for independent and dependent is given by $2n_s E(u^2_{t,i})$. The bias is reduced when $n_s < n$, however, the variance is increased due to discretization. This leads to the well-known bias-variance trade-off.

Under the i.i.d. assumption on $u_{t,i}$, Zhou (1996) proposed to use the first order correlation to correct the bias in RV_0. This estimator RV_Z is unbiased but inconsistent. The version of $\mathrm{RV_Z}$ with different sampling is given by

$$\mathrm{RV}_{\mathrm{Z}}^{s} = \sum_{i=2}^{n_s-1} (r_{t,i}^2 + r_{t,i}r_{t,i+1} + r_{t,i}r_{t,i-1}). \tag{5}$$

Empirical evidence of MS noise can be found by displaying the ACFs of returns. Figure 1 to Figure 4 show the ACFs of different sampling-frequency returns on four selected trading days, respectively. Generally speaking, the ACFs on every selected trading day vary a lot by changing the sampling-frequency. ACFs of tick-by-tick returns for Thyssenkrupp on 07. May 2007 are displayed in Figure 1 (a). It shows that $\hat{\rho}_r(1) < 0$ and the ACFs at lags 2 and 3 are slightly not-zero. This indicates the existence of possible dependent MS noise. From Figure 1 (b) and (c) we can see that the correlations in 1-minute and 5- minute returns are reduced, but several lags are still outside two bounds. There are almost no correlations in 15-minute returns, which can be seen in Figure 1 (d). Figure 2 displays the ACFs of Siemens on 01. Dec. 2009. In Figures 2 (a), (b) and (c) not only $\hat{\rho}_r(1)$ but also the following several lags are slightly not-zero. The correlations are reduced with the decrease of sample frequencies and there are also almost no correlations in 15-minute returns. ACFs of Peugeot on 24. May 2011 are shown in Figure 3. In Figure 3 (a) $\hat{\rho}_r(1)$ is not significant, but some ACFs at higher lags are clearly significant. This indicates the presence of dependent MS noise and that the simple independence assumption on the noise may not be sufficient. Using the DN algorithm to calculate RK may be suitable in such a case. The correlations are clearly reduced by using 1-minute returns in Figure 3 (b). As has been demonstrated before, there are also almost no correlations in 5- and 15-minutes returns. Figure 4 shows the ACFs for Schneider Electric on 24. Jan. 2014. In Figure 4 (a) the noise is very strong with $\hat{\rho}_r(1) < -0.5$ and the several lags, which follow, are also clearly significant. On this day $\mathrm{RV_Z} < 0$, which is one of the special cases for the IN algorithm. The correlations diminish clearly by using 1-minute returns in Figure 4 (b). There are almost no correlations in 5-minute and 15-minute returns. In conclusion, from Figure 1 to Figure 4 it can be seen that in most cases $\rho_r(1)$ is significantly negative and there are possible dependent MS noise in tick-by-tick returns and for some cases there may be still MS noise in 1-minute and 5-minute returns. Furthermore, the correlations reduce with the decrease of sample frequencies. The reason is that the diminished number of observations reduces the bias and extends the two ACF bounds, which makes the daily returns uncorrelated. However, decreasing the sampling frequency toward to increase the variance and reduce the accuracy of volatility estimators. Hence, a realized estimator with an optimal frequencies is necessary to investigate, which can remove the effect of MS noise and grantee the accuracy of the estimation at the same time. The corresponding numerical results of RV_0 and RK computed from different sample frequencies for these four selected days are provided in Table 1. From the first four columns we see that the values of RV_0 for all examples reduce by the decease of the sample frequencies. For instance, for SE on 24. Jan. 2014 the tick-by-tick RV_0 is even three times bigger than the 1-minute RV_0. The last four columns show us that the results of RK computed from different frequencies returns by the IN algorithm are clearly different form the results of tick-by-tick RK using the DN algorithm. These differences are caused by the calculated dependent MS noise with the DN algorithm. One of our purpose in this paper is to investigate which of these realized estimators can estimate the financial market volatility more accurately.

3 Bandwidth selection for realized kernels

3.1 Realized kernels

Realized kernels are introduced by Barndorf-Nielsen et al. (2008, 2009, 2011) and can be thought of as an extension of RV_Z.

$$\mathrm{RK} = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \quad \gamma_h = \sum_{i=|h|+1}^{n} r_i r_{i-|h|}, \tag{6}$$

where $k(x)$ is a kernel weight function, H is the selected bandwidth and γ_h is the h-th realized autocovariance. In this paper we use the Parzen kernel, which can guarantee the nonnegative realized kernels estimate.

Under $H = O(n^\alpha)$ with $1/2 < \alpha < 1$ and further regularity conditions, RK is a consistent estimator of IV with the AMSE (asymptotic mean squared error):

$$\mathrm{AMSE}(H) = [K''(0)]^2\Omega^2 n^2 H^{-4} + 4K_\bullet^{0,0}\,(T\mathrm{IQ})\,Hn^{-1}, \tag{7}$$

where $K_\bullet^{0.0} = \int_0^\infty k(x)^2 dx$ is a constant, IQ$= \int_0^T \sigma_t^4 dt$ is called the integrated quarticity, $\Omega = \Sigma_{h>0}\Omega(h)$ is the long-run variance of $u_{t,i}$ and $\Omega(h) = \Sigma_{i=h+1}^{n} u_{t,i}u_{t,i-h}$. The first part and second part in Eq.(7) are the asymptotic squared bias and variance of RK, respectively.

The asymptotically optimal bandwidth which minimizes the AMSE is provided by

$$H = c_0 \xi^{4/5} n^{3/5} \quad \text{with} \quad c_0 = \left\{ \frac{k''(0)^2}{k_{\cdot}^{0.0}} \right\}^{1/5} \quad \text{and} \quad \xi^2 = \frac{\Omega}{\sqrt{T \mathrm{IQ}}}, \tag{8}$$

where $c_0 = 3.5134$ for the Parzen kernel. The bandwidth H depends on the unknown quantities Ω and IQ.

3.2 Bandwidth selection under i.i.d. noise

The crucial problem when applying RK is the selection of the bandwidth. The asymptotically optimal bandwidth given in Eq. (8) provides a base for selecting the bandwidth. To do this, we need to estimate Ω and IQ. When $u_{t,i}$ is under the i.i.d. noise assumption, Ω reduces to ω^2. Following Barndorf-Nielsen et al. (2009), IQ in Eq. (8) is replaced with IV^2, because the former is not too far from the latter, under conditions, when σ_t^2 does not vary significantly. It can be easy shown that the conclusions for RK in Barndorf-Nielsen et al. (2008, 2009, 2011) are also suitable for RK based on different sampling frequencies (RK_s). Under the i.i.d assumption a biased version of H, called H^{B} constructed at different sampling frequencies is given by

$$H_s^{\mathrm{B}} = c_0 \xi_{\mathrm{B}}^{4/5} n_s^{3/5} \quad \text{with} \qquad \xi_{\mathrm{B}}^2 = \frac{\omega_s^2}{\mathrm{IV_s}}. \tag{9}$$

Now, assuming that $\widehat{\mathrm{IV}}_s$ is at least an unbiased estimator of IV_s, it can be shown that

$$\hat{\omega}_s^2 = \frac{\mathrm{RV}_0^{\mathrm{s}} - \widehat{\mathrm{IV}}_{\mathrm{s}}}{2n_s} \tag{10}$$

is a consistent estimator of ω_s^2. Barndorff-Nielsen et al. (2009) proposed to select the bandwidth by plugging two complex estimates of IV and ω^2. Their proposal is not fully data-driven and the selected bandwidth does not converge to H^{B}. Instead of these two complex estimators, Feng and Zhou (2015b) utilized $\mathrm{RV_Z}$ as an initial value of IV for estimating the starting bandwidth H_1. Then inserting $\hat{H}_1$ into Eqs. (9) and (10), and using the IPI idea to obtain optimal bandwidth $\hat{H}^{\mathrm{B}}$ automatically. This is the first IPI algorithm and it is shown the selected bandwidth is consistent to H^{B}. However, their algorithm does not work, if three following special cases occur. They are $\mathrm{RV_Z}$ is smaller than zero (Case 1), $\mathrm{RV_Z}$ is larger than $\mathrm{RV_0}$ (Case 2), and RK is larger than $\mathrm{RV_0}$ (Case 3). For Case 1 a large starting bandwidth should be manually set, since $\widehat{\mathrm{IV}}_1 < 0$. For Case 2 they set $\hat{H}^{\mathrm{B}} = 0$ and simply use $\mathrm{RV_0}$, because $\hat{\omega}_1^2 < 0$. And for Case 3 they put $\hat{H}^{\mathrm{B}} = \hat{H}_j$ as the selected optimal bandwidth, because $\hat{\omega}_j^2 < 0$. The last two cases imply that there is possibly no strong MS noise on those days. To avoid these three cases, in this paper we let $\hat{\omega}_{s,1}^2 = \frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}$ replace $\hat{\omega}_{s,1}^2 = \frac{\mathrm{RV}_0^s - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}}{2n_s}$. It aims to avoid $\mathrm{RV}_0^s - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}} < 0$. It can be observed that $\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}$ is not far from $\frac{\mathrm{RV}_0^s - \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}}{2n_s}$, if $n \to \infty$. Moreover, end effects in the computation of the RK is also considered with $m = 2$. For more details please refer to section 2.2 in Barndorff-Nielsen et al. (2009).

Let j denote the number of iterations. The IN algorithm unfolds as:

Step 1. To resolve end-effects problem. Let $p_0 = \frac{1}{m}(p_{t,1} + \dots p_{t,m})$, $p_j = p_{t,j+1}$, $j = m \dots, n - m - 1$ and $p_n = \frac{1}{m}(p_{t,n-m+1} + \dots p_{t,n})$, where we put $m = 2$.

Step 2. In the first iteration $j = 1$, let $\widehat{\mathrm{IV}}_{s,1} = \mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}$, calculate $\hat{\xi}_{s,1}^4$ with

$$\hat{\xi}_{s,1}^4 = \left(\frac{\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}}{\mathrm{RV}_{\mathrm{Z}}^{\mathrm{s}}} \right)^2. \tag{11}$$

Insert $\hat{\xi}_{s,1}^4$ into Eq. (9) to obtain $\hat{H}_{s,1}$. Let $j = 2$.

Step 3. In the jth iteration with $j > 1$, calculate $\widehat{\mathrm{IV}}_{s,j}$ with $\hat{H}_{s,j-1}$. Then insert

$$\hat{\xi}_{s,j}^4 = \left(\frac{\frac{\mathrm{RV}_0^{\mathrm{s}}}{2n_s}}{\widehat{\mathrm{IV}}_{s,j}} \right)^2 \tag{12}$$

into Eq. (9) to obtain $\hat{H}_{s,j}$.

Step 4. Iteratively carry out this procedure until convergence has been achieved.

The algorithm will be stopped, if $\hat{H}_{s,j}$ is equal to $\hat{H}_{s,j-1}$. As showed in Theorem 1 in Feng and Zhou (2015b), theoretically, both of $\widehat{\mathrm{RK}}_s$ and $\hat{H}_s^{\mathrm{B}}$ become consistent form in the third iteration, while their rate of convergence can still be improved in the fourth iteration. Thereafter, $\widehat{\mathrm{RK}}_s$ achieves its optimal rate of convergence of the order $O(n^{-1/5})$ and this rate of convergence is also shared by $(\hat{H}_s^{\mathrm{B}} - H_s^{\mathrm{B}})/H_s^{\mathrm{B}}$. An R code is developed for practical implementation of the proposed bandwidth selector.

3.3 Bandwidth selection under dependent noise

From Eq. (7) it can be seen that for dependent noise the estimation of the long-run variance Ω may enable the correction of the leading bias of RK. Following Ikeda (2015) we utilize $MK(G)$ to estimate Ω. Define

$$MK(G) = (|k''(0)|nG^{-2})^{-1}\mathrm{RK}(G), \tag{13}$$

where RK(G) are the realized kernels for bandwidth G. Under the assumption G=$O(n^{\beta})$, $\beta \in (0, 1/2]$ and if $n \to \infty$

$$MK(G) \longrightarrow \Omega + \lim_{n\to\infty} n^{-1}G^2(|k''(0)|)^{-1}\mathrm{IV}_t. \tag{14}$$

In addition, given G=$O(n^{\beta})$ for $1/(2q+1) \leq \beta < 1/3$ or $\beta = 1/3$, where q is the characteristic exponent of $k''(0)$. If $q = 1$ and $\beta = 1/3$ (for Parzen kernel) are allowed for the asymptotic normality of $MK(G)$. According to the related results in Ikeda (2015), the AMSE$_G$ for the Parzen kernel is obtained by

$$\mathrm{AMSE}_G = \left(\frac{\mathrm{IV}_t}{k''(0)}\right)^2 \frac{G^4}{n^2} + O\left(\frac{G}{n}\right) + O\left(\frac{1}{G^2}\right). \tag{15}$$

The asymptotically optimal bandwidth $G = O(n^{1/3})$ for MK(G) can be obtained by minimizing the AMSE$_G$. Two-scale realized kernels can be used to estimate Ω. However, the estimator can be negative. Wang (2014) proposed two IPI algorithms A and B under dependent noise assumption. Algorithm A starts with $H_1 = n^{\alpha}$, $\alpha \in (1/2, 1)$. In the second iteration $\hat{H}_2$ has already achieved the rate of convergence of the order $O(n^{3/5})$ and after only a few iterations $\hat{H}_j$ converges to H^{B}. He also demonstrated that $\hat{H}_j$ is consistent even when α is outside $(1/2, 1)$. Algorithm B is a fully automatic data-driven algorithm and with slight adjustments it is called the DN algorithm in this paper. Let j denote the number of iterations. The DN algorithm unfolds as:

Step 1. To resolve end-effects problem. Let $p_0 = \frac{1}{m}(p_{t,1} + \dots p_{t,m})$, $p_j = p_{t,j+1}$, $j = m \dots, n-m-1$ and $p_n = \frac{1}{m}(p_{t,n-m+1} + \dots p_{t,n})$, where we put $m = 2$.

Step 2. In the first iteration, let $\widehat{\mathrm{IV}}_1 = \mathrm{RV_Z}$, calculate $\hat{\xi}_1^4$ with

$$\hat{\xi}_1^4 = \left(\frac{\frac{\mathrm{RV}_0}{2n}}{\mathrm{RV_Z}}\right)^2. \tag{16}$$

Insert $\hat{\xi}_1^4$ into Eq. (9) to obtain $\hat{H}_1$. Let $j = 2$.

Step 3. In the jth iteration with $j > 1$, calculate $\widehat{\mathrm{IV}}_j$ with $\hat{H}_{j-1}$. Meanwhile, similar to step 2, let $G = H_{j-1}^{5/9}$ to obtain $\widehat{\mathrm{RK}}(G)$. The unknown value $\hat{\xi_j^4}$ is calculated by

$$\hat{\xi_j^4} = \left(12^{-1} n^{-1} G_j^2 \frac{\mathrm{RK(G)}}{\mathrm{IV}_j}\right)^2 \tag{17}$$

and then $\hat{H}_j$ can be obtained.

Step 4. Iteratively carry out this procedure until convergence has been achieved.

Like in the IN algorithm IV^2 is also used to estimate IQ. Let $\Omega_1 = \frac{\mathrm{RV}_0}{2n}$, so that a negative Ω_1 is avoided. Please note that for both IN and DN algorithms $\hat{H}_j$ is obtained by truncating the integer part of selected optimal bandwidth and subsequently adding 1. The proof for the convergence of this algorithm was given in Wang (2014). Theoretically, in the fourth iteration G and H have achieved to their optimal rate of convergence, respectively. After several iterations the resulted RK achieves its optimal rate of convergence of the order $O(n^{-1/5})$.

4 Application

The datasets of ten European companies (Air France (AF), Allianz (ALV), BMW, German Bank (DBK), Michelin (MIC), Peugeot (PSA), RWE, Schneider Electric (SE), Siemens (SIE) and Thyssenkrupp (THK)) from 02. Jan. 2006 to 30. Sept. 2014 are used as data examples, which were downloaded from "Thomson Reuters" Corporation. The total number of trading days for the six German companies is 2226 and for the four French companies it is 2239. The trading times for all ten companies are from 9:00 to 17:30. We eliminate the data outside this time interval and also the data containing clerical errors. In addition to tick-by-tick data, we also employ 1-minute, 5-minute and 15-minute data, which are computed based on tick-by-tick data. Two realized measures, RV_0 and RK are considered. The IN algorithm is utilized to calculate RK obtained based on tick-by-tick, 1-minute and 5-minute returns. The DN algorithm is used to calculate tick-by-tick RK. We consider end effects in the computation of the RK with $m = 2$. The absolute difference for the ten European companies are less than 0.878% on average, which confirms the conclusion in Barndorf-Nielsen et al. (2009) that the end effect can be ignored in practice.

4.1 Implementation of algorithms

Figure 5 shows the processes of the IN and DN algorithms starting with different bandwidths for four selected days. In Figure 5, "u" and "l" represent the selected bandwidths in each iteration when a fixed upper and a lower starting bandwidths are used, respectively. And "z" stands for selected bandwidths in each iteration, when algorithms start with $\hat{H}_1$ estimated by Eq. (9). Figure 5 (a) displays the process of the IN algorithm based on tick-by-tick returns for THK on 07. May 2007. We choose 50 as the upper starting bandwidth and the selected bandwidths $\hat{H}_{s,j}$ in 3 iterations are 50, 14, 14, respectively. When 5 is chosen as the lower starting bandwidth, the selected bandwidths in 4 iterations are 5, 12, 14, 14, respectively. The calculated starting bandwidth $\hat{H}_1$ is 16 by using $\mathrm{RV}_{\mathrm{Z}}^{s}$ as the estimator of $\mathrm{IV}_{s,1}$, after 3 iterations ($\hat{H}_{s,j}$=16, 14, 14), the optimal bandwidth is also 14. Likewise, the processes of the IN algorithm based on 1-minute and 5-minute returns for MIC and RWE are plotted in Figure 5 (b) and (c), respectively. Figure 5 (d) plot the process of the DN algorithm based on tick-by-tick returns for SIE on 01. Dec. 2009. In Figure 5 (d) the upper starting bandwidth is 100, the selected bandwidths in 5 iterations are 100, 39, 46, 51, 51, respectively. We choose 5 as the lower starting bandwidth, the selected bandwidths in 6 iterations are 5, 32, 41, 46, 51, 51, respectively. The calculated starting bandwidth is 19 by using $\mathrm{RV}_{\mathrm{Z}}^{s}$ as the estimator of $\mathrm{IV}_{s,1}$, after 5 iterations ($\hat{H}_{s,j}$=19, 35, 45, 51, 51), $\hat{H}^{\mathrm{B}}$ is also equal to 51. From Figure 5 we can see that both algorithms work very well and only a few iterations are required. In addition, no matter what the starting bandwidths are, the selected bandwidths as of the second iteration are very close and that the final selected bandwidths are indeed the same.

Figure 6 displays the histograms of selected bandwidths and iteration numbers of RK for DBK. The histograms of $\hat{H}^{\mathrm{B}}$ calculated by the IN algorithm for tick-by-tick, 1-minute and 5-minute RK are plotted in Figures 6 (a), (b) and (c), respectively. Figure 6 (d) provides the histogram of $\hat{H}^{\mathrm{B}}$ for the tick-by-tick RK calculated by the DN algorithm (RK-DN-tick). The corresponding histograms of iteration numbers are displayed in Figures 6 (e), (f), (g) and (h), respectively. The selected bandwidths in Figure 6 (d) are bigger than those in Figure 6 (a), the reason is that the dependent MS noise is taken into account. The selected bandwidths for 1-minute RK in Figure 6 (b) are smaller than those for tick-by-tick RK, but larger than those for 5-minute RK. This corresponds to their ACFs results, that the correlations in 5-minute returns are smaller than the correlations in 1-

minute and tick-by tick returns. In R code a maximal number of iterations $J = 20$. For the datasets under consideration this limit is never achieved. The commonly required number of iterations for DBK-IN-tick, DBK-IN-1min, DBK-IN-5min and DBK-DN-tick are 3, 3, 3, 5, respectively, which are also true for the other nine companies. These confirm that both IN and DN algorithms work very well in practice and only a few iterations are required. The illustrative results for MIC can be found in Figure 7. The histograms for the other eight companies show the same conclusions as in Figures 6 and 7, and are hence omitted to save space. Please note that the IN algorithm does not work for 15-minute RK of all the ten stocks, because on some days the number of observations is smaller than the selected bandwidths ($n < \hat{H}^{\mathrm{B}}$). For the same reason, 5-minute is not recommended using for calculating RK.

Table 2 lists the number of the three special cases ($\mathrm{RV_Z} < 0$, $\mathrm{RV_Z} > \mathrm{RV_0}$ and $\mathrm{RK} > \mathrm{RV_0}$) mentioned in Feng and Zhou (2015b) for RK-IN-tick, RK-IN-1min, RK-IN-5min and RK-DN-tick of the ten companies, respectively. We see that for all companies Case 1 occur rarely, however, Case 2 and Case 3 happen with a very big chance. For instance, Case 2 occur on 116 days for MIC-IN-tick and Case 3 occur on 226 days for SE-DN-tick. This shows us the necessity of adjusting Feng and Zhou (2015b)'s algorithm by replacing $\hat{\omega}^2_{s,1} = \frac{\mathrm{RV}^s_0 - \mathrm{RV}^s_{\mathrm{Z}}}{2n_s}$ with $\hat{\omega}^2_{s,1} = \frac{\mathrm{RV}^s_0}{2n_s}$. After this adjustment, the proposed IN and DN algorithms work automatically very well for all days. Meanwhile, we find that for all the companies Case 3 occur much more often for DN-tick than that for IN-tick. For instance, Case 3 in THK occur on 100 days for DN-tick and only on 26 days for IN-tick. This difference shows that the obtained RK results change obviously if the dependent noise takes into account. In addition, we see that Case 2 and Case 3 occur much more often when the sampling frequencies decline.

Table 3 lists the mean and standard deviations of $\mathrm{RV_0}$ and RK obtained based on tick-by-tick, 1-minute, 5-minute and 15-minute returns for the ten companies over the whole period, respectively. The mean values and standard deviations for tick-by-tick $\mathrm{RV_0}$ are larger than any other realized estimators in Table 3. The mean values of $\mathrm{RV_0}$ diminish with the decline of sampling frequency. The standard deviations of 1-minute, 5-minute and 15-minute $\mathrm{RV_0}$ for all companies are much smaller than those of tick-by-tick $\mathrm{RV_0}$. As mentioned before, the algorithm IN does not work for 15-minute RK of all companies, because on some days $n < \hat{H}^{\mathrm{B}}$. Therefore, the results are not listed in Table 3. For the same reason, 5-minute returns are not recommended using to calculate RK. In most cases, the standard deviations of RK-DN-tick are smaller than those of RK-IN-tick and of RK-IN-1min. The mean values of RK-DN-tick lie between the mean values of RK-IN-tick and RK-IN-1min. The differences between the means of RK-DN-tick and RK-IN-tick indicate the bias cased by estimated dependent MS noise.

4.2 Comparison of realized estimators based on Value-at-Risk

Further comparison of these realized estimators is carried out by assessing their performances in the computation of Value-at-Risk (VaR). Financial risk managers often report the risk of investments using the concept of VaR, which estimates the maximum loss at given confidence interval in a certain period. VaR is widely used by investors and regulators in the financial industry to measure the amount of assets needed to cover possible losses and has been considered as a expectable banking risk measure. The Basel Committee demands banks use the VaR in establishing the minimum capital necessary at investments in order to reduce the fragility of international active banks. After its proposal it becomes a popular risk assessment tool in financial service firms. In the literature, Beltratti and Morana (2005) as well as Giot and Laurent (2004) investigated the performances of VaR measures based on the GARCH and lnRV-FARIMA models. Their results show that the lnRV-FARIMA class provide a superior performance by computing VaR. Feng and Zhou (2015a) showed the logarithmic RK can be well described by a long-memory process and may also exhibit a deterministic nonparametric trend. Hence, in our framework we compute the one step ahead VaR by means of the semiparametric FARIMA model (SEMIFAR also called the Semi-FI-Log-ACD model in Feng and Zhou, 2015a) based on the different logarithmic realized estimators. The VaR calculated based on the the SEMIFAR model at a given confidence interval $(1-\alpha)$ is obtained by

$$\mathrm{VaR}^t_{1-\alpha} = s(\tau_t)\sigma_t Z_{1-\alpha}, \tag{18}$$

where $s(\tau_t)$ is the local variance, σ_t is the standard deviation (volatility) of investments at time t and $Z_{1-\alpha}$ is loss distribution for the corresponding N(0,1) quantile. In the first step we use the intraday returns of different sampling frequencies to compute the realized estimators mentioned in section 4.1. In the second step we estimate the conditional mean using the SEMIFAR model based on the logarithmic realized estimators. The estimated total means in the original data $s(\tau_t)\sigma_t$ is obtained through the exponential transformation of these estimated deterministic trend and the estimated conditional mean.

A wealth of literature focuses on measuring the quality of the VaR calculations. Backtesting is a statistical procedure, where the loss forecast calculated by VaR is compared with the actual losses. Kupiec (1995) proposed a basic tests to examine the frequency of losses in excess of VaR. By means of a simple backtesting Peitz (2016) compared VaR calculations based on the parametric with semiparametric models. In his proposal the observed amount of exceptions (Points over VaR) is compared to the expected amount/benchmark $n * \alpha$. We utilize the simple backtesting in Peitz's work. Let $\alpha = 5\%$, n is 2226 for the six German companies and 2239 for the four French companies. The corresponding benchmarks for all ten companies are 111. Figure 8 to Figure 17 show the one day dynamic ahead 95% VaR based on the lnRV-SEMIFAR model for the ten example companies from 02. Jan. 2006 to 30. Sept. 2014. The loss distribution is assumed normal. In each figure the black line shows the loss/negative returns and the red line depicts the 95% VaR values by means of the SEMIFAR model based on the different logarithmic realized estimators. Table 4 lists the corresponding numerical results of points over VaR and the deviations from the benchmark for Figure 8 to Figure 17. Table 5 summarizes the total absolute deviation and total deviation from the benchmark of 95% VaR based on the lnRV-SEMIFAR model. In the empirical implementation, RK-IN-tick with the smallest total absolute deviation 121 is the best estimator in the computation of VaR. RK-DN-tick is the second best estimator with the total absolute deviation 145. This confirms that the independent MS noise assumption is more suitable on the most trading days in our data examples. From Figure 1 to 4 we can also see that in most cases the ACFs at lag 1 are clearly significant and the results imply that some at higher lags slightly significant ACFs occur only occasionally. Meanwhile, the difference between the total absolute deviation of RK-IN-tick and that of RK-DN-tick is not big. Both can be considered as the good volatility estimators. In addition, the worst realized estimator is RV_0-tick. Its deviations for all ten companies are negative, the corresponding total deviation is -568. From the figures (a) of all ten companies we can see that the red lines are far over the daily losses, which confirms the numerical result that the 95% VaR based on the ln(RV_0-tick)-SEMIFAR seriously overestimate the financial risk. Next overestimated realized estimator is RV_0-1min with the total deviation -167. The deviations of RK-IN-5min for all ten companies are however positive, which means the underestimation of the financial risk. This fact can be confirmed in the figures (g) of all ten companies, where a few black lines are outside the red lines. The same results are also obtained for RK-IN-1min, RV_0-5min, RV_0-15min, based on them the financial risk can be seriously underestimated. Moreover, the performance of RV_0-15min is worse than that of RV_0-1min, while the performance of RK-IN-5min is worse than that of RK-IN-1min and RK-IN-tick. This shows again the necessity of choosing a realized estimator together with a suitable frequency.

In summary, after comparing the performances of different realized estimators in the computation of VaR based on the SEMIFAR model we find that the performances of RK-IN-tick and RK-DN-tick are better than any other realized estimator and are hence recommended using as the estimators of IV in practice.

5 Final remarks

In this paper we improved the IPI algorithms for RK under independent and dependent noise assumptions. End effects are considered for both algorithms. The data from ten European companies of 9 years are used as data examples. We apply the IN algorithm to tick-by-tick, 1-minute, 5-minute and 15-minute returns and apply the DN algorithm to tick-by-tick returns. Both algorithms are fully automatic and now work very well for all data examples. Furthermore, we compare these RK estimators with RV_0 calculated by tick-by-tick, 1-minute, 5-minute and 15-minute returns. The IN algorithm does not work well for the data-frequencies over 5-minute, because on same days the number of observations is smaller than the selected bandwidths. Further comparison of these realized measures is carried out by assessing their performances in the computation of VaR based on the SEMIFAR model. It is found that RK-IN-tick and RK-DN-tick estimators have good performances and are hence recommended using as the estimators of IV in practice. However, a more direct method to rate these realized estimators should be proposed by comparing their MSE values. This is also one of our future research directions.

Acknowledgments: I would like to express my deepest gratitude to Prof. Dr. Yuanhua Feng, who gives me some useful suggestions to improve the quality of this paper.

Reference

Aït-Sahalia, Y. and Mancini, L. (2008). Out of sample forecasts of quadratic variation. *Journal of Econometrics*, 147(1), 17-33.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4, 115-158.

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39 (4), 885 -905.

Andersen, T. G.; Bollerslev, T.; Diebold, F. X. and Ebens, H. (2001a). The distribution of stock return volatility. *Journal of Financial Economics*, 61, 43-76.

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

Andersen, T. G.; Bollerslev, T.; Diebold, F. X. and Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica*, 71, 579-625.

Andersen, T. G.; Bollerslev, T. and Dobrev, D. (2007). No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics*, 138(1), 125-180.

Andersen, T. G.; Bollerslev, T.; Frederiksen, P. and ørregaard Nielsen, M. (2010). Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25(2), 233-261.

Bandi, F. M. and Russell, J. R. (2006). Separating market microstructure noise from volatility. *Journal of Financial Economics*, 79, 655-692.

Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized volatility, and optimal sampling. *Review of Economic Studies*, 75, 655-692.

Bandi, F. M. and Russell, J. R. (2011). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics*, 160, 145-159.

Barndorff-Nielsen, O. E.; Hansen, P.; Lunde, A. and Shephard, N. (2008). Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. *Econometrica*, 76, 1481-1536.

Barndorff-Nielsen, O. E.; Hansen, P.; Lunde, A. and Shephard, N. (2009). Realised kernels in practice: Trades and Quotes. *Econometrics Journal*, 12, C1-C32.

Barndorff-Nielsen, O. E.; Hansen, P.; Lunde, A. and Shephard, N. (2011). Multivariate realized kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and nonsynchronous trading. *Journal of Econometrics*, 162, 149-169.

Beltratti, A. and Morana, C. (2005). Statistical benefits of Value-at-risk with long memory. *Journal of Risk*, 7, 21-45.

Bollen, B. and Inder, B. (2002). Estimating daily volatility in financial markets utilizing intraday data. *Journal of Empirical Finance*, 9, 551-562.

Corsi, F.; Zumbach, G.; Müler, U. A. and Dacorogna, M. M. (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes*, 30(2), 183-204.

Dacorogna, M.; Gencay, R.; Mller, U.; Olsen, R. B. and Pictet, O. V. (2001). An introduction to high-frequency finance. Academic Press, London.

Feng, Y. and Zhou, C. (2015a). Forecasting financial market activity using a semiparametric fractionally integrated Log-ACD. *International Journal of Forecasting*, 31, 349-363.

Feng, Y. and Zhou, C. (2015b). An iterative plug-in algorithm for realized kernels. Preprint, University of Paderborn.

French, K. R.; Schwert, G. W. and Stambaugh, R. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19, 3-29.

Gasser, T.; Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.*, 86, 643-652.

Giot, P. and Laurent, S.(2004). Modeling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, 11(3), 379-398.

Griffin, J. E. and Oomen, R. C. A.(2008). Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Reviews*, 27(3), 230-253.

Hansen, P. R. and Lunde, A. (2004). An unbiased measure of realized variance. Unpublished manuscript. Stanford University.

Hansen, P. R. and Lunde, A. (2006a). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, 24, 127-218.

Hansen, P. R. and Lunde, A. (2006b). Consistent ranking of volatility models. *Journal of Econometrics*, 131, 97-121.

Hsieh , D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance*, 46(5), 1839-1877.

Ikeda, S. (2015). Two scale realized kernels: a univariate case. Boston University. *Journal of Financial Econometrics*, 13(1), 126-165.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3, 73-84.

Large, J. (2007). Estimating quadratic variation when quoted prices jump by a constant increment. *SSRN Electronic Journal*, 160(1), 2-11.

Liu, L.; Patton, A. and Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187, 293-311.

Peitz, C. (2016). Die parametrische und semiparametrische Analyse von Finanzzeitreihen-Neue Methoden, Modelle und Anwendungsmöglichkeiten. Springer Gabler.

Peters, R. T. and Vilder, R. G. (2006). Testing the continuous semimartingale hypothesis for the S&P 500. *Journal of Business & Economic Statistics*, 24, 444-454.

Wang, J. (2014). Research on ACD model and its application. *Doctorial Dissertation*. Zhongnan University of Economics and Law.

Wasserfallen, W. and Zimmermann, H. (1985). The behavior of intraday exchange rates. *Journal of Banking and Finance*, 9, 55-72.

Zhang, L.; Mykland, P. A. and AÏt-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, 100, 1394-1411.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12, 1019-1043.

Zhou, B. (1996). High-Frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, 14, 45-52.

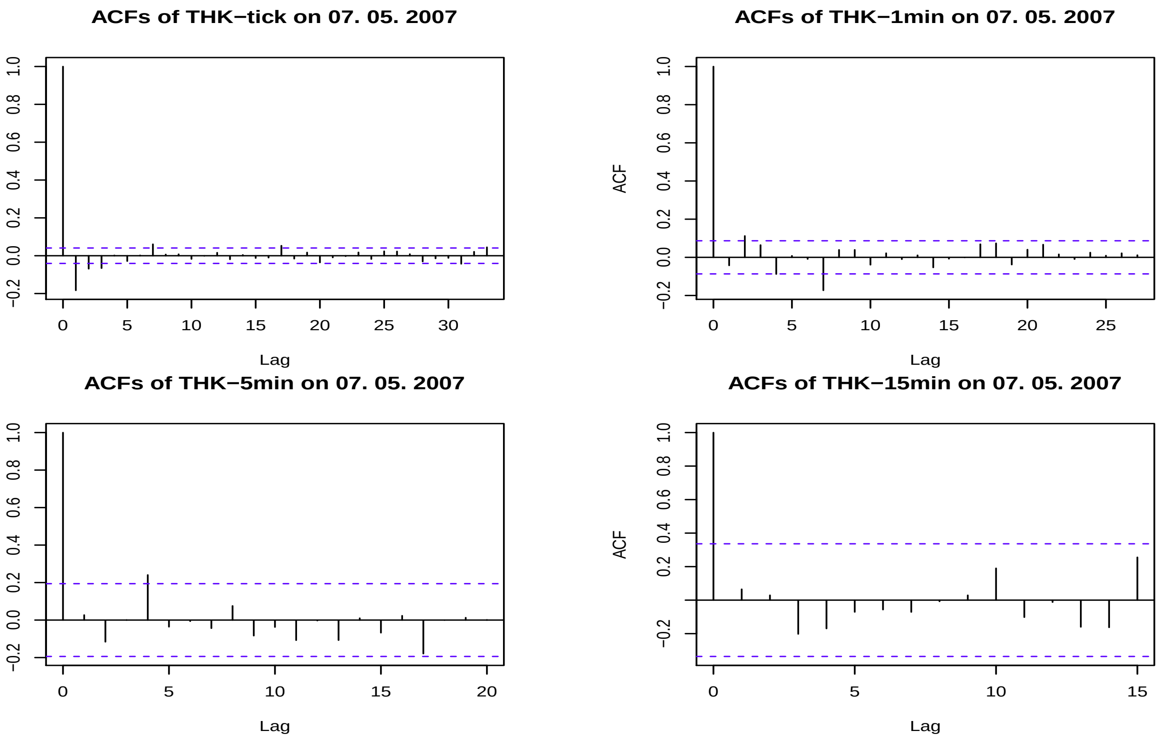

Figure 1: ACFs of Thyssenkrupp on 20. Jan. 2012.

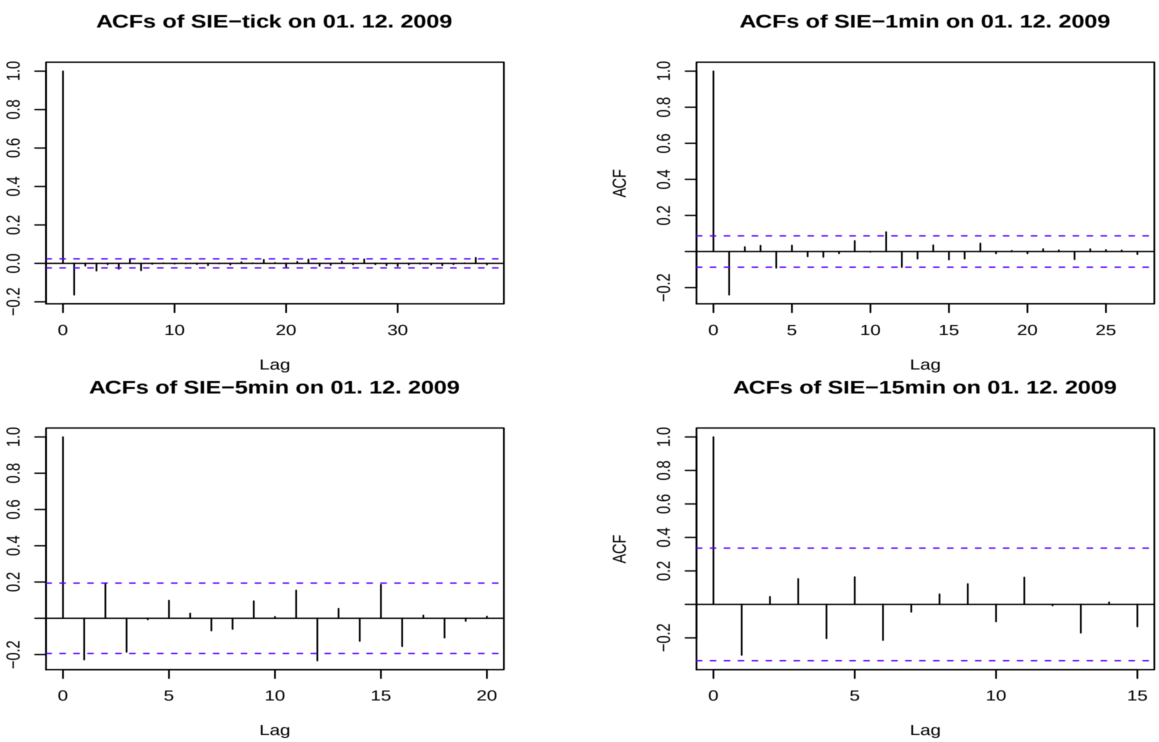

Figure 2: ACFs of Siemens on 13. Jun. 2008.

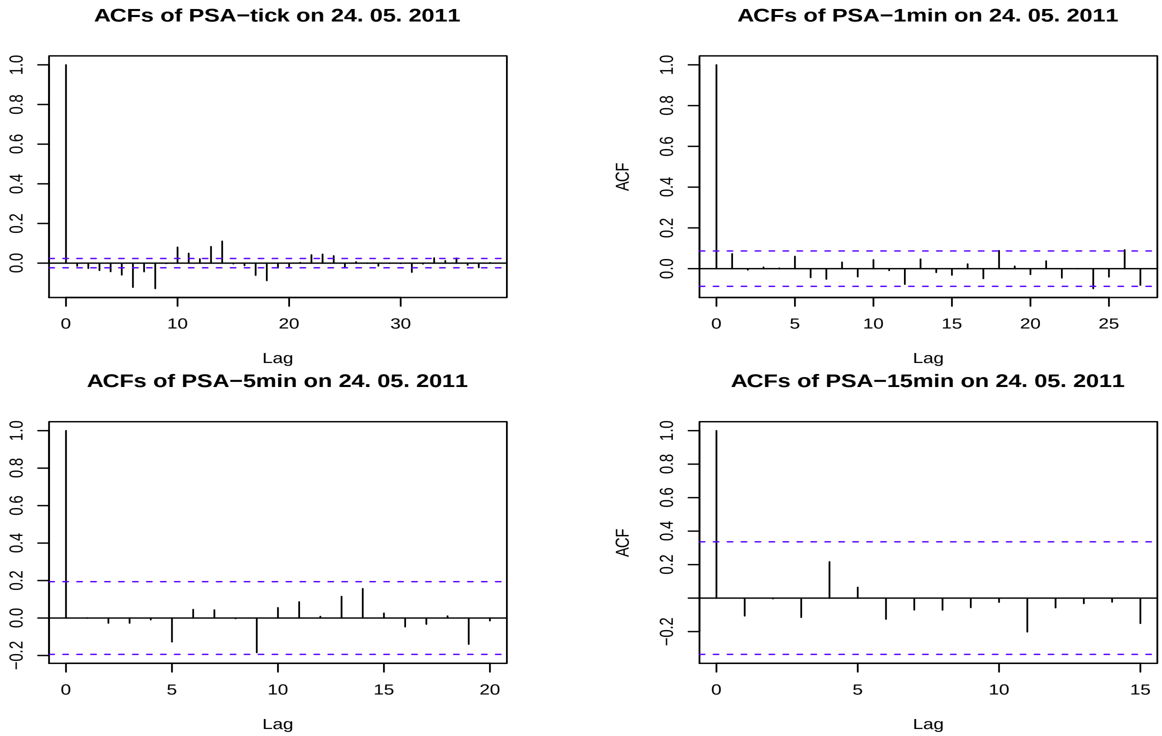

Figure 3: ACFs of Peugeot on 24. May 2011.

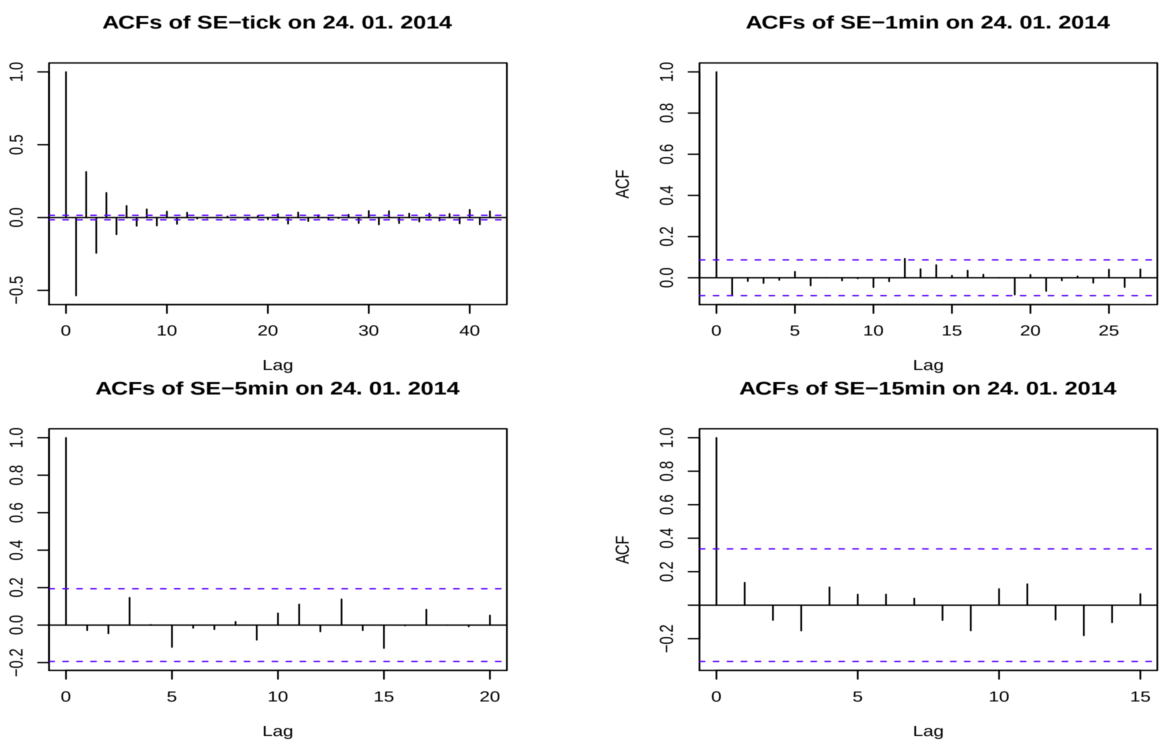

Figure 4: ACFs of Schneider Electric on 24. Jan. 2010.

Table 1: Detailed results of different realized estimators ($*10^4$) for selected examples

	RV_0				RK			
	tick	1-min	5-min	15-min	IN-tick	IN-1min	IN-5min	DN-tick
THK 07. 05. 2007	3.867	2.413	2.393	2.362	1.558	2.425	2.439	1.611
SIE 01. 12. 2009	3.761	2.518	1.635	0.906	2.258	1.941	1.278	1.744
PSA 24. 05. 2011	5.064	2.699	2.551	1.765	2.657	2.924	2.864	2.448
SE 24. 01. 2014	7.660	2.472	2.044	2.028	2.150	2.114	2.016	2.147

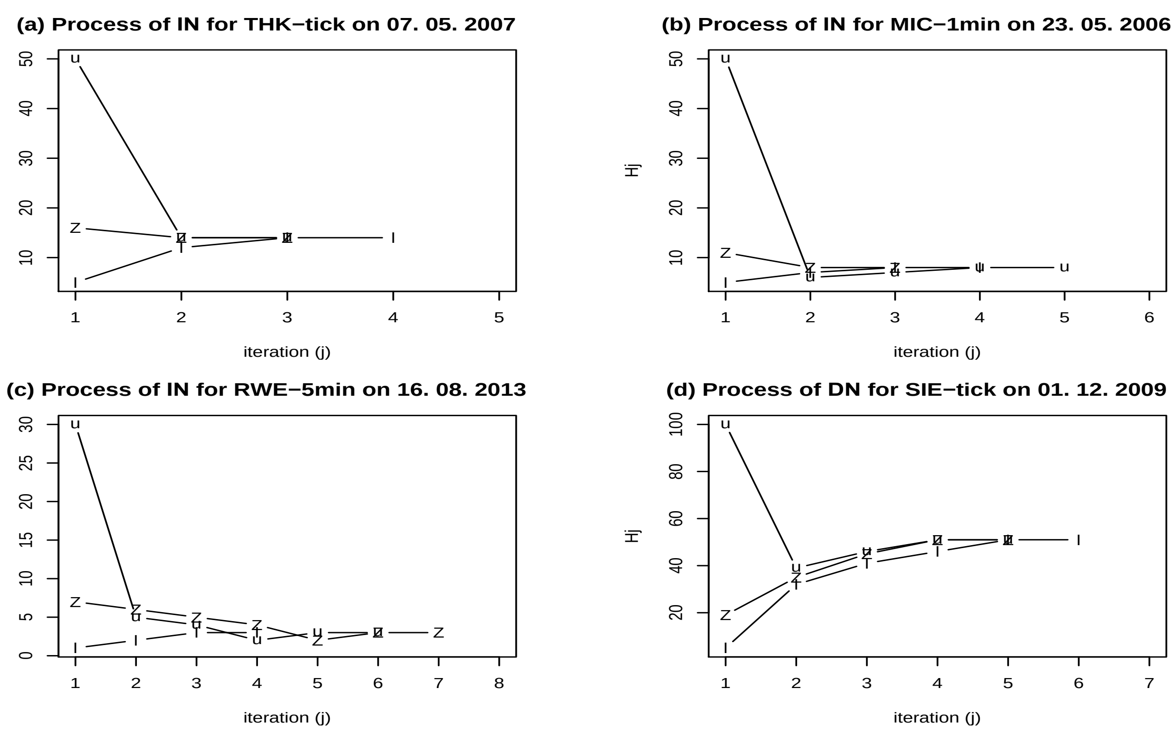

Figure 5: Process of algorithms for four selected days.

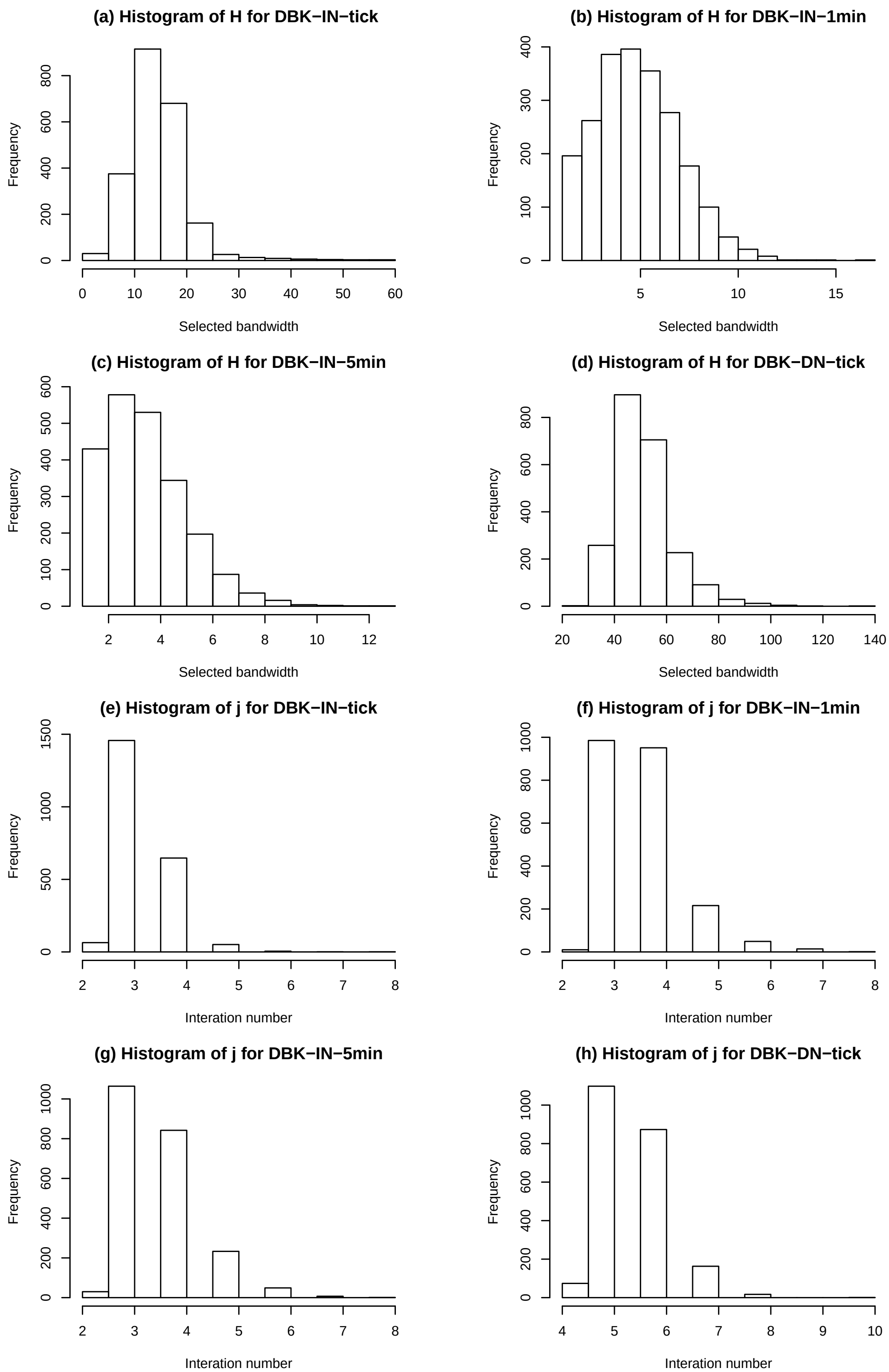

Figure 6: Histograms of selected bandwidth and interation number for German Bank.

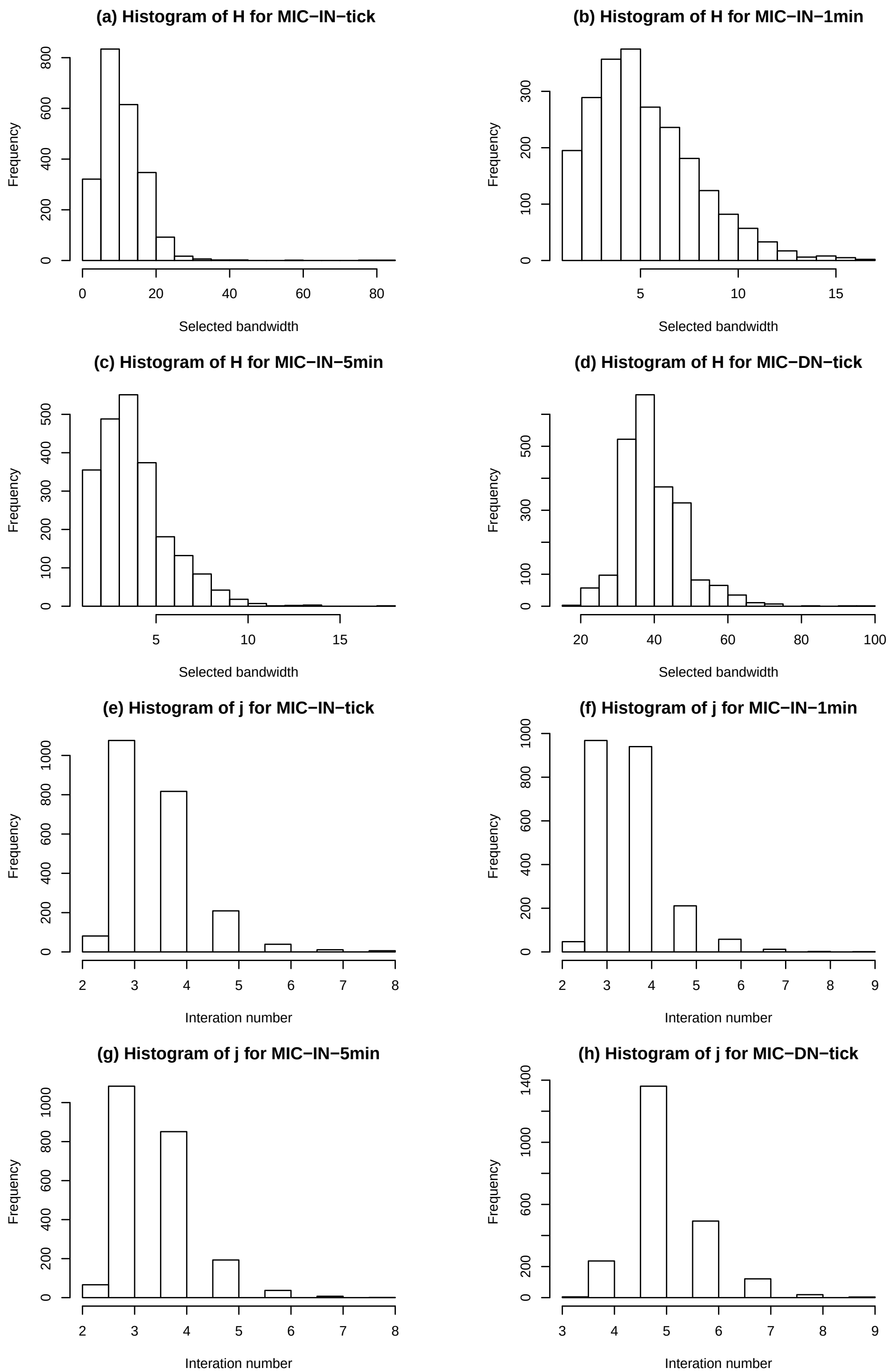

Figure 7: Histograms of selected bandwidth and interation number for Michelin.

Table 2: The number of different cases for the ten companies

Firm	Cases	**IN-tick**	**IN-1min**	**IN-5min**	**DN-tick**
AF	$RV_Z < 0$	0	2	1	0
	$RV_Z > RV_0$	2	308	698	2
	$RK > RV_0$	2	332	748	65
ALV	$RV_Z < 0$	3	0	1	3
	$RV_Z > RV_0$	1	234	683	1
	$RK > RV_0$	0	244	717	10
BMW	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	43	463	748	43
	$RK > RV_0$	52	477	810	156
DBK	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	2	436	818	2
	$RK > RV_0$	5	467	877	21
MIC	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	116	499	748	116
	$RK > RV_0$	133	505	805	356
PSA	$RV_Z < 0$	1	0	1	1
	$RV_Z > RV_0$	6	431	725	6
	$RK > RV_0$	11	437	786	110
RWE	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	25	334	662	25
	$RK > RV_0$	33	336	729	74
SE	$RV_Z < 0$	1	1	1	1
	$RV_Z > RV_0$	68	375	697	68
	$RK > RV_0$	77	391	753	226
SIE	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	20	345	750	20
	$RK > RV_0$	21	344	802	29
THK	$RV_Z < 0$	0	0	0	0
	$RV_Z > RV_0$	22	393	720	22
	$RK > RV_0$	26	403	779	100

Table 3: Mean ($*10^4$) and standard deviation ($*10^4$) of RV_0 and RK based on the different sampling frequencies for the ten companies

		AF	ALV	BMW	DBK	MIC	PSA	RWE	SE	SIE	THK
RV_0-tick	mean	11.253	8.289	6.323	17.031	8.059	11.861	5.051	7.491	6.092	7.848
	s.d.	13.077	20.982	10.094	143.075	13.343	13.382	7.559	13.962	14.040	12.776
RV_0-1min	mean	7.042	4.744	4.653	6.219	5.534	8.135	3.396	4.693	3.694	5.592
	s.d.	6.274	9.902	5.951	11.649	6.239	7.646	4.757	6.561	6.996	7.086
RV_0-5min	mean	5.949	4.188	4.303	5.679	4.804	7.171	2.976	3.839	3.324	4.918
	s.d.	5.710	9.479	6.221	10.227	5.651	7.302	4.530	4.957	6.910	6.279
RV_0-15min	mean	5.458	3.988	4.105	5.544	4.565	6.748	2.805	3.526	3.169	4.645
	s.d.	5.663	9.558	6.172	11.019	5.608	7.695	4.884	4.400	7.132	5.869
RK-IN-tick	mean	5.858	4.290	4.293	5.811	4.751	7.119	3.042	4.119	3.468	5.203
	s.d.	5.748	8.804	5.833	9.533	5.695	7.074	4.137	6.357	6.454	7.492
RK-IN-1min	mean	5.518	3.995	4.091	5.472	4.562	6.843	2.783	3.573	3.211	4.688
	s.d.	5.169	8.999	5.786	10.065	5.371	6.895	3.859	4.796	7.130	5.893
RK-IN-5min	mean	4.836	3.679	3.733	4.977	4.114	6.063	2.472	3.177	2.901	4.243
	s.d.	4.914	9.153	5.847	9.191	5.185	6.833	3.750	4.214	6.895	5.705
RK-DN-tick	mean	5.579	4.106	4.194	5.533	4.701	6.980	2.864	3.876	3.328	4.830
	s.d.	5.165	8.477	5.579	9.586	5.508	6.916	3.984	5.608	6.377	6.074

Table 4: "Points over 95% VaR" (benchmark: 111) by means of the SEMIFAR model based on the different logarithmic realized measures for the ten companies

		AF	ALV	BMW	DBK	MIC	PSA	RWE	SE	SIE	THK
RV_0-tick	PoV	53	41	53	52	62	61	47	57	38	78
	deviation	-58	-70	-58	-59	-49	-50	-64	-54	-73	-33
RV_0-1min	PoV	92	96	80	102	89	111	88	102	81	102
	deviation	-19	-15	-31	-9	-22	0	-23	-9	-30	-9
RV_0-5min	PoV	214	134	107	136	122	147	115	144	111	138
	deviation	+103	+23	-4	+25	+11	+36	+4	+33	0	+27
RV_0-15min	PoV	157	138	107	131	126	153	123	146	116	153
	deviation	+46	+27	-4	+20	+15	+42	+12	+35	+5	+42
RK-IN-tick	PoV	138	109	86	110	112	131	101	117	84	113
	deviation	+27	-2	-25	-1	+1	+20	-10	+6	-27	+2
RK-IN-1min	PoV	148	137	103	130	116	144	113	139	110	141
	deviation	+37	+26	-8	+19	+5	+33	+2	+28	-1	+30
RK-IN-5min	PoV	187	158	127	154	152	176	147	176	139	169
	deviation	+76	+47	+16	+43	+41	+65	+36	+65	+28	+58
RK-DN-tick	PoV	143	122	87	119	112	131	110	124	92	127
	deviation	+32	+11	-24	+8	+1	+20	-1	+13	-19	+16

Table 5: Summary of "Points over 95% VaR" based on the lnRV-SEMIFAR

Realized measure	**Total absolute deviation**	**Total deviation**
RK-IN-tick	121	-9
RK-DN-tick	145	57
RV_0-1min	167	-167
RK-IN-1min	189	171
RV_0-15min	248	240
RV_0-5min	266	258
RK-IN-5min	475	475
RV_0-tick	568	-568

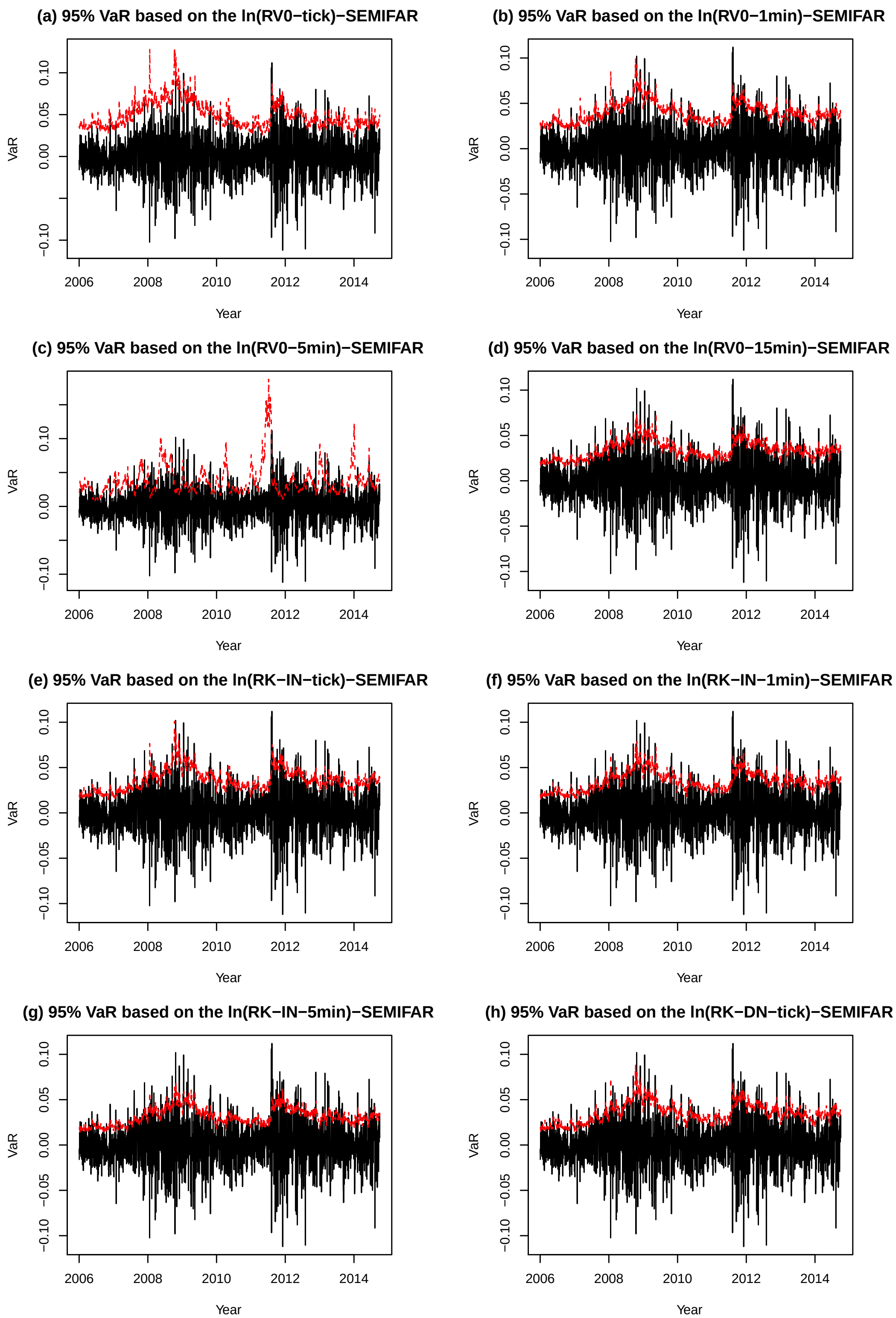

Figure 8: Loss and 95% VaR based on the lnRV-SEMIFAR model for Air France.

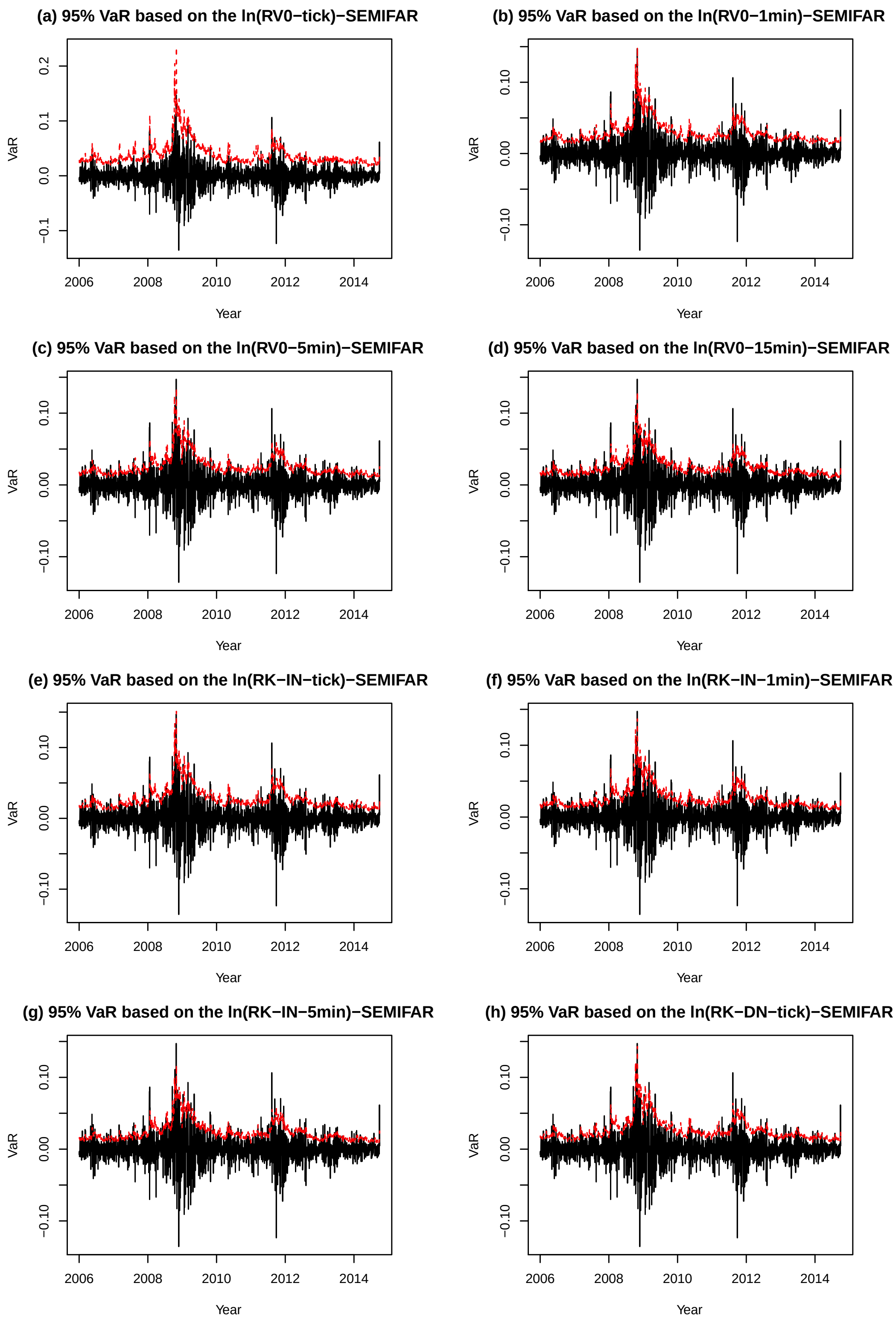

Figure 9: Loss and 95% VaR based on the lnRV-SEMIFAR model for Allianz.

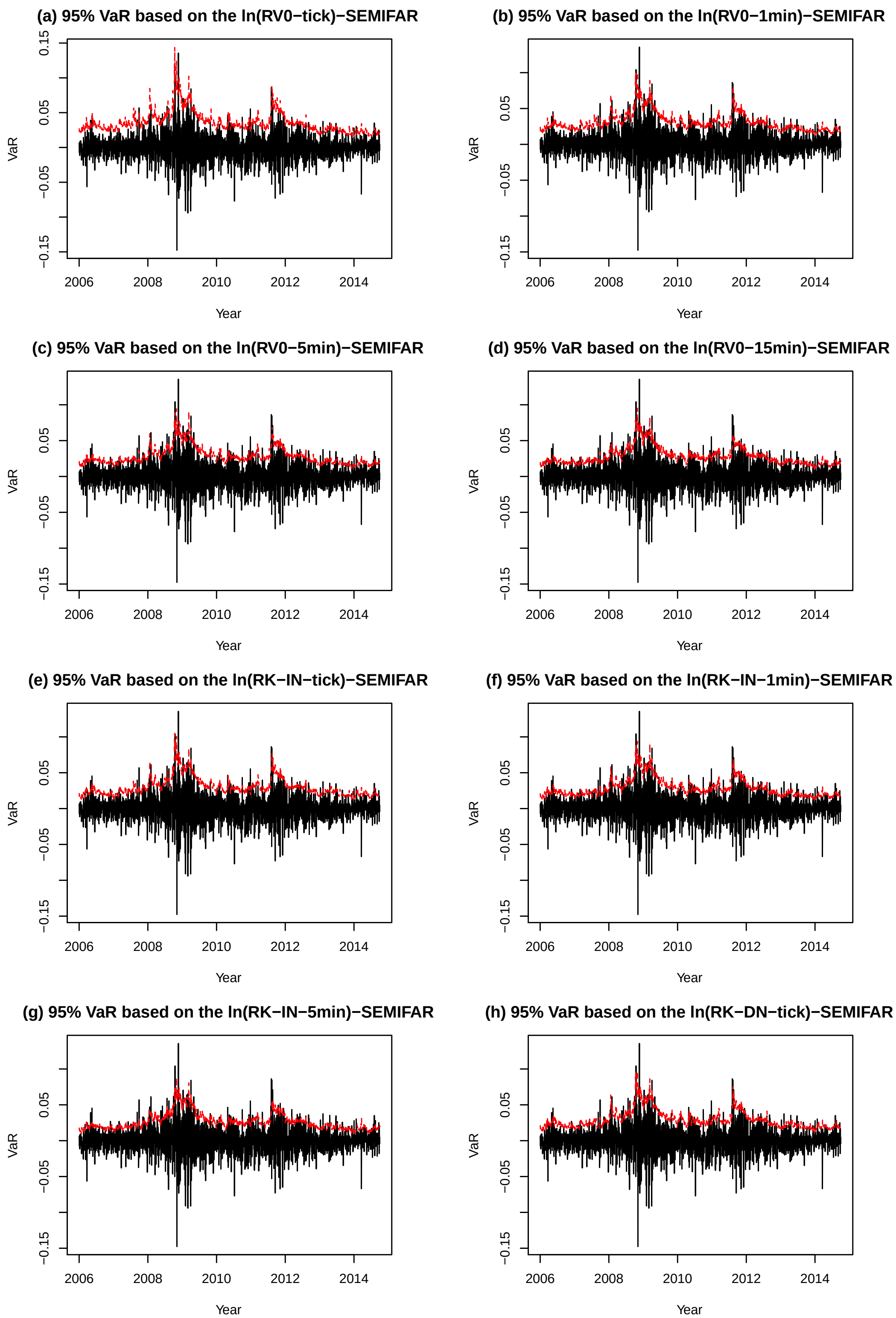

Figure 10: Loss and 95% VaR based on the lnRV-SEMIFAR model for BMW.

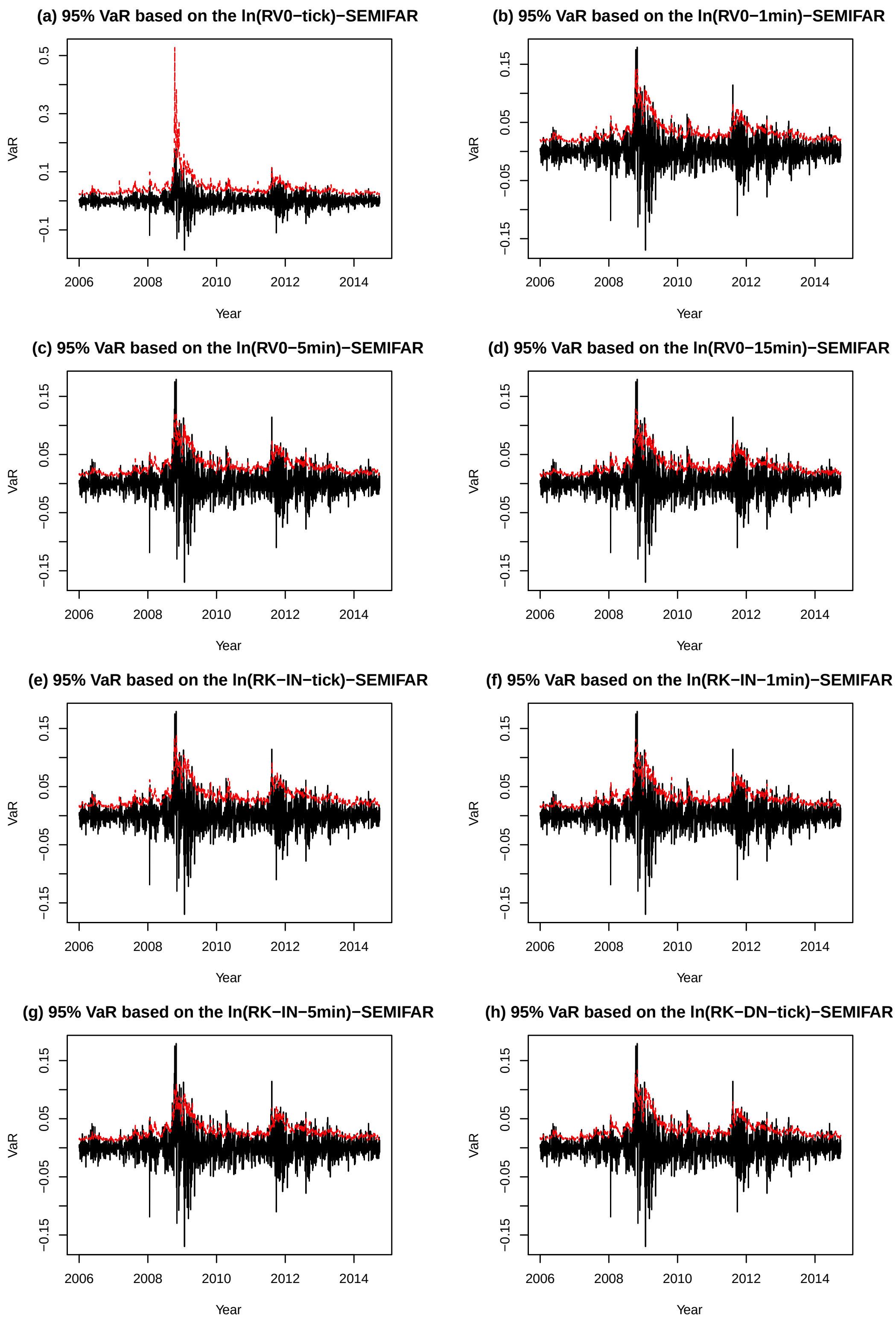

Figure 11: Loss and 95% VaR based on the lnRV-SEMIFAR model for German Bank.

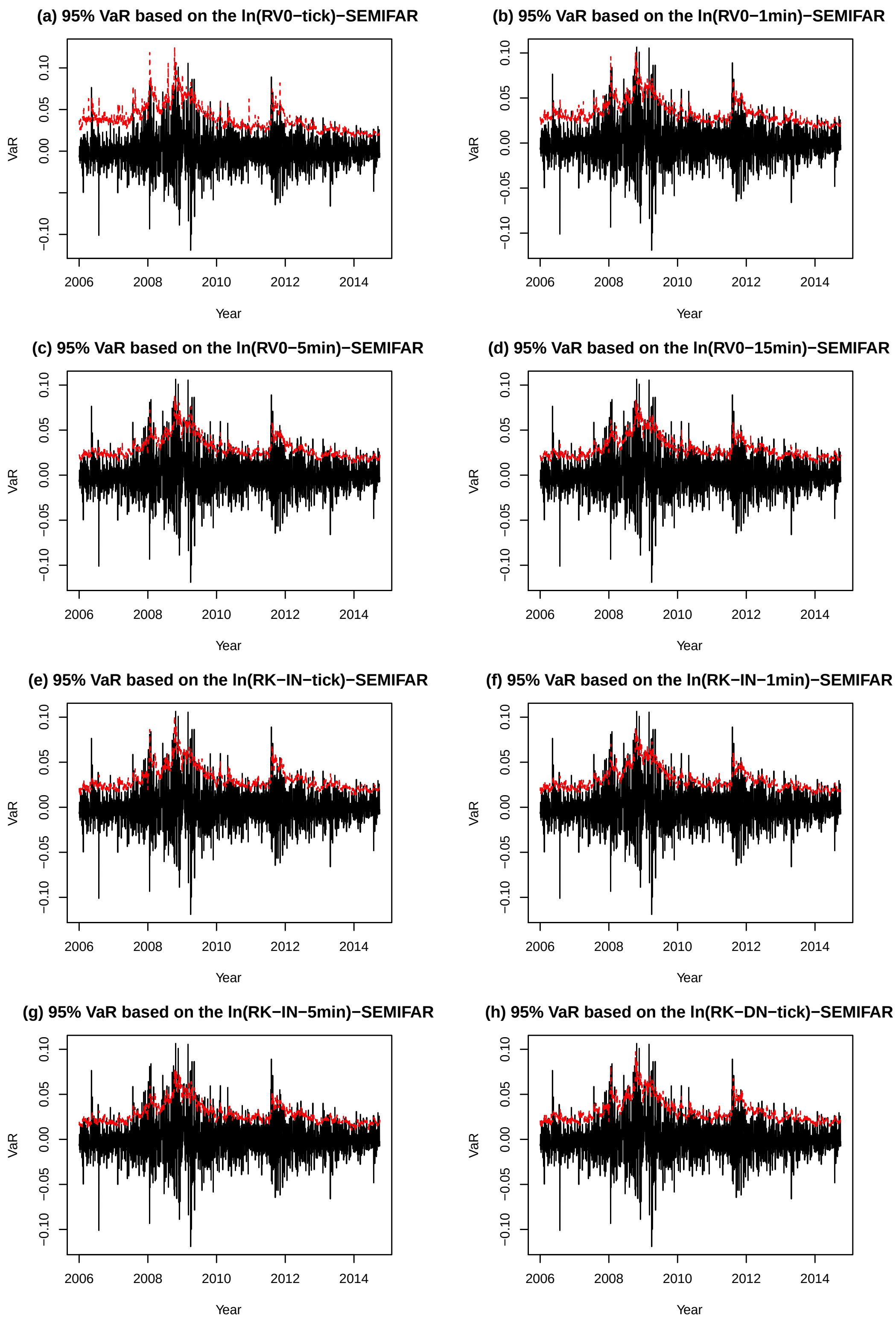

Figure 12: Loss and 95% VaR based on the lnRV-SEMIFAR model for Michelin.

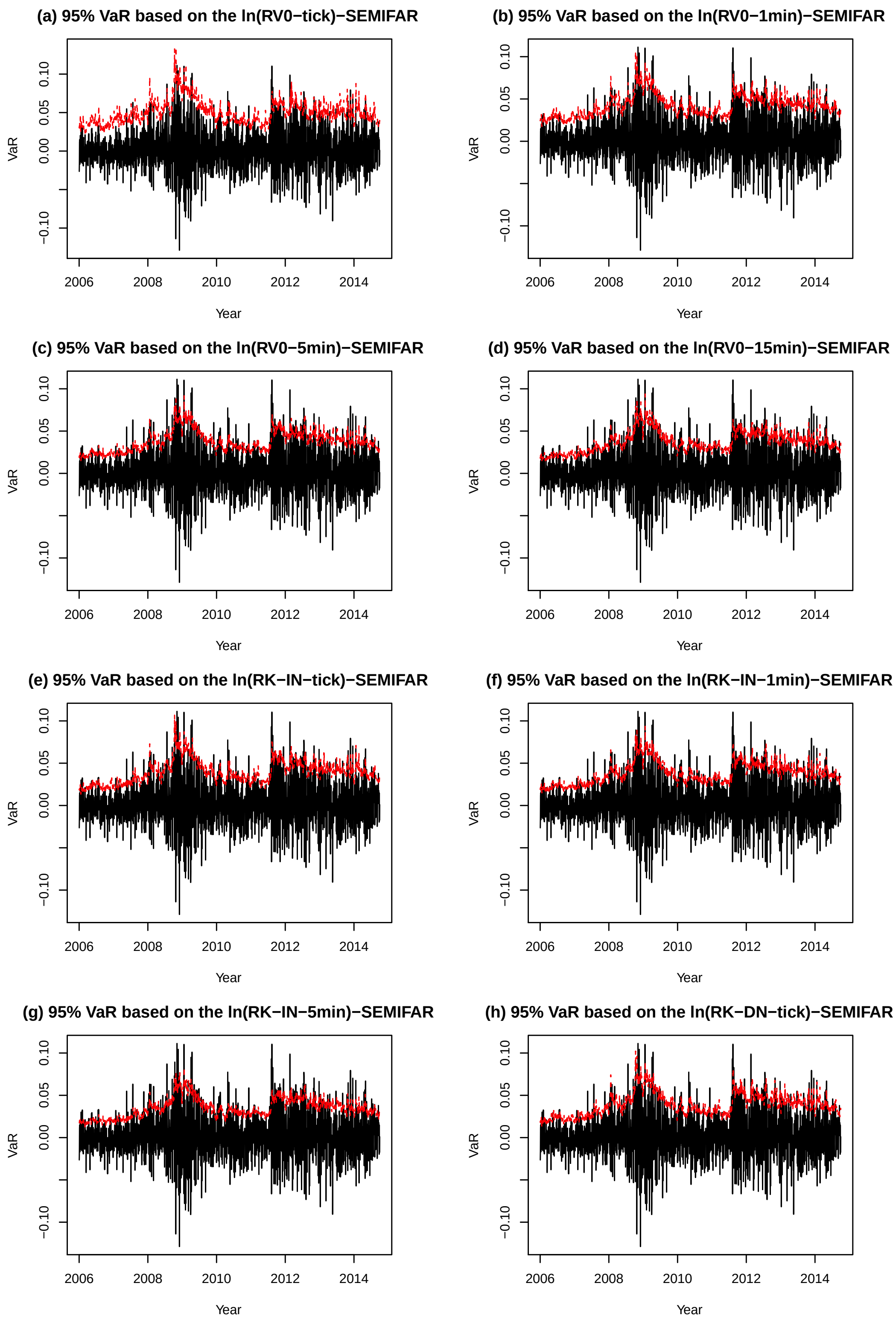

Figure 13: Loss and 95% VaR based on the lnRV-SEMIFAR model for Peugeot.

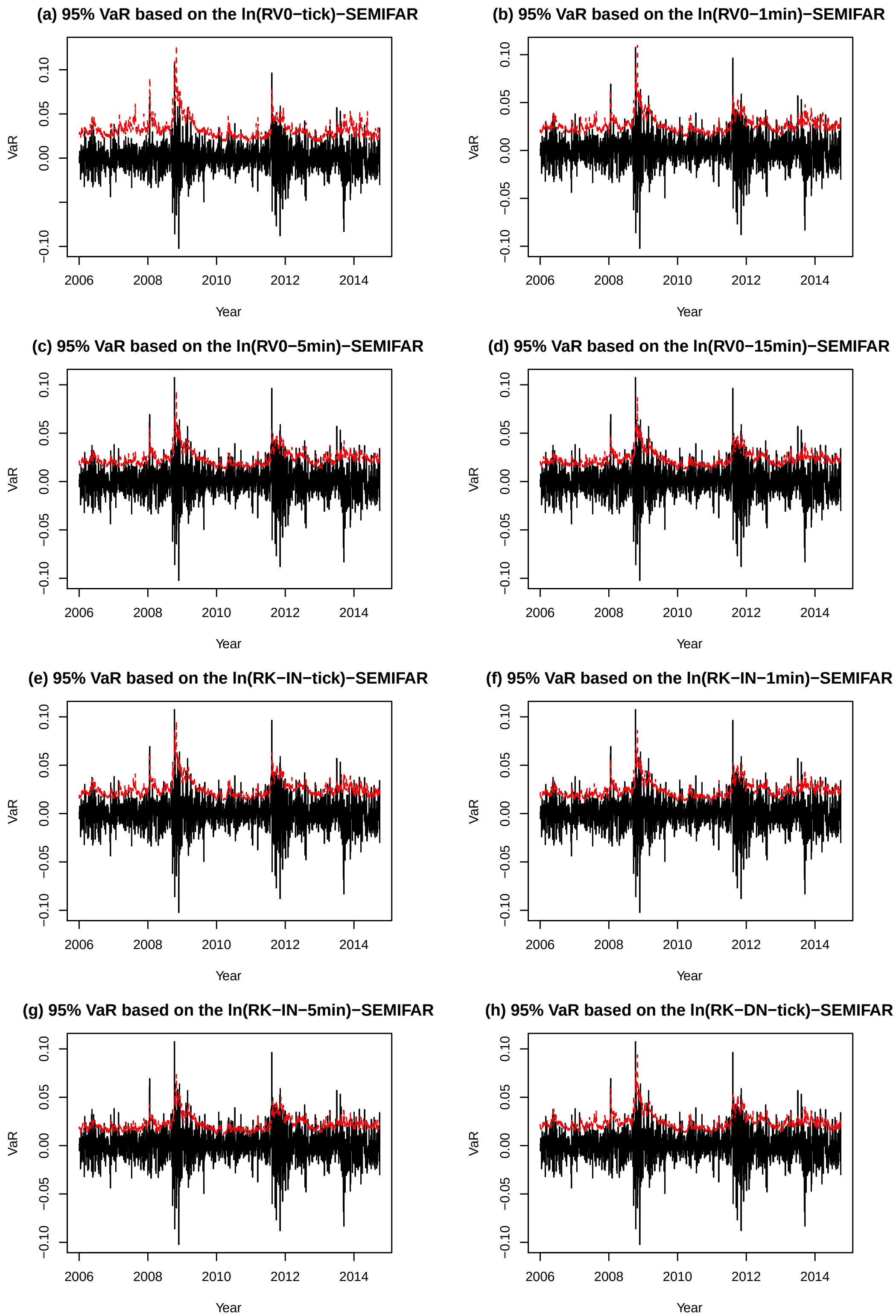

Figure 14: Loss and 95% VaR based on the lnRV-SEMIFAR model for RWE.

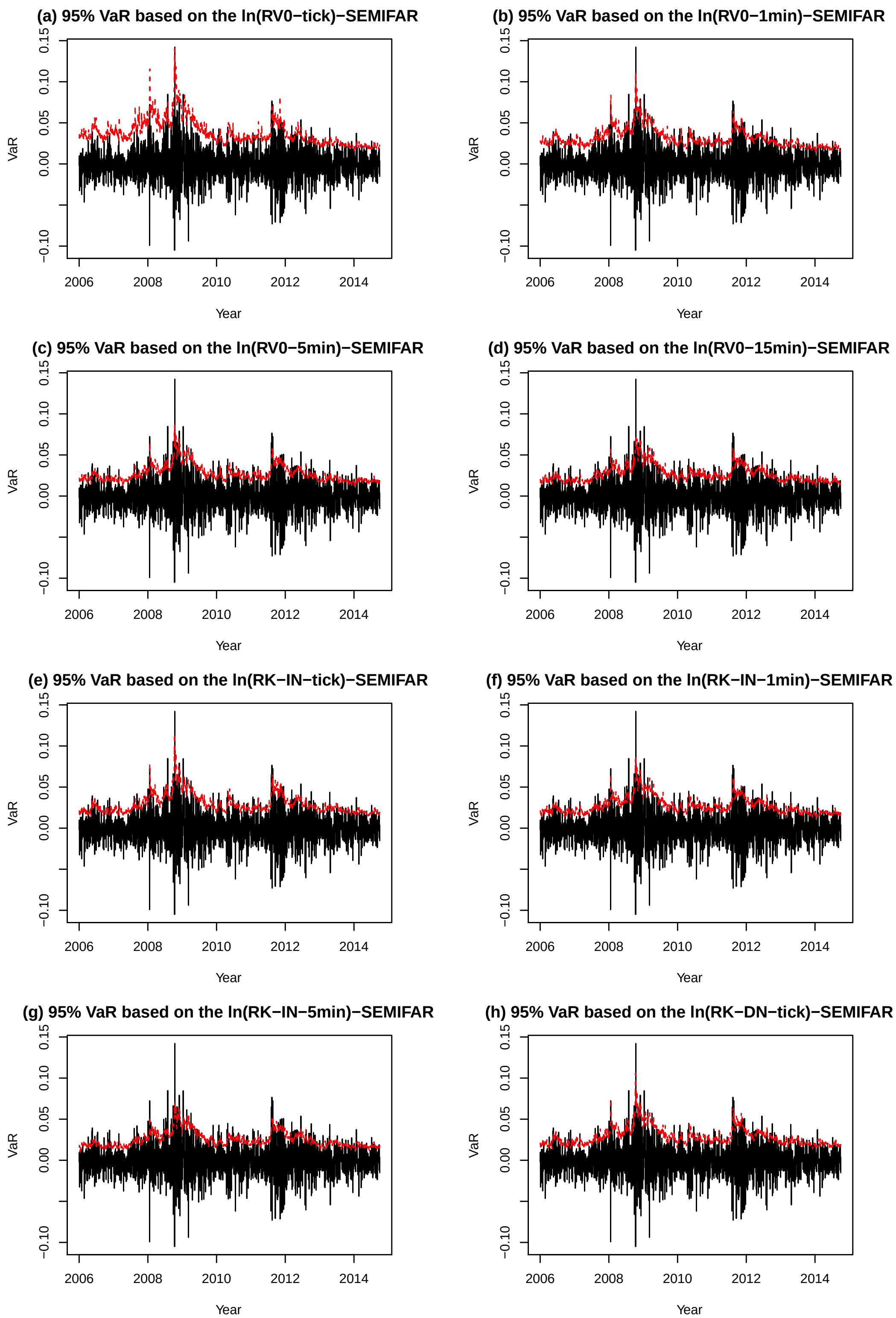

Figure 15: Loss and 95% VaR based on the lnRV-SEMIFAR model for Schneider Electric.

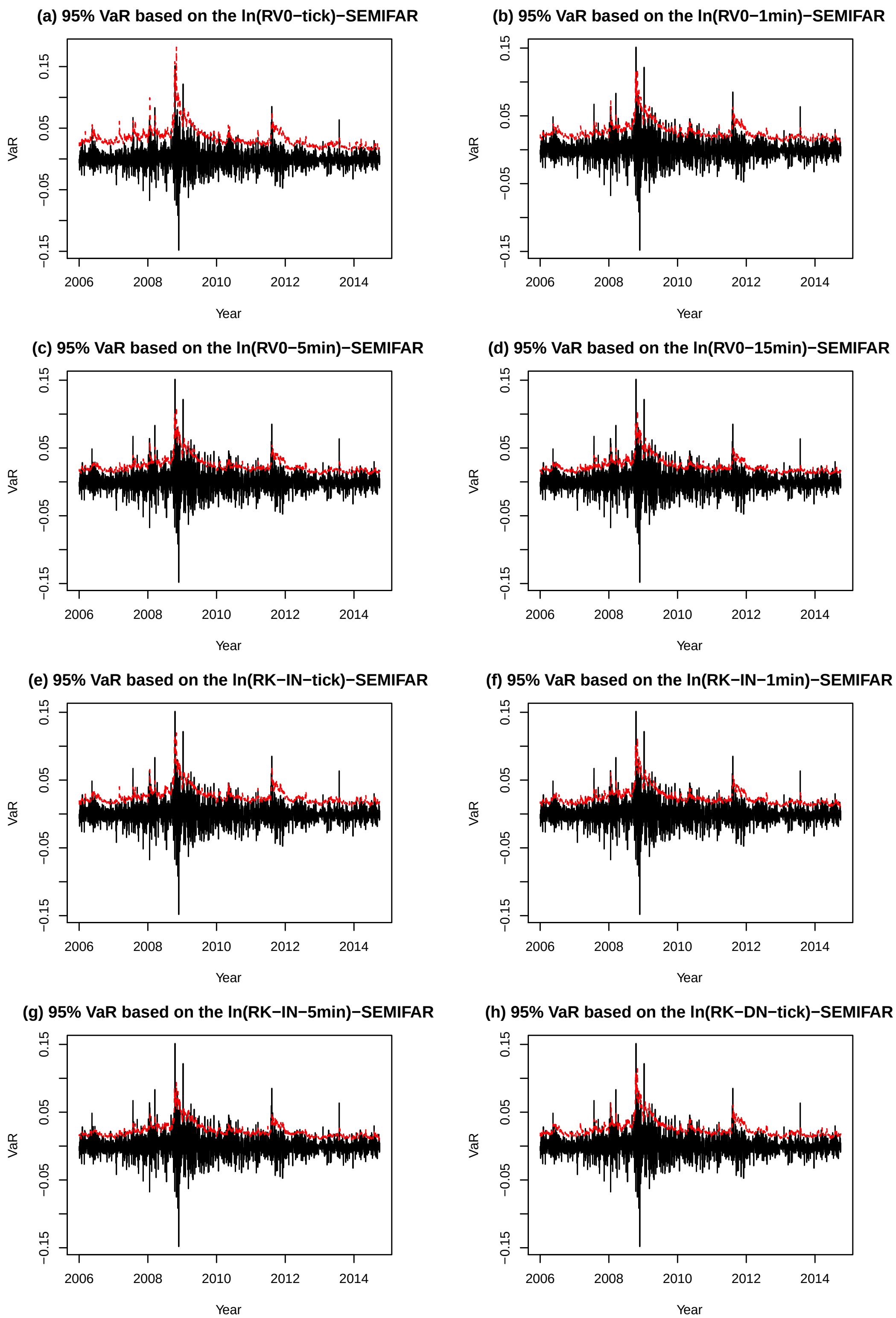

Figure 16: Loss and 95% VaR based on the lnRV-SEMIFAR model for Siemens.

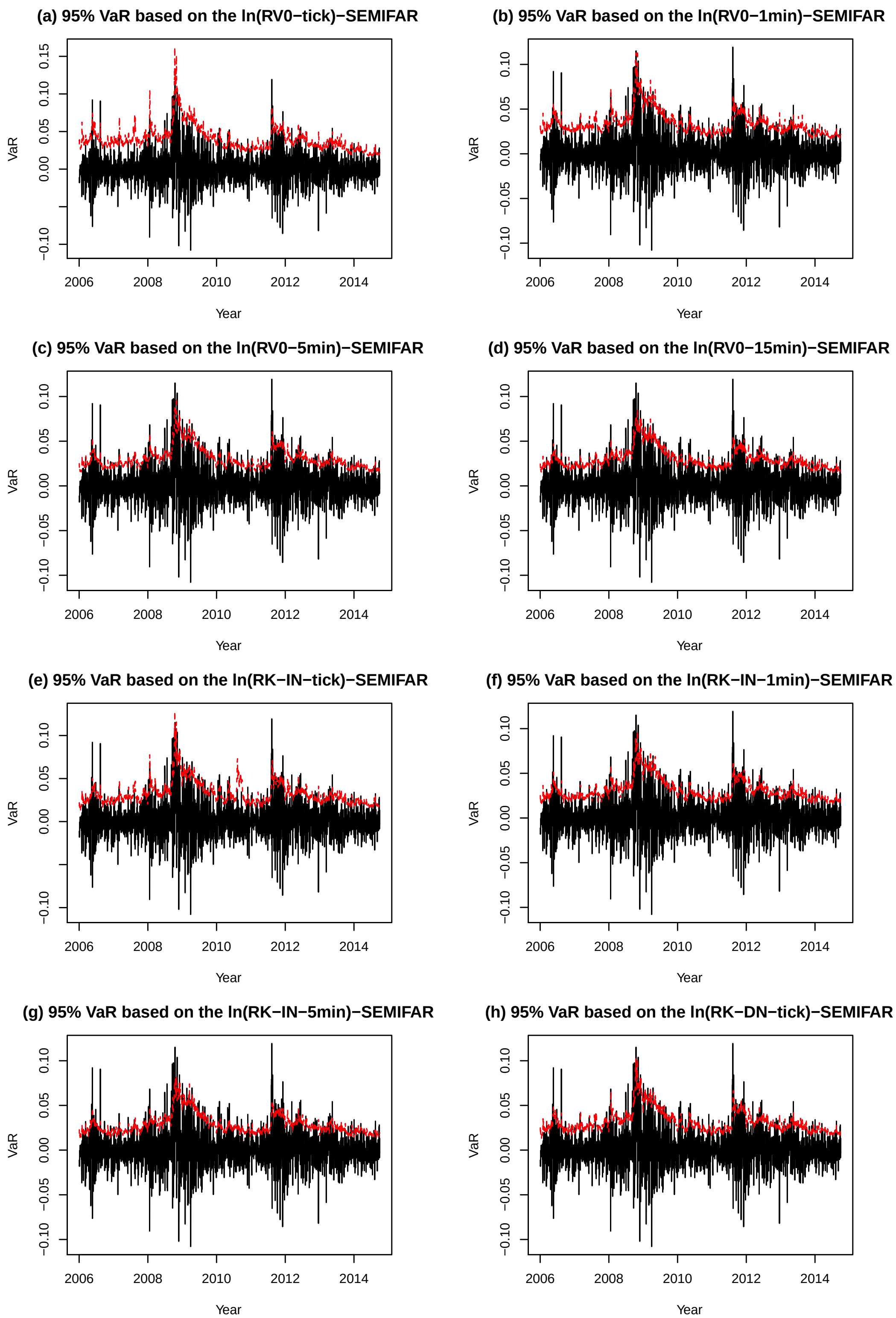

Figure 17: Loss and 95% VaR based on the lnRV-SEMIFAR model for Thyssenkrupp.