Recent developments in the SIML estimation of integrated volatility with high frequency financial data *

Hiroumi Misaki †

October, 2016

1. Introduction

Estimating the volatility and covariance of asset prices has been a key issue in finance, since it is very important for option pricing, asset allocation, risk management, and so on. By now it is possible to use a large number of high-frequency data in financial markets and considerable interest has been paid on the estimation problem of the volatility and covariances by using high-frequency data in financial econometrics.

One of the conventional methods to estimate the volatility and covariance is the realized volatility, introduced by Andersen et al. (2001). The realized volatility is defined by simply summing up the intraday squared returns. They have argued that under some appropriate assumptions the realized volatility converges to the integrated volatility, which is a natural measure of volatility. However, it has been well known that the realized volatility works poorly when there exist micro-market noise, of which we cannot ignore the affects in actual markets.

To deal with this problem several new statistical estimation methods have been developed. See Ait-Sahalia et al. (2005), Zhang et al. (2005), Bandorff-Nielsen et

---

*This is an abstract for International Symposium on Statistical Analysis for Large Complex Data, Tsukuba City, Japan, November 2016.

†Faculty of Engineering, Information and Systems, University of Tsukuba, Tennodai 1-1-1, Tsukuba City, Ibaraki 305-8577, JAPAN, hmisaki@risk.tsukuba.ac.jp
al. (2008), and Malliavin and Mancino (2009) as recent literatures on the related topics among many.

In this respect Kunitomo and Sato (2011, 2013) have proposed a new statistical method called the Separating Information Maximum Likelihood (SIML) method for estimating the integrated volatility and the integrated covariance by using high frequency data under the presence of micro-market noises. They have shown that the SIML estimator has reasonable asymptotic properties as well as finite sample properties.

Misaki and Kunitomo (2015) has further investigated the properties of the SIML estimation when we have the micro-market noises and randomly sampled data at the same time. In actual high frequency financial transaction data they are usually recorded irregularly which can be regarded as observations at random times and the effects of the randomness could be significant when we have micro-market noises.

When we wish to estimate the integrated covariance and its functionals by the SIML estimation in multivariate case, we need synchronized data of two or more series of assets, but actual transactions are usually non-synchronously observed. Kunitomo, Misaki and Sato (2015) has studied for this situation and reported reasonable results, though we shall mainly focus on univariate case in this presentation.

2. Micro-market noise models and the estimation of the integrated volatility

We introduce the high frequency financial price models with micro-market noise and the Separating Information Maximum Likelihood (SIML) method employed in Misaki and Kunitomo (2015).

2.1 A General Formulation

Let \( y(t^n_i) \) be the \( i \)-th observation of the (log-) price at \( t^n_i \) for \( 0 = t^n_0 < t^n_1 < \cdots < t^n_{n^*} \leq t^n_n = 1 \) and \( t^n_{n^*} = \max_{i \leq 1} t^n_i \). (We shall use \( n \) and \( n^* \) as the indices of sample size; \( n^* \) is a stochastic index while \( n \) is a constant index.) We set \( y_n = (y(t^n_i)) \) be
a vector of \( n^{*} \) observations. We consider the situation that the high-frequency data are observed at random times \( t_{i}^{n} \) and for the simplicity we assume some conditions on random sampling.

**Assumption 2.1**: There exists a positive constant \( c \) such that

\[
\lim_{n \to \infty} t_{i}^{n} = 1, \quad \frac{n^{*}}{n} \to c
\]

for \( 0 = t_{0}^{n} < t_{1}^{n} < \cdots < t_{n}^{n} = 1 \) and

\[
E \left[ |t_{i}^{n} - t_{i-1}^{n}| \right] = O(n^{-1})
\]

as \( n \to \infty \).

**Assumption 2.2**: The underlying stochastic process \( X(t) \) \((0 \leq t \leq 1)\) is independent of the sequence \( t_{i}^{n} \) \((i \geq 1)\).

We assume that the underlying continuous process \( X(t) \) \((0 \leq t \leq 1)\) is not necessarily the same as the observed (log-)price at \( t_{i}^{n} \) \((i = 1, \cdots, n^{*})\) and

\[
X(t) = X(0) + \int_{0}^{t} \sigma_{x}(s)dB(s) \quad (0 \leq t \leq 1),
\]

where \( B(s) \) is the standard Brownian motion, \( \sigma_{x}(s) \) is the instantaneous volatility function adapted to the \( \sigma \)-field \( \mathcal{F}(x, B(r), r \leq s) \). The main statistical objective is to estimate the integrated volatility

\[
\int_{0}^{1} \sigma_{x}^{2}(s)ds
\]

of the underlying continuous process \( X(t) \) \((0 \leq t \leq 1)\) from the set of discretely observed prices \( y(t_{i}^{n}) \) with the condition that \( \sigma_{x}(s) \) is a progressively measurable function and \( \sup_{0 \leq s \leq 1} E[\sigma_{x}^{4}(s)] < \infty \).

Then we consider the situation when the observed (log-)price \( y(t_{i}^{n}) \) is a sequence of discrete stochastic process generated by

\[
y(t_{i}^{n}) = h \left( X(t_{i}^{n}), y(t_{i-1}^{n}), u(t_{i}^{n}) \right),
\]
where \( h(\cdot) \) is a measurable function, the (unobservable) continuous martingale process \( X(t) \) \((0 \leq t \leq 1)\) is defined by (2.3) and \( u(t^n_i) \) is the micro-market noise process.

The simple additive (signal-plus-noise) measurement error model can be represented by

\[
y(t^n_i) = X(t^n_i) + u(t^n_i),
\]

as a special case of (2.5).

### 2.2 The SIML Method

We consider the basic situation of (2.3) and (2.5) with fixed observation intervals when we have \( y(t^n_i) = X(t^n_i) + u(t^n_i) \) and \( h_n = t^n_i - t^n_{i-1} = 1/n \), where \( X(t) \) \((0 \leq t \leq 1)\) and \( u(t^n_i) \) \((i = 1, \ldots, n)\) are independent with \( \sigma_x^2(s) = \sigma_x^2 \) (a positive constant).

Under the assumption that \( u(t^n_i) \) are independently, identically and normally distributed as \( N(0, \sigma_u^2) \) given the initial condition \( y_0 \), we have

\[
y_n \sim N_n \left(y(0)1_n, \sigma_u^2 I_n + h_n \sigma_x^2 C_n C_n' \right),
\]

where an \( n \times n \) matrix

\[
C_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}
\]

and a \( 1 \times n \) vector \( 1'_n = (1, \ldots, 1) \).

When the observations are randomly sampled, the durations \( \tau^n_i = t^n_i - t^n_{i-1} \) are a sequence of random variables and we take \( n^* \) instead of \( n \) in (2.7) and (2.8). By transforming \( y_{n^*} \) to \( z_{n^*} \) \((z_k)\) by

\[
z_{n^*} = \sqrt{n^*} P_{n^*} C_{n^*}^{-1} (y_{n^*} - \bar{y}_0)
\]

where \( \bar{y}_0 = y(0)1_{n^*} \), \( P_{n^*} = (p_{jk}) \) and for \( j, k = 1, \ldots, n^* \),

\[
p_{jk} = \sqrt{\frac{2}{n^* + 1}} \cos \left( \frac{2\pi}{2n^* + 1} \left( j - \frac{1}{2} \right) \left( k - \frac{1}{2} \right) \right).
\]
Then the transformed variables $z_k$ ($k = 1, \cdots, n^*$) given $n^*$ are mutually independent and $z_k \sim N(0, \sigma_x^2 + a_k, n^* \sigma_u^2)$, where

$$a_{k, n^*} = 4n^* \sin^2 \left( \frac{\pi}{2} \frac{2k - 1}{2n^* + 1} \right) \quad (k = 1, \cdots, n^*) . \tag{2.11}$$

Let $m$ and $l$ be positive integers, which are dependent on $n^*$ and we write $m_{n^*}$ and $l_{n^*}$. Then we define the SIML estimator of $\hat{\sigma}_x^2$ and $\hat{\sigma}_u^2$ by

$$\hat{\sigma}_x^2 = \frac{1}{m_{n^*}} \sum_{k=1}^{m_{n^*}} z_k^2 \quad \text{and} \quad \hat{\sigma}_u^2 = \frac{1}{l_{n^*}} \sum_{k=n^*+1-l_{n^*}}^{n^*} a_{k, n^*}^{-1} z_k^2 , \tag{2.12}$$

respectively.

The numbers of terms $m_{n^*}$ and $l_{n^*}$ are dependent on $n^*$ such that $m_{n^*}, l_{n^*} \to \infty$ as $n \to \infty$. We impose the order requirements that $m_{n^*} = O_p(n^\alpha)$ $(0 < \alpha < \frac{1}{2})$ and $l_{n^*} = O_p(n^\beta)$ $(0 < \beta < 1)$ for $\hat{\sigma}_x^2$ and $\hat{\sigma}_u^2$, respectively.

3. Simulation Studies

We have investigated the finite sample properties of the SIML estimator for the integrated volatility based on a set of simulations. See Misaki and Kunitomo (2015) for the details.

In our basic simulations we consider three cases when the observations are the sum of signal and micro-market noise, which correspond to the models with (2.3) and (2.6). We use the Poisson Sampling as the basic stochastic sampling with the parameter $\lambda = n$. In Case 1 and 2, the signal is Brownian motion with the deterministic volatility function

$$\sigma_x^2(s) = \sigma(0)^2 \left[ a_0 + a_1 s + a_2 s^2 \right] , \tag{3.1}$$

where $a_i$ ($i = 0, 1, 2$) are constants and we need the restriction such that $\sigma_x(s)^2 > 0$ for $s \in [0, 1]$. In this case the integrated volatility is given by

$$\sigma_x^2 = \int_0^1 \sigma_x(s)^2 ds = \sigma_x(0)^2 \left[ a_0 + \frac{a_1}{2} + \frac{a_2}{3} \right] . \tag{3.2}$$

In this example we take flat (or constant) volatility as Case1 and the U-shaped volatility movements as Case 2.
In Case 3 we take that the instantaneous volatility follows the stochastic volatility model
\[ dh(t) = \gamma h(t) \, dt + c \, dW(t) \quad \text{and} \quad \sigma_x(t)^2 = \sigma^2 e^{h(t)}, \]
where \( W(t) \) is another Brownian Motion which is independent of \( B(t) \) and \( u(t) \) \((0 \leq t \leq 1)\) and \( \sigma^2 \) is a constant.

By extending the basic simulation framework reported above, we have conducted a large number of simulations. As a stochastic duration model we use the EACD(1,1) model as Case 4, which was originally proposed by Engle and Russel (1997). It represents the dependent structure on the duration process of transactions, which is often observed in actual financial markets. Let
\[ \tau^n_i = t^n_i - t^n_{i-1} \quad \text{and} \quad \tau^n_i = \psi^n_i \varepsilon^n_i, \]
where \( \psi^n_i = \omega + a \tau^n_{i-1} + b \psi^n_{i-1} \) and \( \varepsilon^n_i \) are a sequence of i.i.d. exponential random variables with \( a > 0, b > 0 \) and \( \omega > 0 \).

We further have adopted the similar micro-market situations investigated by Sato and Kunitomo (2011). In our simulations we use several non-linear transformation models in the form of (2.5). Among them Case 5 corresponds to the linear price adjustment model such that
\[ y(t^n_i) = P(t^n_i) \quad \text{and} \quad P(t^n_{i-1}) - P(t^n_i) = g \left[ X(t^n_i) - P(t^n_{i-1}) \right] + u(t^n_i), \]
where \( u(t^n_i) \) is an i.i.d. sequence of micro-market noises with \( \mathcal{E}[u(t^n_i)] = 0, \mathcal{E}[u(t^n_i)^2] = \sigma^2_u \) and \( g \) is a constant adjustment coefficient.

Case 6 corresponds to the micro-market models with the round-off errors. We set
\[ P(t^n_i) - P(t^n_{i-1}) = g_\eta \left[ X(t^n_i) - P(t^n_{i-1}) + u(t^n_i) \right], \]
where \( u(t^n_i) \) is an i.i.d. sequence of micro-market noises with \( \mathcal{E}[u(t^n_i)] = 0, \mathcal{E}[u(t^n_i)^2] = \sigma^2_u, \quad g_\eta(x) = \eta \left[ x/\eta \right] \) is the round-off part of \( x \), \( \left[ x \right] \) is the largest integer being less than \( x \) and \( \eta \) is a (small) positive constant. This formulation corresponds to the micro-market model with the restriction of the minimum price change with \( \eta \). Hence Case 6 is the basic round-off model while Case 7 corresponds to its variant when
\[ P(t^n_i) - P(t^n_{i-1}) = g_\eta [X(t^n_i) - P(t^n_{i-1})] + u(t^n_i). \]

Case 8 corresponds to the nonlinear price adjustment model with micro-market
noise when the nonlinear transformation is given by

\begin{equation}
    g(x) = g_1 x I(x \geq 0) + g_2 x I(x < 0),
\end{equation}

where \( g_i (i = 1, 2) \) are some constants and \( I(\cdot) \) is the indicator function.

In all cases, the estimates obtained by realized volatility are badly-biased, which have been well known in the analysis of high frequency financial data. The SIML estimate, on the other hand, gives reasonable estimate and the variance of the SIML estimator is within a reasonable range for practical purposes.

4. Empirical Analysis

We have analysed high frequency financial data in the Japanese stock market. Our main purpose of this empirical study is to estimate daily volatility, covariance and other related quantities by using SIML estimator and to compare them to some alternative estimators.

This study is significant for both econometric and practical reasons. First, this is the first research investigating the SIML estimation on randomly observed multivariate data. The property of the SIML estimator on irregularly spaced and non-synchronous data has not been evident. Then we shall examine the SIML estimator by using actual transaction data in the Japanese market.

Second, the SIML estimation has not ever applied to individual stocks. Accurate estimation of the daily integrated volatility covariance of individual stocks is beneficial to financial practice such as asset pricing, risk management and portfolio allocation. Since the SIML method is relatively simple and easy to implement, it is rather important to examine the usefulness of the method in this respect.

The data possibly have the market microstructure noise and irregular and unsynchronous observation time. The SIML estimator is so simple that it can be applied to both univariate and multivariate time series with market microstructure noise. For estimating the realized volatility of each single stock, we have constructed the SIML estimator with the whole of the high-frequency series, while in the mul-
tivariate case we have used the refresh time scheme in order to synchronize the objective series.

We have found that the SIML estimation provides reasonable results in any case whereas most of the examined alternatives are severely biased. In addition, we have found that the SIML estimates are similar to the realized volatilities and covariances based on relatively long intervals in respect of the summary statistics across the sample period. Our detailed analysis, however, have indicated they are not always coincide to each other. It may be considered as rather reasonable to employ the SIML estimation since it uses hundreds or thousands observations whereas the realized volatility and covariance based on, say, 20 minute returns can uses only 14 samples per day.

In conclusion, our investigation suggest that the SIML estimation is useful to estimate the daily integrated volatility, covariance, and other related quantities in actual markets.

5. Finite Sample Variance of the SIML Estimator

When we wish to carry out statistical inference, we need an estimate of the variance or standard deviation of estimators. Although Kunitomo and Sato (2011, 2018), Misaki and Kunitomo (2015) and Kunitomo, Misaki and Sato (2015) have given the asymptotic properties of the SIML estimator, the finite sample properties are yet to investigate.

One of the possible approach to obtain the variance and its functionals of an estimator in complex models is to utilize the bootstrap methods, introduced by Efron (1979). The initial bootstrap method is based on independent variables but later extended to dependent variables including time series by several researchers. See Kreiss and Paparoditis (2011) and Kreiss and Lahiri (2012) for reviewing several types of bootstrap methods for time series data such as the moving block bootstrap (MBB) and its variants.

In our formulation there would be two possible approaches to exploit bootstrap methods for the SIML estimator. One is to resample the transformed variables $z_k$
which is mutually independent (but not identical), and the other is to apply the
MBB to the difference of log price $y_t$. However it is not straightforward to apply the
MBB to our situation since we have to check whether the assumptions of the MBB
methods are satisfied, select the optimal size of the moving blocks, and so on. In
this presentation the preliminary report on estimating the finite sample variance of
the SIML estimator will be given.

6. Concluding Remarks

In this presentation, we have shown that the SIML estimator has the asymptotic
robustness when the high frequency financial data are randomly sampled. They
include not only the cases when we have the micro-market noises but also the cases
when the micro-market structure has the nonlinear adjustments and the round-off errors under a set of reasonable assumptions. By conducting a large number
of simulations, we have confirmed that the SIML estimator has reasonable robust
properties in finite samples even in the non-standard situations. We have also shown
the reasonable results in empirical analysis by using the tick data in the Japanese
financial market, and the preliminary report on estimating the finite sample variance
of the SIML estimator has been given.

References

Continuous-Time Process in the Presence of Market Microstructure Noise,”

: determining integrated volatility with noisy-high frequency data,” Journal of
the American Statistical Association, Vol. 100(472), 1394-1411.


realized kernels to measure the ex-post variation of equity prices in the
presence of noise,” Econometrica, Vol.76-6, 1481-1536.


