A State Space Approach to Estimating Integrated Variance and Microstructure Noise*

Daisuke Nagakura † and Toshiaki Watanabe ‡

October, 2008

Abstract

In this paper, we consider estimation of integrated variance (IV) as well as microstructure noise by using a state space method. Our method is based on the result obtained in Meddahi (2003) and Barndorff-Nielsen and Shephard (2002), namely, when the underlying true log-price process follows a certain class of continuous-time stochastic volatility (SV) models, the IV follows an ARMA(1, 1) process. Assuming that an observed log-price is the sum of the true log-price and an i.i.d. microstructure noise, we show that the realized variance (RV) calculated from the observed log-prices, which we call "noise-contaminated RV (NCRV)", follows an ARMA(1, 2) process. We represent the NCRC by a state space form showing that the model parameters are identifiable. We estimates the IV and microstructure noises by applying the Kalman filtering to the state space form. The proposed method is applied to yen/dollar exchange rate data. We find that the magnitude of a microstructure noise is, in average, about 23% $\sim$ 32% of the NCRV.

Key Words: Realized Variance; Integrated Variance; Microstructure Noise; State Space

---

*This research is partially supported by Grant-in-Aid for Scientific Research 18203901 from the Japanese Ministry of Education, Science, Sports, Culture and Technology. Views expressed in this paper are those of authors and do not necessarily reflect the official views of the Bank of Japan

†Institute for Monetary and Economic Studies, Bank of Japan, E-mail: daisuke.nagakura@boj.or.jp

‡Institute of Economic Research, Hitotsubashi University, E-mail: watanabe@ier.hit-u.ac.jp

1 Introduction

The variance of financial asset returns are known to change over time. More specifically, the variance, or volatility, the square root of the variance, tends to be large (small) following successive large (small) variances in previous periods. This phenomenon is known as "volatility clustering." The model, known as the "stochastic volatility (SV) model", in which the volatility is changing over time, have been used for modeling this kind of volatility dynamics of financial assets. Based on the SV model with estimated model parameters, researchers can estimate the changing variances. See, for example, Ghysels, Harvey and Renault (1996) for the reviews of some of the older papers on SV models and Shephard (2005) for a list of selected papers in the SV literature.

Recently, a new class of estimators for the changing variances has been developed by Barndorff-Nielsen and Shephard (2001, 2002), Andersen, Bollerslev, Diebold and Ebens (2001) and Andersen, Bollerslev, Diebold and Labys (2001). The estimator is called the "realized variance (RV)". The RV employs very high frequency financial data such as minute-by-minute return data or entire records of quote or transaction price data. Under certain assumptions, the RV is a consistent estimator for the "integrated variance (IV)". The formal definition of IV will be given in Section 2. Intuitively, it is a measure of a variability of financial assets over a period specified by researchers (for example, a day). The RV is a model-free estimator in the sense that we do not have to specify the volatility dynamics.

One of the key assumptions needed for the consistency of the RV is that there are no measurement errors in observed log-prices. The measurement error is particularly called "microstructure noise" and is due to, for example, discreteness of prices, bid-ask bounce and irregular trading. When this assumption is violated, the RV is no longer a consistent estimator of the IV. It can be shown that under the existence of microstructure noises, the RV diverges as the sampling frequency tends to be high. Several alternative estimators of the RV, that are consistent even under the existence of microstructure noises, have been proposed by Zhou (1996), Zhang, Mykland and Aït-Sahalia (2005), Hansen and Lunde (2006) and Bandi and Russell (2006). See also Bandi and Russell (2008) which consider a mean-squared-error optimal sampling theory for reducing the effect of microstructure noise.

In this paper, we propose a new approach for estimating the IV under the existence of microstructure noises. Our method is based on the result obtained in Meddahi (2003) and Barndorff-Nielsen and Shephard (2002), namely, when the underlying true log-price process follows a certain class of continuous-time SV models, the IV follows an ARMA(1, 1) process. Assuming that an observed log-price is the sum of the true log-price and an i.i.d. microstructure noise, we show that the RV calculated from the observed log-prices, which we call "noise-contaminated RV (NCRV)", follows an ARMA(1, 2) process. Then, we represent the NCRC by a state space form showing that the model parameters are identifiable. We estimates the IV as well as microstructure noises by applying the Kalman filtering to the state space form. The proposed method is applied to yen/dollar spot exchange rate data. We find that the magnitude of a microstructure noise is, in average, about 23% $\sim$ 32% of the NCRV for each day depending on the sampling frequency.

The rest of the paper is organized as follows. In the next section, we introduce a class of SV models employed in this paper. We also define formally the RV, IV, and microstructure noise in this section. In Section 3, we explain our approach for estimating IV and microstructure noises in details. Section 4 provides an empirical analysis applying our method to yen/dollar spot exchange rate. The last section provides summary and concluding remarks.

2 Continuous Time SV model and RV

2.1 One factor SR-SARV model

Let $p(t)$ be the log of (efficient) spot price at time t. We suppose that $p(t)$ follows the class of continuous-time SV model considered in Meddahi (2003), which is termed as "one factor square-root stochastic autoregressive variance (SR-SARV) model." In this class of models, $p(t)$ follows the diffusion process defined as

$$dp(t) = \sigma(t)dW(t), \quad \sigma^2(t) = \sigma_0^2 + \omega_0 P(f(t)), \tag{1}$$

where $f(t)$ is a state-variable process and the function $P(\cdot)$ is defined so that

$$\begin{aligned} &E[P(f(t))] = 0, \quad \mathrm{var}[P(f(t))] = 1, \\ &E[P(f(t+h))|f(s), p(s), s \leq t] = \exp(-\lambda_0 h)P(f(t)), \quad \forall h > 0. \end{aligned} \tag{2}$$

Note that $E[\sigma^2(s)] = \sigma_0^2$ and $\mathrm{var}[\sigma^2(s)] = \omega_0^2$. This specification of $\sigma^2(t)$ includes many known models such as constant elasticity of volatility (CEV) processes, GARCH diffusion models (Nelson, 1990), eigenfunction stochastic volatility (ESV) models (Meddahi, 2001) and positive Ornstein-Uhlenbeck Levy-driven models (Barndorff-Nielsen and Shephard, 2001).

2.2 Integrated and realized variance

The integrated variance (IV) is defined as

$$IV_t \equiv \int_{t-1}^{t} \sigma^2(s)ds, \quad t = 1, 2, ..., \tag{3}$$

where the unit of t is determined depending on the objective of research. For example, if the researcher is interested in changes in the variances of daily returns, t is interpreted as a day.

Under certain conditions, IV can be consistently estimated by the estimator known as the realized variance (RV), which is defined as

$$RV_t^{(m)} \equiv \sum_{i=1}^{m} r_{t-1+\frac{i}{m}}^{(m)2}, \tag{4}$$

where $r_t^{(m)} \equiv p(t) - p(t - \frac{1}{m}) = \int_{t-\frac{1}{m}}^{t} \sigma(s)dW(s)$ and m is a positive integer. Here, and hereafter, the notation "(m)" implies that its value depends on the sampling frequency m. For example, if t denotes a day and we take observations every five minutes, then $m = 288$ and $r_t^{(288)}$ denotes a five-minute return. It is well known that, as $m \to \infty$, $RV_t^{(m)} \xrightarrow{p} IV_t$.

Meddahi (2003, Proposition 3.1) shows that if the true process of $p(t)$ belongs to the one-factor SR-SARV model introduced in the previous section, then IV_t follows an ARMA(1, 1) process:

$$IV_t = c_{IV} + \phi IV_{t-1} + \eta_t + \theta_{IV}\eta_{t-1}, \tag{5}$$

where η_t is a white noise with $\text{var}(\eta_t) = \sigma_\eta^2$ and $\text{cov}(\eta_t, d_s^{(m)}) = 0$, $d_t^{(m)} \equiv RV_t^{(m)} - IV_t$, for all t and s. The ARMA(1, 1) model parameters, c_{IV}, ϕ, θ_{IV} and σ_η^2 are expressed, in terms of the one-factor SR-SARV model parameters, σ_0^2, ω_0^2 and λ_0, as

$$\begin{aligned}
&\phi = \exp(-\lambda_0), \quad c_{IV} = (1-\phi)\sigma_0^2, \\
&\theta_{IV} = \frac{1-\sqrt{1-4\rho^2}}{2\rho}, \quad \rho = \frac{-\phi + \text{corr}[IV_t, IV_{t-1}]}{1+\phi^2 - 2\phi\text{corr}[IV_t, IV_{t-1}]}, \\
&\sigma_\eta^2 = \frac{(1+\phi^2)\text{var}[IV_t] - 2\phi\text{cov}[IV_t, IV_{t-1}]}{1+\theta_{IV}^2},
\end{aligned} \tag{6}$$

where $\text{var}(IV_t)$, $\text{cov}(IV_t, IV_{t-1})$ and $\text{corr}(IV_t, IV_{t-1})$ are given as

$$\begin{aligned}
&\text{var}[IV_t] = \frac{2\omega_0^2}{(\log\phi)^2}(\phi - \log\phi - 1), \quad \text{cov}[IV_t, IV_{t-1}] = \frac{\omega_0^2}{(\log\phi)^2}(1-\phi)^2, \\
&\text{corr}[IV_t, IV_{t-1}] = \frac{(1-\phi)^2}{2(\phi - \log\phi - 1)}.
\end{aligned} \tag{7}$$

Also, $\sigma_d^{2(m)} \equiv \text{var}[d_t^{(m)}]$ is given as

$$\sigma_d^{2(m)} = \frac{2\sigma_0^4}{m} + \frac{4\omega_0^2 m}{(\log\phi)^2}\left(\phi^{\frac{1}{m}} - 1 - \log\phi^{\frac{1}{m}}\right), \quad \forall m \geq 1. \tag{8}$$

See Meddahi (2003) for the above results. Meddahi (2003, Proposition 3.2) shows that $RV_t^{(m)} = IV_t + d_t^{(m)}$ also follows an ARMA(1, 1) process. Note that the four ARMA(1, 1) model parameters, c_{IV}, ϕ, θ_{IV} and σ_η^2, are completely determined as functions of the three one-factor SR-SARV model parameters, σ_0^2, ω_0 and λ_0.

2.3 Microstructure noise

Now suppose that the observed log-price $p^*(t)$ is contaminated by a measurement error or a microstructure noise so that

$$p^*(t) = p(t) + \varepsilon(t). \tag{9}$$

We assume the following properties on the microstructure noise $\varepsilon(t)$.

Assumption 1

(a) $\varepsilon(t) \sim i.i.d.(0, \sigma_\varepsilon^2)$ with $E[\varepsilon^4(t)] < \infty$.

(b) $\varepsilon(t)$ is independent from $p(t)$ for all s and t.

We do not assume any distribution for $\varepsilon(t)$. The observed return is defined as

$$r_t^{*(m)} \equiv p^*(t) - p^*\left(t - \tfrac{1}{m}\right) = r_t^{(m)} + e_t^{(m)}, \tag{10}$$

where $e_t^{(m)} = \varepsilon(t) - \varepsilon(t - \frac{1}{m})$. It is easy to show that $E\left[e_t^{(m)}\right] = 0$,

$$\mathrm{var}\left[e_t^{(m)}\right] = 2\sigma_\varepsilon^2 \quad \text{and} \quad \mathrm{cov}\left[e_t^{(m)}, e_{t-\frac{i}{m}}^{(m)}\right] = \begin{cases} -\sigma_\varepsilon^2, & i = 1, \\ 0 & i \geq 2. \end{cases} \tag{11}$$

Note that $\mathrm{var}[e_t^{(m)}]$ and $\mathrm{cov}[e_t^{(m)}, e_{t-\frac{1}{m}}^{(m)}]$ do not depend on m. We define the "noise-contaminated RV (NCRV)", denoted by $RV_t^{*(m)}$, as $RV_t^{*(m)} \equiv \sum_{i=1}^{m} r_{t-1+\frac{i}{m}}^{*(m)2}$. Write

$$\begin{aligned} RV_t^{*(m)} &= \sum_{i=1}^{m} \left(r_{t-1+\frac{i}{m}}^{(m)} + e_{t-1+\frac{i}{m}}^{(m)}\right)^2, \\ &= RV_t^{(m)} + u_t^{(m)}. \end{aligned} \tag{12}$$

where

$$u_t^{(m)} \equiv 2\sum_{i=1}^{m} r_{t-1+\frac{i}{m}}^{(m)} e_{t-1+\frac{i}{m}}^{(m)} + \sum_{i=1}^{m} e_{t-1+\frac{i}{m}}^{(m)2}. \tag{13}$$

Note that, unlike $RV_t^{(m)}$ and $RV_t^{*(m)}$, $u_t^{(m)}$ is not necessarily positive since the first term of $u_t^{(m)}$ can be negative. We call $u_t^{(m)}$ "microstructure noise in NCRV" where we need to distinguish it from $\varepsilon(t)$. Otherwise, where there is no ambiguity, we simply call $u_t^{(m)}$ microstructure noise. This microstructure noise will be estimated in the later section.

In the Appendix, we show that

$$\begin{aligned} E\left[u_t^{(m)}\right] &= 2m\sigma_\varepsilon^2, \quad \text{and} \\ \mathrm{cov}\left[u_t^{(m)}, u_s^{(m)}\right] &= \begin{cases} 8\sigma_\varepsilon^2\sigma_0^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4 & t = s, \\ \omega_\varepsilon^2 & t = s \pm 1, \\ 0 & \text{otherwise}, \end{cases} \end{aligned} \tag{14}$$

where $\omega_\varepsilon^2 = \mathrm{var}[\varepsilon^2(t)] = E[\varepsilon^4(t)] - \sigma_\varepsilon^4$. Thus, $u_t^{(m)}$ has the autocovariance structure of a MA(1) process. Suppose that the MA(1) process is represented as

$$u_t^{(m)} = c_u^{(m)} + \xi_t^{(m)} + \theta_u^{(m)}\xi_{t-1}^{(m)}, \quad \xi_{t-1}^{(m)} \sim WN(0, \sigma_\xi^{2(m)}). \tag{15}$$

The mean and autocovariances of $u_t^{(m)}$ in terms of $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$ are

$$\begin{aligned} E\left[u_t^{(m)}\right] &= c_u^{(m)}, \quad \text{and} \\ \mathrm{cov}\left[u_t^{(m)}, u_s^{(m)}\right] &= \begin{cases} (1 + \theta_u^{2(m)})\sigma_\xi^{2(m)} & t = s, \\ \theta_u^{(m)}\sigma_\xi^{2(m)} & t = s \pm 1, \\ 0 & \text{otherwise}. \end{cases} \end{aligned} \tag{16}$$

Later, we utilize these two different expressions of the moments of $u_t^{(m)}$ to derive the relationships among parameters.

3 State Space Approach

Our approach is in the same spirit of the state space method used in Barndorff-Nielsen and Shephard (2002), where they consider the situation without microstructure noise. The existence of microstructure noises requires additional efforts for checking identification of state space model parameters. In this section, we describe the identification problem in estimating the IV from the NCRC by a state space method.

3.1 State space form of NCRV

Since $RV_t^{(m)} = IV_t + d_t^{(m)}$, we have

$$RV_t^{*(m)} = IV_t + d_t^{(m)} + u_t^{(m)}. \tag{17}$$

From (5), (15) and (17), we have the following state space form of $RV_t^{*(m)}$:

(Observation equation)

$$RV_t^{*(m)} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t \end{bmatrix} + d_t^{(m)}, \tag{18}$$

(State equation)

$$\begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} c_{IV} \\ c_u^{(m)} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \phi & 0 & \theta_{IV} & 0 \\ 0 & 0 & 0 & \theta_u^{(m)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_{t-1} \\ u_{t-1}^{(m)} \\ \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}, \tag{19}$$

where

$$\begin{bmatrix} d_t^{(m)} \\ \eta_t \\ \xi_t \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_d^{2(m)} & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\xi^{2(m)} \end{bmatrix} \right). \tag{20}$$

Given the values of ϕ, c_{IV}, θ_{IV}, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$, we can estimate IV_t and $u_t^{(m)}$ by applying the Kalman filtering to the state space form. One problem of the state space form is how to estimate those parameters. One may simply think that we can estimate them from the state space form by, e.g., quasi-maximum likelihood estimation under Gaussian noise assumption. This is; however, not always possible. It is known that in general the parameters of state space form are not necessarily identified (see, for example, Hamilton, 1994, p.388). More precisely, there may be infinitely many combinations of state space model parameters that give the same autocovariance structure. We have to confirm whether the model parameters are uniquely identified before parameter estimation. We consider this problem in the next section. In fact, we show that the above parameters in the state space form cannot be uniquely identified.

3.2 Identification of model parameters

Since $RV_t^{*(m)}$ is the sum of three components: IV_t (an ARMA(1, 1) process), $d_t^{(m)}$ (a white noise process) and $u_t^{(m)}$ (an MA(1) process), $RV_t^{*(m)}$ itself follows an ARMA(1, 2) process (see Granger and Morris, 1976) so that it is expressed as

$$(1-\phi L)RV_t^{*(m)} = c^{(m)} + (1+\delta_1^{(m)}L + \delta_2^{(m)}L^2)\tau_t, \quad \tau_t \sim WN(0, \sigma_\tau^{2(m)}). \tag{21}$$

Note that the AR coefficient ϕ is the same as that of IV_t in (5). The ARMA model of a state space representation is commonly referred to as a reduced form or ARMA reduced form. Note that the parameters of any ARMA reduced form are always identified (and hence can be estimated). From (5), (17) and (15), we have

$$\begin{aligned}(1-\phi L)RV_t^{*(m)} = \;& (1-\phi L)IV_t + (1-\phi L)d_t^{(m)} + (1-\phi L)u_t^{(m)} \\ = \;& (1-\phi)\sigma_0^2 + \eta_t + \theta_{IV}\eta_{t-1} + d_t^{(m)} - \phi d_{t-1}^{(m)} + \xi_t \\ & +(1-\phi)c_u^{(m)} + (\theta_u^{(m)} - \phi)\xi_{t-1} - \phi\theta_u^{(m)}\xi_{t-2}.\end{aligned} \tag{22}$$

These two expressions on the right-hand sides in (21) and (22) are of the same model and hence their mean and autocovariances must be identical. The autocovariances of the MA parts in (21) are given as $\gamma_0^{(m)} = (1+\delta_1^{(m)2} + \delta_2^{(m)2})\sigma_\tau^{2(m)}$, $\gamma_1^{(m)} = (\delta_1^{(m)} + \delta_1^{(m)}\delta_2^{(m)})\sigma_\tau^{2(m)}$, $\gamma_2^{(m)} = \delta_2^{(m)}\sigma_\tau^{2(m)}$, and $\gamma_j = 0 \quad \text{for} \quad j \geq 3$. It is shown in the Appendix that the autocovariance functions of the MA parts in (22) are

$$\gamma_0^{(m)} = (1+\theta_{IV}^2)\sigma_\eta^2 + (1+\phi^2)\sigma_d^{2(m)} + [1 + (\theta_u^{(m)} - \phi)^2 + \phi^2\theta_u^{(m)2}]\sigma_\xi^{2(m)}, \tag{23a}$$

$$\gamma_1^{(m)} = \theta_{IV}\sigma_\eta^2 - \phi\sigma_d^{2(m)} + (\theta_u^{(m)} - \phi - \phi\theta_u^{(m)2} + \phi^2\theta_u^{(m)})\sigma_\xi^{2(m)} \tag{23b}$$

$$\gamma_2^{(m)} = -\phi\theta_u^{(m)}\sigma_\xi^{2(m)}, \tag{23c}$$

and $\gamma_j = 0 \quad \text{for} \quad j \geq 3$. By equating the means of the MA parts in (21) and (22), we obtain

$$c^{(m)} = (1-\phi)\left(\sigma_0^2 + c_u^{(m)}\right). \tag{23d}$$

Given ARMA(1, 2) model parameters in (21), we can obtain $c^{(m)}$, ϕ, and $\gamma_j^{(m)}$, $j = 0, 1, 2$. Then, unknown parameters in the equations in (23a)$\sim$(23d) are θ_{IV}, σ_0^2, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$. Observe that the number of unknown parameters are more than the number of equations. Hence, we cannot uniquely identify these parameters from these equations. In other word, for a given ARMA(1, 2) model, there are infinitely many combinations of parameters θ_{IV}, σ_0^2, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ that give the same autocovariance structure as the ARMA(1, 2) model.

3.3 Estimation

We reduce the number of unknown parameters as follows. Equations (6) and (7) imply that θ_{IV}, σ_η^2 and σ_d^2 are functions of ϕ, σ_0^2, ω_0^2. Below we show that $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$ can also be expressed as functions of σ_0^2, ω_ε^2 and σ_ε^2.[1] Substituting these

[1] They also depend on m as the notation implies.

functions into (23a)$\sim$(23d), we can reduce the number of unknown parameters. To derive the relationships among those parameters, we utilized the expressions given in (14) and (16).

In views of (14) and (16), we obtain the following equations:

$$c_u^{(m)} = 2m\sigma_\varepsilon^2, \tag{24a}$$

$$(1+\theta_u^{2(m)})\sigma_\xi^{2(m)} = 8\sigma_0^2\sigma_\varepsilon^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4, \tag{24b}$$

$$\theta_u^{(m)}\sigma_\xi^{2(m)} = \omega_\varepsilon^2. \tag{24c}$$

Assuming the MA parameter satisfies the invertibility condition, i.e., $|\theta_u^{(m)}| < 1$, we can solve the equations (24a) $\sim$ (24c) for $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$ as

$$c_u^{(m)} = 2m\sigma_\varepsilon^2, \quad \sigma_\xi^{2(m)} = \frac{\omega_\varepsilon^2}{\theta_u^{(m)}} \quad \text{and} \quad \theta_u^{(m)} = A - \sqrt{A^2-1}, \tag{25}$$

where $A = 4\frac{\sigma_0^2\sigma_\varepsilon^2}{\omega_\varepsilon^2} + 2m - 1 + 2m\frac{\sigma_\varepsilon^4}{\omega_\varepsilon^2}$. The detail of the calculation is given in the Appendix. Note that $0 < \theta_u^{(m)} < 1$ since $A > 1$.

From (6), (7) and (25), we see that c_{IV}, θ_{IV}, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ can be expressed as functions of ϕ, σ_0^2, ω_0^2, σ_ε^2 and ω_ε^2. To emphasize this, we may write them as

$$\begin{aligned} &c_{IV} = c_{IV}(\phi, \sigma_0^2), \quad \theta_{IV} = \theta_{IV}(\phi), \quad \sigma_\eta^2 = \sigma_\eta^2(\phi, \omega_0^2), \quad c_u^{(m)} = c_u^{(m)}(\sigma_\varepsilon^2), \\ &\theta_u^{(m)} = \theta_u^{(m)}(\sigma_0^2, \sigma_\varepsilon^2),\ \sigma_d^{2(m)} = \sigma_d^{2(m)}(\phi, \sigma_0^2, \omega_0^2) \text{ and } \sigma_\xi^{2(m)} = \sigma_\xi^{2(m)}(\sigma_0^2, \sigma_\varepsilon^2, \omega_\varepsilon^2). \end{aligned} \tag{26}$$

Note that θ_{IV} is a function of only ϕ, hence can be assumed to be known (since ϕ is identified from the reduced form). Substituting the expressions in (26) into the equations in (23a)$\sim$(23d), eventually, we have four equations for the four unknown parameters σ_0^2, ω_0^2, ω_ε^2 and σ_ε^2. Hence, the order condition for identification is satisfied. However, this result only does not necessarily mean that for a given autocovariance structure, we can uniquely identify σ_0^2, ω_0^2, ω_ε^2 and σ_ε^2. For the uniqueness of the identification, we must check if the rank condition is also satisfied. To show the uniqueness of the identification, we explicitly derive the solutions for the parameters in terms of c, ϕ, $\gamma_j^{(m)}$ $j = 0, ..., 2$.

In the Appendix, we show that, given c, ϕ, $\gamma_j^{(m)}$ $j = 0, ..., 2$ and (26), the equations in (23a)$\sim$(23d) can be *uniquely*[2] solved for σ_0^2, ω_0^2, σ_ε^2 and ω_ε^2 as

$$\omega_\varepsilon^2 = -\frac{\gamma_2^{(m)}}{\phi}, \quad \omega_0^2 = \frac{(\log\phi)^2[\phi\gamma_0^{(m)} + (1+\phi^2)\gamma_1^{(m)} + \frac{1+\phi^4}{\phi}\gamma_2^{(m)}]}{(1-\phi)^3(1+\phi)}, \tag{27a}$$

$$\sigma_\varepsilon^2 = \sqrt{\frac{c^2}{2m^2(1-\phi)^2} - \frac{(2m-1)\gamma_2^{(m)}}{2m\phi} - \frac{\gamma_0^{(m)} - 2\omega_0^2 D - 2\gamma_2^{(m)}}{4m(1+\phi^2)}}, \tag{27b}$$

[2]More precisely, under the condition $\sigma_\varepsilon^2 > 0$.

and

$$\sigma_0^2 = \frac{c}{1-\phi} - 2m\sigma_\varepsilon^2, \tag{27c}$$

where $D = B + (1+\phi^2)C$,

$$B \equiv \frac{\phi^2 - 1 - (1+\phi^2)\log\phi}{(\log\phi)^2} \quad \text{and} \quad C \equiv \frac{2m\left(\phi^{\frac{1}{m}} - 1 - \log\phi^{\frac{1}{m}}\right)}{(\log\phi)^2}. \tag{27d}$$

The equations in (27a) $\sim$ (27d) show that the four parameters, σ_0^2, ω_0^2, σ_ε^2 and ω_ε^2 can be uniquely identified from the ARMA(1, 2) reduced form of the state space form in (18) and (19). It should be emphasized that this does not imply that we can directly estimate the state space model parameters in (18) and (19), but implies that we can estimate the above four parameters directly by replacing the state space model parameters with the functions of the four parameters. We estimate these four parameters by the quasi-maximum likelihood estimation of ARMA(1, 2) model, where innovations are (conveniently) assumed to be Gaussian.

Here, we summarize how to calculate the likelihood.

(Summary on how to calculate the likelihood)

(1) For a given m, calculate $RV_t^{*(m)}$.

(2) Given ϕ, σ_0^2, ω_0^2, σ_ε^2 and ω_ε^2, calculate c_{IV}, θ_{IV}, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ according to (6), (7) and (25).

(3) With c_{IV}, θ_{IV}, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ obtained in Step (2), compute $\gamma_j^{(m)}$, $j = 1, \ldots 3$ and $c^{(m)}$ in (23a)$\sim$(23d).

(4) Applying a result in Meddahi (2002), (representations of MA(2) parameters in terms of the autocovariance functions) calculate corresponding ARMA(1, 2) parameters, $c^{(m)}$, $\delta_1^{(m)}$, $\delta_2^{(m)}$ and $\sigma_\xi^{2(m)}$ from $\gamma_j^{(m)}$, $j = 1, \ldots 3$ obtained in Step (3).

(5) Calculate the Gaussian ARMA(1, 2) log-likelihood for RV_t^* with ϕ, $c^{(m)}$ $\delta_1^{(m)}$, $\delta_2^{(m)}$ and $\sigma_\xi^{2(m)}$.

We can obtain the estimates of σ_0^2, ω_0^2, σ_ε^2 and ω_ε^2 by maximizing the above log-likelihood with respect to the four parameters. The method provides consistent estimators for the four parameters. Using the proposed methodology, we conduct an empirical analysis for exchange rate data in the next section.

4 Empirical Analysis

4.1 Data description

The yen/dollar spot exchange rate series we use are the mid-quote prices observed every 1 minute, which are obtained from Olsen and Associates. The full sample covers the period from 1 January 2000 to December 31 2006. Figure 1 plots the daily returns calculated from the data over the period.

We interpolate missing prices by the previous tick method, i.e., interpolating the previous prices. Also, following Andersen, Bollerslev, Diebold and Labys (2001), we remove the data of inactive trading days. Whenever we did so, we always cut from 21:01 GMT on one night to 21:00 the next evening. For details on the motivation of this definition of "day", see Andersen, Bollerslev, Diebold and Labys (2001), Andersen and Bollerslev (1998) and Bollerslev and Domowitz (1993) . We cut the data according to the following criterions, which is similar to the criterions used in Beine *et al.* (2007):

(1) the days that miss more than 500 price observations,

(2) the days where, *in total*, there are more than 1000 minutes of zero returns

(3) the days where the price did not change for more than 35 minutes.

By these criterions, we can remove all weekend data. However, the days such as U.S. holidays that Andersen, Bollerslev, Diebold and Labys (2001) and Beine *et al.* (2007) removed are not necessarily cut by these criterions. This is because even when the U.S. market is closed, the transactions are made in other markets. Eventually, we are left with 1809 complete days, or $1809 \times 1440 = 2604960$ price observations from which we calculate the 1-min and 5-min returns.

With these returns, we calculate the series of daily NCRVs. We call the NCRVs 1-min NCRV ($m = 1440$) or 5-min NCRV ($m = 288$) depending on the intervals of the returns. Table 1 reports the descriptive statistics of 1-min and 5-min NCRVs. Figure 2 plots these two NCRVs. The mean of 1-min NCRV is higher than that of 5-min NCRV. This can be due to microstructure noises since the mean of NCRV increases as the sampling frequency tends to be high, or $m \to \infty$, as shown in (23d) and (24a). The first order autocorrelation of NCRVs are somewhat lower than usually expected for variances of financial time series (0.4794 for 1-min NCRV and 0.4177 for 5-min RV).

4.2 Estimation of parameters, IV and microstructure noise

For these two series of NCRVs, we estimate parameters, ϕ, σ_0^2, ω_0^2, σ_ε^2 and ω_ε^2, as described in the previous section. Note that, in general, the values of 1-min and 5 min NCRVs are different and thus the estimates from these two NCRVs are different. Table 2 shows the estimates of the above parameters and the values of state space form parameters in (18) and (19) computed from the estimates. The notation "(m)" implies that those values depend on the value of m. Notice that even though we used the series of NCRVs with different m's for the estimations, the estimated values of parameters that do not depend on m are very similar. For example, the estimated value of σ_0^2, the unconditional mean of the IV_t, is 0.3581 for $m = 1440$ and 0.3781 for $m = 288$. The estimated value of ω^2, the unconditional variance of IV_t, is 0.0301 for for $m = 1440$ and 0.0279 for $m = 288$. The estimates of the first order autocorrelation of IV are significantly higher than those of NCRVs. This suggests that the observed low autocorrelations of the NCRVs are due to microstructure noises. One interesting observation is that although we do not assume any distribution for the microstructure noise $\varepsilon(t)$, the kurtosis of $\varepsilon(t)$, calculated from the estimated variances of $\sigma_\varepsilon^2 = \mathrm{var}[\varepsilon(t)]$ and $\omega_\varepsilon^2 = \mathrm{var}[\varepsilon(t)^2]$, are close to 3 in the case of 5-min NCRV.

We display the estimates of the IV by Kalman smoothing in Figure 3(a) for 5-min NCRV, 3(b) for 1-min NCRV and 3(c) for the both. Note that both are the estimate of the same IV series. These two series of IV estimates are very similar as seen in Figure 3(c). Figure 4 plots the smoothed estimates of microstructure noise $u_t^{(m)}$'s. The results are natural. We find that 1-min NCRV has a larger bias than 5-min NCRV. The mean of the microstructure noises for 1-min NCRV is higher than that for 5-min NCRV. The small biases of 5-min NCRV is consistent with the use of 5-min NCRV in the previous literature.

Last, we calculate the ratio of microstructure noises to the NCRVs, i.e., $R(m) = \widehat{u}_t^{(m)}/RV_t^{*(m)}$ for each m, $t = 1, ..., 1809$, where $\widehat{u}_t^{(m)}$ is the smoothed estimates of $u_t^{(m)}$. The maximum and minimum values are, respectively, 0.6970 and -4.2179 for $R(288)$(5 min NCRV) and 0.7832 and -0.7201 for $R(1440)$(1 min NCRV). We also calculate the average magnitude of microstructure noise $u_t(m)$'s as the mean of $R(m)$ and $|R(m)|$. The values of the means of $R(m)$ and $|R(m)|$ are, respectively, -0.0458 and 0.2341 for $m = 288$, and 0.3073 and 0.3273 for $m = 1440$.

5 Summary and Concluding Remarks

In this paper, we proposed a new approach, using a state space form of the realized variance, to estimating the integrated variance. Our method is based on the result in Meddahi (2001), which shows that when the prices follows a certain class of stochastic volatility models, the integrated variance follows a ARMA(1, 1) model. We showed that under the existence of microstructure noises, the observed realized variance follows a ARMA(1 ,2) model. We represented the ARMA(1, 2) model by a state space form and established the uniqueness of the identification of state space form parameters. The proposed method is applied to yen/dollar exchange rate data, where we find that the RV calculated with 5-min returns are less biased than with 1-min returns. The two series of IV estimates obtained by the proposed method with 1-min returns and 5-min returns are very similar. The method was also used for estimating microstructure noises.

In the estimation, we constructed the likelihood using only either 1-min or 5-min RV. It is more desirable to use both RVs for estimating common parameters. Investigating the optimal way for combining RVs of different scales is a subject of future research. It is also interesting to apply our method to stock return data and examine the effects of microstructure noises.

Appendix

Hereafter, we suppress "(m)" in the notations $r_t^{(m)}$, $u_t^{(m)}$ and $e_t^{(m)}$, for simplicity.

Derivation of (14)

First, we derive $E[u_t]$. Since r_t and e_t are independent by Assumption 1 and $\mathrm{var}(e_t) = 2\sigma_\varepsilon^2$, we have

$$\begin{aligned} E[u_t] = & \; 2\sum_{i=1}^{m} E\left[r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}\right] + \sum_{i=1}^{m} E\left[e^2_{t-1+\frac{i}{m}}\right] \\ = & \; \sum_{i=1}^{m} E\left[r_{t-1+\frac{i}{m}}\right] E\left[e_{t-1+\frac{i}{m}}\right] + m\mathrm{var}(e_t) \\ = & \; 2m\sigma_\varepsilon^2. \end{aligned} \tag{28}$$

The following results are used for deriving $\mathrm{var}[u_t]$ in (34) and $\mathrm{cov}[u_t, u_{t-1}]$ in (35) below: For any s and t, we have

$$\begin{aligned} \mathrm{cov}\left[r_s e_s, r_t e_t\right] = & \; E\left[r_s e_s r_t e_t\right] - E\left[r_s e_s\right] E\left[r_t e_t\right] \\ = & \; E\left[e_s e_t\right] E\left[r_s r_t\right] - E\left[r_s\right] E\left[e_s\right] E\left[r_t\right] E\left[e_t\right]. \end{aligned} \tag{29}$$

Thus, when $t = s$ we have

$$\begin{aligned} \mathrm{cov}\left[r_s e_s, r_t e_t\right] = & \; 2\sigma_\varepsilon^2 E\left[r_t^2\right] \\ = & \; 2\sigma_\varepsilon^2 E[(\textstyle\int_{t-1/m}^{t} \sigma(s) dW(s))^2] \\ = & \; 2\sigma_\varepsilon^2 E[\textstyle\int_{t-1/m}^{t} \sigma^2(s) ds] \\ = & \; 2\sigma_\varepsilon^2 E\left[\textstyle\int_{t-\frac{1}{m}}^{t} \sigma^2(s) ds\right] \\ = & \; \tfrac{2\sigma_\varepsilon^2 \sigma_0^2}{m}. \end{aligned} \tag{30a}$$

The third equality comes from the Ito isometry. When $t \neq s$, we have

$$\mathrm{cov}\left[r_s e_s, r_t e_t\right] = E\left[e_s e_t\right] E\left[r_s\right]\left[r_t\right] - E\left[r_s\right] E\left[e_s\right] E\left[r_t\right] E\left[e_t\right] = 0. \tag{30b}$$

Next, we derive $\mathrm{cov}\left[e_t^2, e_s^2\right]$. When $t = s$, we have

$$\begin{aligned} \mathrm{cov}[e_t^2, e_t^2] = & \; \mathrm{var}[e_t^2] \\ = & \; E[e_t^4] - (E[e_t^2])^2 \\ = & \; E\left[\varepsilon_t^4 - 4\varepsilon_t^3 \varepsilon_{t-\frac{1}{m}} + 6\varepsilon_t^2 \varepsilon_{t-\frac{1}{m}}^2 - 4\varepsilon_t \varepsilon_{t-\frac{1}{m}}^3 + \varepsilon_{t-\frac{1}{m}}^4\right] - 4\sigma_\varepsilon^4 \\ = & \; 2E[\varepsilon_t^4] + 2\sigma_\varepsilon^4. \end{aligned} \tag{31}$$

When $t = s \pm \frac{1}{m}$, we have

$$\begin{aligned} \mathrm{cov}\left[e_s^2, e_{s-\frac{1}{m}}^2\right] = & \; \mathrm{cov}\left[e_{s+\frac{1}{m}}^2, e_s^2\right] \\ = & \; \mathrm{cov}\left[\varepsilon_{s+\frac{1}{m}}^2 - 2\varepsilon_{s+\frac{1}{m}}\varepsilon_s + \varepsilon_s^2, \varepsilon_s^2 - 2\varepsilon_s \varepsilon_{s-\frac{1}{m}} + \varepsilon_{s-\frac{1}{m}}^2\right] \\ = & \; \mathrm{var}[\varepsilon_s^2] \\ = & \; \omega_\varepsilon^2. \end{aligned} \tag{32}$$

When $t = s \pm \frac{i}{m}$ for $i \geq 2$, we have $\text{cov}[e_t, e_s] = 0$. We have $\text{cov}[r_t e_t, e_s^2] = 0$ for any t and s since

$$\begin{aligned}
\text{cov}[r_t e_t, e_s^2] = & \; E[r_t e_t e_s^2] - E[r_t e_t]E[e_s^2] \\
= & \; E[r_t]E[e_t e_s^2] - E[r_t]E[e_t]E[e_s^2] \\
= & \; 0.
\end{aligned} \tag{33}$$

From (30) $\sim$ (33) we have.

$$\begin{aligned}
\text{var}\,[u_t] = & \; \text{var}\left[2\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}} + \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
= & \; 4\text{var}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}\right] + \text{var}\left[\sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
& +4\text{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
= & \; 4\text{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}\right] + \text{cov}\left[\sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
& +4\text{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
= & \; 4\sum_{i=1}^{m}\sum_{j=1}^{m} \text{cov}\left[r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, r_{t-1+\frac{j}{m}} e_{t-1+\frac{j}{m}}\right] + \sum_{i=1}^{m}\sum_{j=1}^{m} \text{cov}\left[e^2_{t-1+\frac{i}{m}}, e^2_{t-1+\frac{j}{m}}\right] \\
& +4\sum_{i=1}^{m}\sum_{j=1}^{m} \text{cov}\left[r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, e^2_{t-1+\frac{i}{m}}\right] \\
= & \; 8\sigma_\varepsilon^2\sigma_0^2 + 2m(E[\varepsilon_t^4] + \sigma_\varepsilon^4) + 2(m-1)(E[\varepsilon_t^4] - \sigma_\varepsilon^4) \\
= & \; 8\sigma_\varepsilon^2\sigma_0^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4,
\end{aligned} \tag{34}$$

and

$$\begin{aligned}
\text{cov}[u_t, u_{t+1}] = & \; \text{cov}\left[2\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}} + \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}, 2\sum_{i=1}^{m} r_{t+\frac{i}{m}} e_{t+\frac{i}{m}} + \sum_{i=1}^{m} e^2_{t+\frac{i}{m}}\right] \\
= & \; 4\text{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} r_{t+\frac{i}{m}} e_{t+\frac{i}{m}}\right] \\
& +2\text{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t+\frac{i}{m}}\right] \\
& +2\text{cov}\left[\sum_{i=1}^{m} r_{t+\frac{i}{m}} e_{t+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] + \text{cov}\left[\sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t+\frac{i}{m}}\right] \\
= & \; \text{cov}\left[e_t^2, e^2_{t+\frac{1}{m}}\right] \\
= & \; \omega_\varepsilon^2.
\end{aligned} \tag{35}$$

It is easy to check $\text{cov}[u_t, u_{t\pm i}] = 0$ for $i \geq 2$, and hence we have (14).

Derivation of (23a)∼(23c)

$$\begin{aligned}
\gamma_0 = \;& \mathrm{cov}\{\eta_t + \theta_{IV}\eta_{t-1} + d_t - \phi d_{t-1} + \xi_t + (\theta_u - \phi)\xi_{t-1} - \phi\theta_u\xi_{t-2}, \\
& \eta_t + \theta_{IV}\eta_{t-1} + d_t - \phi d_{t-1} + \xi_t + (\theta_u - \phi)\xi_{t-1} - \phi\theta_u\xi_{t-2}\} \\
= \;& \sigma_\eta^2 + \theta_{IV}^2\sigma_\eta^2 + \sigma_d^2 + \phi^2\sigma_d^2 + \sigma_\xi^2 + (\theta_u - \phi)^2\sigma_\xi^2 + \phi^2\theta_u^2\sigma_\xi^2 \\
= \;& (1 + \theta_{IV}^2)\sigma_\eta^2 + (1 + \phi^2)\sigma_d^2 + [1 + (\theta_u - \phi)^2 + \phi^2\theta_u^2]\sigma_\xi^2,
\end{aligned}$$

$$\begin{aligned}
\gamma_1 = \;& \mathrm{cov}\{\eta_t + \theta_{IV}\eta_{t-1} + d_t - \phi d_{t-1} + \xi_t + (\theta_u - \phi)\xi_{t-1} - \phi\theta_u\xi_{t-2}, \\
& \eta_{t-1} + \theta_{IV}\eta_{t-2} + d_{t-1} - \phi d_{t-2} + \xi_{t-1} + (\theta_u - \phi)\xi_{t-2} - \phi\theta_u\xi_{t-3}\} \\
= \;& \theta_{IV}\sigma_\eta^2 - \phi\sigma_d^2 + (\theta_u - \phi - \phi\theta_u^2 + \phi^2\theta_u)\sigma_\xi^2
\end{aligned}$$

$$\begin{aligned}
\gamma_2 = \;& \mathrm{cov}\{\eta_t + \theta_{IV}\eta_{t-1} + d_t - \phi d_{t-1} + \xi_t + (\theta_u - \phi)\xi_{t-1} - \phi\theta_u\xi_{t-2}, \\
& \eta_{t-2} + \theta_{IV}\eta_{t-3} + d_{t-2} - \phi d_{t-3} + \xi_{t-2} + (\theta_u - \phi)\xi_{t-3} - \phi\theta_u\xi_{t-4}\} \\
= \;& -\phi\theta_u\sigma_\xi^2.
\end{aligned}$$

Derivation of (25)

From (24c), we have $\sigma_\xi^2 = \omega_\varepsilon^2/\theta_u$. Substituting this into (24b), we have

$$\begin{aligned}
& (1 + \theta_u^2)\tfrac{\omega_\varepsilon^2}{\theta_u} = 8\sigma_0^2\sigma_\varepsilon^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4 \\
\Leftrightarrow\; & \omega_\varepsilon^2\theta_u^2 - [8\sigma_0^2\sigma_\varepsilon^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4]\theta_u + \omega_\varepsilon^2 = 0 \\
\Leftrightarrow\; & \theta_u^2 - 2[4\tfrac{\sigma_0^2\sigma_\varepsilon^2}{\omega_\varepsilon^2} + 2m - 1 + 2m\tfrac{\sigma_\varepsilon^4}{\omega_\varepsilon^2}]\theta_u + 1 = 0 \\
\Leftrightarrow\; & \theta_u = A \pm \sqrt{A^2 - 1},
\end{aligned} \tag{36}$$

where $A = 4\frac{\sigma_0^2\sigma_\varepsilon}{\omega_\varepsilon^2} + 2m - 1 + 2m\frac{\sigma_\varepsilon^4}{\omega_\varepsilon^2}$. Since $A > 1$ for $m \geq 1$, we have $A + \sqrt{A^2 - 1} > 1$. Assuming that θ_u satisfies the invertibility condition, we obtain θ_u in (25).

Derivation of (27a) and (27c)

From (23c) and (24c), we have $\omega_\varepsilon^2 = -\frac{\gamma_2}{\phi}$. From (6) and (7), we have

$$\sigma_\eta^2 = \frac{2\omega_0^2}{1 + \theta_{IV}^2}B \quad \text{and} \quad \sigma_d^2 = \frac{2\sigma_0^4}{m} + 2\omega_0^2 C, \tag{37}$$

where B and C are given in (27d). From $\omega_\varepsilon^2 = \theta_u\sigma_\xi^2$ in (24c), we have

$$\begin{aligned}
(1 + \theta_u^2 - 2\theta_u\phi + \phi^2 + \phi^2\theta_u^2)\sigma_\xi^2 = \;& \left(\tfrac{1}{\theta_u} + \theta_u - 2\phi + \tfrac{\phi^2}{\theta_u} + \phi^2\theta_u\right)\omega_\varepsilon^2 \\
= \;& \left[\left(\tfrac{1}{\theta_u} + \theta_u\right)(1 + \phi^2) - 2\phi\right]\omega_\varepsilon^2,
\end{aligned} \tag{38}$$

and

$$\begin{aligned}
(\theta_u - \phi - \phi\theta_u^2 + \phi^2\theta_u)\sigma_\xi^2 = \;& \left(1 - \tfrac{\phi}{\theta_u} - \phi\theta_u + \phi^2\right)\omega_\varepsilon^2 \\
= \;& \left[1 + \phi^2 - \left(\tfrac{1}{\theta_u} + \theta_u\right)\phi\right]\omega_\varepsilon^2.
\end{aligned} \tag{39}$$

Substituting (37), (38) and (39) into (23a) and (23b), we have

$$\gamma_0 = 2\omega_0^2 D + 2\sigma_0^4\frac{1 + \phi^2}{m} + \left[\left(\frac{1}{\theta_u} + \theta_u\right)(1 + \phi^2) - 2\phi\right]\omega_\varepsilon^2, \tag{40a}$$

and

$$\gamma_1 = 2\omega_0^2 E - 2\sigma_0^4 \frac{\phi}{m} - \left[\left(\frac{1}{\theta_u} + \theta_u \right) \phi - (1+\phi^2) \right] \omega_\varepsilon^2, \tag{40b}$$

where $D = B + (1+\phi^2)C$ and $E = \frac{\theta_{IV}}{1+\theta_{IV}^2} B - \phi C$. From (6), we have

$$\begin{aligned} \frac{\theta_{IV}}{1+\theta_{IV}^2} &= \frac{2\rho(1-\sqrt{1-4\rho^2})}{4\rho^2 + (1-\sqrt{1-4\rho^2})^2} \\ &= \frac{2\rho(1-\sqrt{1-4\rho^2})(1+\sqrt{1-4\rho^2})}{4\rho^2(1+\sqrt{1-4\rho^2}) + (1-\sqrt{1-4\rho^2})^2(1+\sqrt{1-4\rho^2})} \\ &= \rho, \end{aligned} \tag{41}$$

and

$$\rho B = \frac{-2\phi(\phi - \log\phi - 1) + (1-\phi)^2}{2(\log\phi)^2} \tag{42}$$

Hence, we have

$$\begin{aligned} \phi\gamma_0 + (1+\phi^2)\gamma_1 &= 2\left[\phi D + (1+\phi^2)E\right]\omega_0^2 + \left[(1+\phi^2)^2 - 2\phi^2\right]\omega_\varepsilon^2, \\ &= 2\left[\phi + (1+\phi^2)\rho\right] B\omega_0^2 + (1+\phi^4)\omega_\varepsilon^2, \\ &= \frac{(1-\phi)^3(1+\phi)}{(\log\phi)^2}\omega_0^2 + (1+\phi^4)\omega_\varepsilon^2, \end{aligned} \tag{43}$$

or

$$\omega_0^2 = \frac{(\log\phi)^2[\phi\gamma_0 + (1+\phi^2)\gamma_1 - (1+\phi^4)\omega_\varepsilon^2]}{(1-\phi)^3(1+\phi)}. \tag{44}$$

Substituting $\omega_\varepsilon^2 = -\frac{\gamma_2}{\phi}$, we obtain (27a). Next, note that from (25), we have

$$\begin{aligned} \frac{1}{\theta_u} + \theta_u &= \frac{1+\theta_u^2}{\theta_u} \\ &= \frac{1+(A-\sqrt{A^2-1})^2}{A-\sqrt{A^2-1}} \\ &= \frac{A+\sqrt{A^2-1}+(A-\sqrt{A^2-1})^2(A+\sqrt{A^2-1})}{(A-\sqrt{A^2-1})(A+\sqrt{A^2-1})} \\ &= 2A. \end{aligned} \tag{45}$$

From (23d) and (24a), we have

$$c = (1-\phi)\left(\sigma_0^2 + 2m\sigma_\varepsilon^2\right) \quad \text{or} \quad \sigma_\varepsilon^2 = \frac{c-(1-\phi)\sigma_0^2}{2(1-\phi)m}. \tag{46}$$

Substituting σ_ε^2 in (46) into A, we have

$$\begin{aligned} 2A &= 2\left[4\sigma_0^2 \left(\tfrac{c-(1-\phi)\sigma_0^2}{2(1-\phi)m}\right)\tfrac{1}{\omega_\varepsilon^2} + 2m - 1 + 2m\left(\tfrac{c-(1-\phi)\sigma_0^2}{2(1-\phi)m}\right)^2 \tfrac{1}{\omega_\varepsilon^2}\right] \\ &= 2\left[\tfrac{2c\sigma_0^2}{(1-\phi)m\omega_\varepsilon^2} - \tfrac{2\sigma_0^4}{m\omega_\varepsilon^2} + 2m - 1 + \left(\tfrac{c^2-2(1-\phi)c\sigma_0^2+(1-\phi)^2\sigma_0^4}{2(1-\phi)^2 m\omega_\varepsilon^2}\right)\right] \\ &= \tfrac{4c\sigma_0^2}{(1-\phi)m\omega_\varepsilon^2} - \tfrac{4\sigma_0^4}{m\omega_\varepsilon^2} + 2(2m-1) + \tfrac{c^2}{(1-\phi)^2 m\omega_\varepsilon^2} - \tfrac{2c\sigma_0^2}{(1-\phi)m\omega_\varepsilon^2} + \tfrac{\sigma_0^4}{m\omega_\varepsilon^2} \\ &= \tfrac{2c\sigma_0^2}{(1-\phi)m\omega_\varepsilon^2} - \tfrac{3\sigma_0^4}{m\omega_\varepsilon^2} + 2(2m-1) + \tfrac{c^2}{(1-\phi)^2 m\omega_\varepsilon^2}. \end{aligned} \tag{47}$$

From (40), (45) and (47), we have

$$\gamma_0 = 2\omega_0^2 D - \frac{(1+\phi^2)}{m}\sigma_0^4 + \frac{2(1+\phi^2)c}{(1-\phi)m}\sigma_0^2 + 2(2m-1)(1+\phi^2)\omega_\varepsilon^2 + \frac{(1+\phi^2)c^2}{(1-\phi)^2 m} - 2\phi\omega_\varepsilon^2. \tag{48}$$

Multiplying both sides in (48) by $m/(1+\phi^2)$ and arranging, we have

$$\sigma_0^4 - \frac{2c}{1-\phi}\sigma_0^2 - \frac{c^2}{(1-\phi)^2} + \frac{m\left(\gamma_0 - 2\omega_0^2 D + 2\phi\omega_\varepsilon^2\right)}{1+\phi^2} - 2m(2m-1)\omega_\varepsilon^2 = 0. \tag{49}$$

Solving this quadratic equation for σ_0^2, we have

$$\sigma_0^2 = \frac{c}{1-\phi} \pm \sqrt{\frac{2c^2}{(1-\phi)^2} + 2m(2m-1)\omega_\varepsilon^2 - \frac{m\left(\gamma_0 - 2\omega_0^2 D + 2\phi\omega_\varepsilon^2\right)}{(1+\phi^2)}}. \tag{50}$$

From $\sigma_\varepsilon^2 > 0$ and (46), we must have $\frac{c}{1-\phi} > \sigma_0^2$. Hence, the sign of the second term in (50) is negative. From (46), we have

$$\sigma_\varepsilon^2 = \frac{1}{2m}\sqrt{\frac{2c^2}{(1-\phi)^2} + 2m(2m-1)\omega_\varepsilon^2 - \frac{m\left(\gamma_0 - 2\omega_0^2 D + 2\phi\omega_\varepsilon^2\right)}{(1+\phi^2)}}. \tag{51}$$

From the above results, we obtain (27b) and (27c).

References

Andersen, T. G., T. Bollerslev, (1998) Deutche Mark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies, *Journal of Finance*, 1. 219–265

Andersen, T. G., T. Bollerslev, F. X. Diebold and H. Ebens (2001) The distribution of realised stochk return volatility, *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2001) The distribution of exchange rate volatility, *Journal of the American Statistical Association*, 92, 42–55.

Barndorff-Nielsen, O.E. and N. Shephard (2001), Non-Gaussian OU based models and some of their uses in financial economics, *Journal of the Royal Statistical Society, B* 63, 167–241.

Barndorff-Nielsen, O.E. and N. Shephard (2002), Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of Royal Statistical Association*, Series B, 64, Part 2, 253–280.

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

Bandi, F. M. and J. R. Russell (2008), Microstructure noise, realized variance, and optimal sampling *Review of Economic Studies*, 75, 339–369.

Beine, M., J. Lahaye, S. Laurent, C.J. Neely, and F.C. Palm (2007) Central bank intervention and exchange rate volatility, its continuous and jump components, *International journal of finance and economics*, 12, 201–223.

Bollerslev, T., and I. Domowitz (1993), Trading Patterns and Prices in the Interbank Foreign Exchange Market, *Journal of Finance*, 48, 1421-1443.

Granger, W. J. C. and M. J. Morris (1976) Time Series Modelling And Interpretation, *Journal of the Royal Statistical Society. Series A (General)* Vol. 139, 246–257.

Ghysels, E. A. C. Harvey, and E. Renault (1996), Stochastic Volatility, 119-192, in *handbook of statistics* 14, edited by G.S. Maddala and C.R. Rao.

Hamilton DJ. 1994. *Time series analysis.* Princeton University Press: Princeton, NJ.

Hansen, P. R. and A. Lunde (2006) Realized variance and market microstructure noise, *Journal of Business & Economic Statistics*, 24, 127–161.

Meddahi, N. (2001), An eigenfunction approach for volatility modeling, CIRANO Working Paper, 2001s–70.

Meddahi, N. (2002), ARMA representation of two-factor models. CIRANO working paper, 2002s–92.

Meddahi, N. (2003), ARMA representation of integrated and realized variances, *Econometrics Journal*, 6, 335–356.

Nelson, D.B. (1990), ARCH models as diffusion approximations, *Journal of Econometrics* 45, 7–39.

Shephard, N. (2005), *Stochastic Volatility, Selected Readings.*

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

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

Table 1: Descriptive statistics of the NCRV

	1-min NCRV	5-min NCRV
m	1440	288
Mean	0.5317	0.4039
Variance	0.0629	0.0620
S.D.	0.2507	0.2490
SACORR(1)	0.4794	0.4177
SACORR(2)	0.3628	0.3292
SACORR(3)	0.3261	0.2819
SACORR(4)	0.3294	0.2595
SACORR(5)	0.3246	0.2577

Note: the table reports the mean, variance and standard deviation of RV calculated with different m. SACORR(k) denotes the sample autocorrelation of order k.

Table 2: Estimates of parameters

	1-min NCRV	5-min NCRV
$\widehat{\phi}$	0.9301	0.8849
	(0.0516)	(0.0329)
$\widehat{\sigma}_0^2$	0.3581	0.3781
	(0.0895)	(0.2155)
$\widehat{\omega}_0^2$	0.0301	0.0279
	(0.0084)	(0.0324)
$\widehat{\sigma}_\varepsilon^2$	6.0915×10^{-5}	4.5457×10^{-5}
	(5.4324×10^{-5})	(1.9086×10^{-4})
$\widehat{\omega}_\varepsilon^2$	5.8662×10^{-6}	2.9568×10^{-5}
	(9.6479×10^{-7})	(7.5890×10^{-6})
L	240.0671	181.187
$\mathrm{corr}(IV_t, IV_{t-1})$	0.9531	0.9225
$\widehat{c}_{IV}$	0.0372	0.0435
$\widehat{\theta}_{IV}$	0.2679	0.2677
$\widehat{\sigma}_\eta^2$	0.0025	0.0038
m	1440	288
$\widehat{c}_u^{(m)}$	0.1754	0.0262
$\widehat{\theta}_u^{(m)}$	1.7267×10^{-4}	8.6660×10^{-4}
$\widehat{\sigma}_\xi^{2(m)}$	0.0340	0.0341
$\widehat{\sigma}_d^{2(m)}$	2.199×10^{-4}	0.0012
$\dfrac{\mathrm{var}[IV_t]}{\mathrm{var}[IV_t] + \mathrm{var}[u_t^{(m)}] + \mathrm{var}[d_t^{(m)}]}$	0.4618	0.4313
$\dfrac{\mathrm{var}[u_t^{(m)}]}{\mathrm{var}[IV_t] + \mathrm{var}[u_t^{(m)}] + \mathrm{var}[d_t^{(m)}]}$	0.5348	0.5496
$\dfrac{\widehat{\sigma}_\eta^2}{\widehat{\sigma}_\eta^2 + \widehat{\sigma}_\xi^2}$	0.0690	0.0992
$\dfrac{\widehat{\sigma}_\xi^2}{\widehat{\sigma}_\eta^2}$	0.0742	0.1101
$E[\varepsilon_t^4]/\sigma_\varepsilon^4$	1.9239	3.0907

Note: the robust standard errors are inside the parenthesis. L is the log-likelihood.

Figure 1 : Daily returns of yen/dollar exchange rate

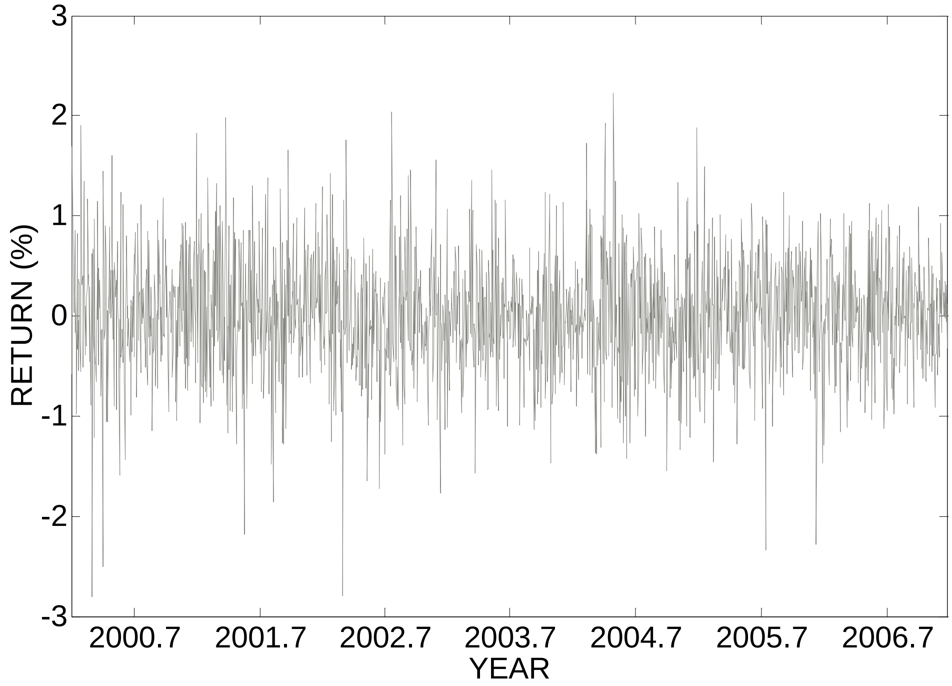


Figure 2: 1-min NCRV and 5 min NCRV

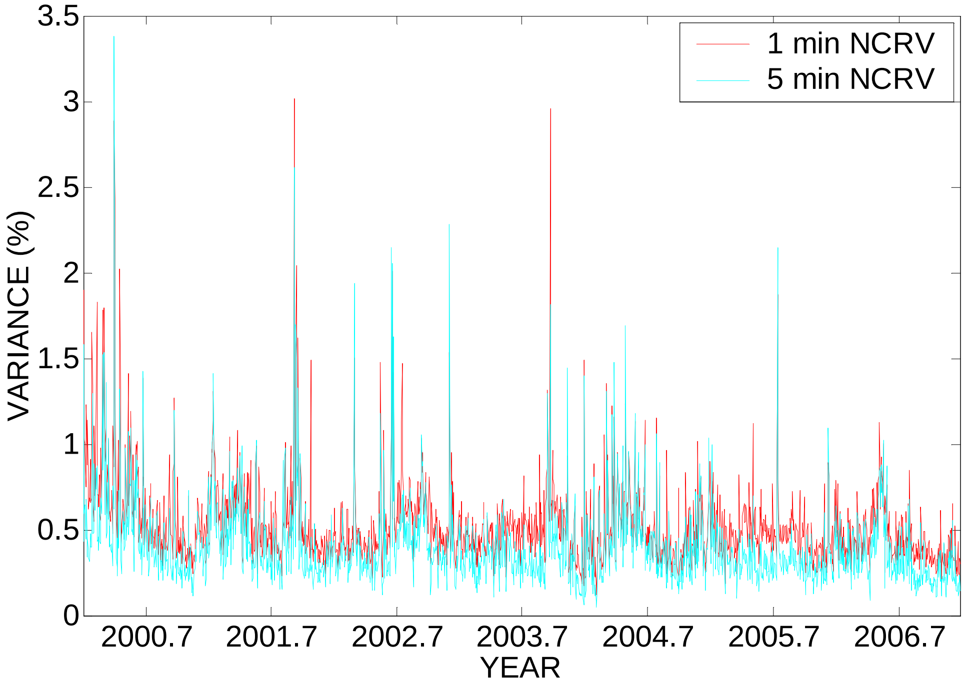
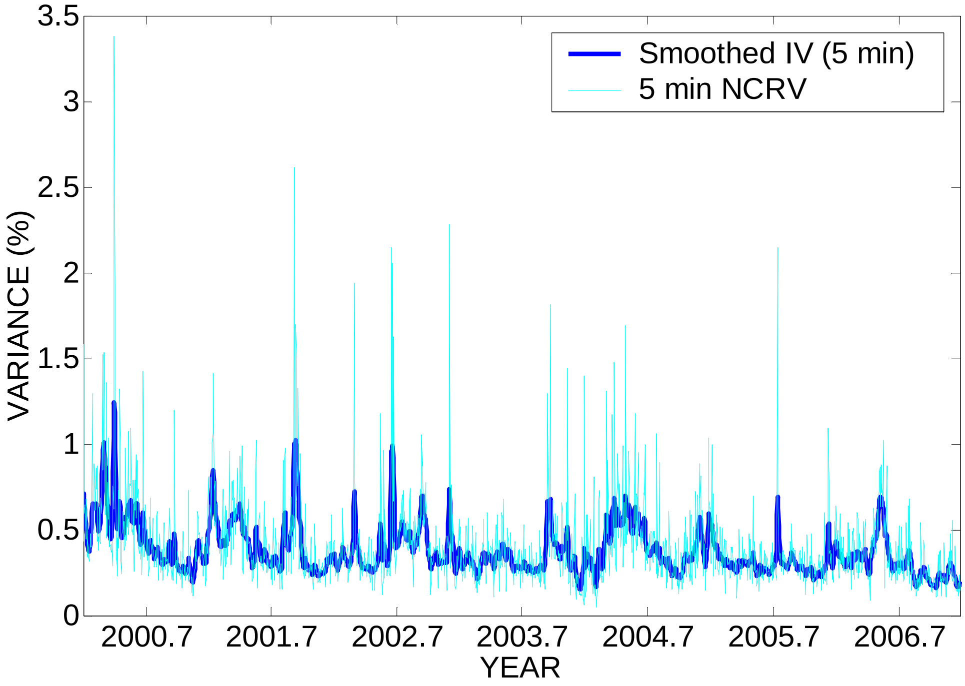
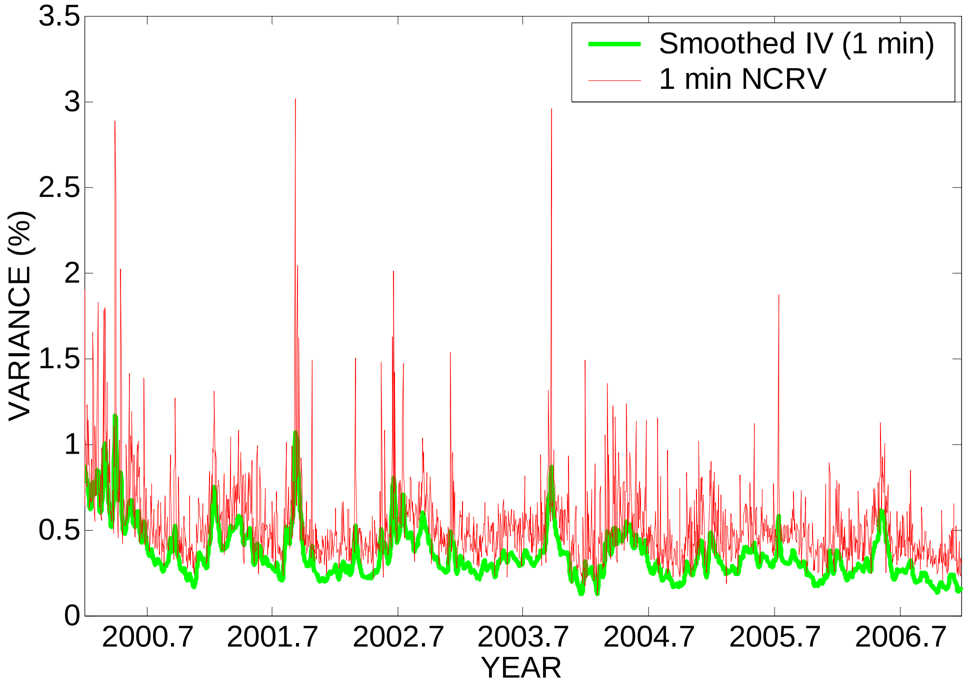


Figure 3: Smoothed IV estimates for 1 min and 5 min NCRVs

(a) 5 min



(b) 1 min



(c) 1 min and 5 min

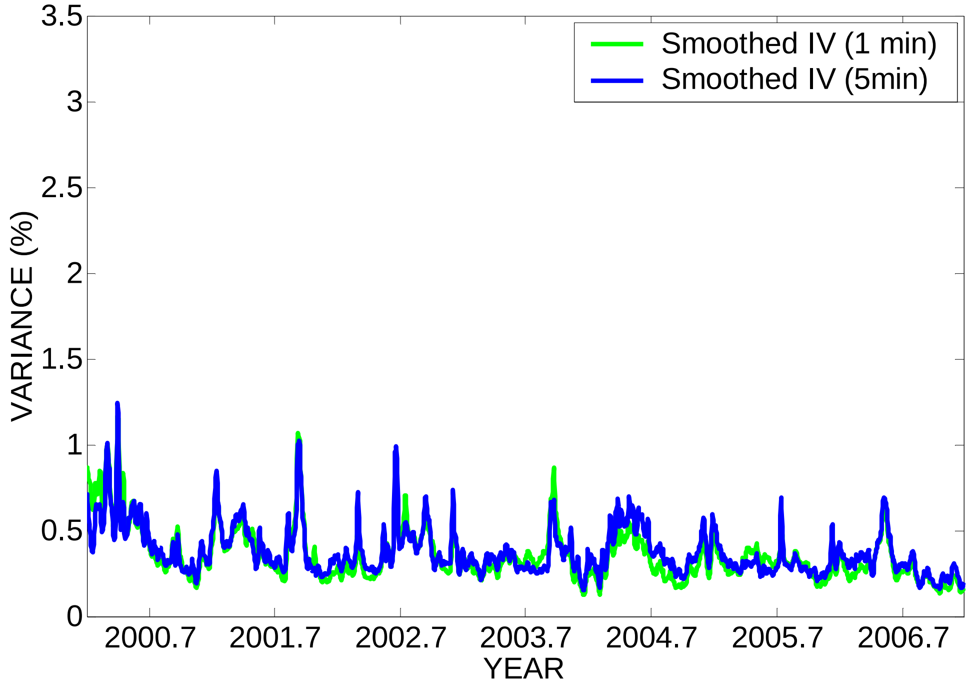


Figure 4: Estimate of microstructure noise for 1 min and 5 min NCRVs