Hitotsubashi University

Global COE Hi-Stat Discussion Paper Series 055

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

A State Space Approach to Estimating the Integrated Variance and Microstructure Noise Component

Daisuke Nagakura
Toshiaki Watanabe

April 2009

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

Hi-Stat Discussion Paper

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

Daisuke Nagakura † and Toshiaki Watanabe ‡

SUMMARY

We call the realized variance (RV) calculated with observed prices contaminated by microstructure noises (MNs) the *noise-contaminated RV* (NCRV) and refer to the component in the NCRV associated with the MNs as the *MN component*. This paper develops a state space method for estimating the integrated variance (IV) and MN component simultaneously. We represent the NCRV by a state space form and show that the state space form parameters are not identifiable; however, they can be expressed as functions of fewer identifiable parameters. We illustrate how to estimate these parameters. The proposed method is applied to yen/dollar exchange rate data.

Key Words: Realized Variance; Integrated Variance; Microstructure Noise; State Space; Identification; Exchange Rate

*The authors would like to thank Torben Andersen, Siem Jan Koopman, Peter Reinhard Hansen, David Hendry, Koichi Maekawa, Yasuhiro Omori, Yushi Yoshida and the participants in International Conference "High Frequency Data Analysis in Financial Markets", International Conference on Econometrics and World Economy, Conference "Recent Development in Finance and Econometrics" and Seminar at Hiroshima University of Economics for their useful comments. Financial support from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government through Grants-in-Aid for Scientific Research (No. 18203901) and the Global COE program "Research Unit for Statistical and Empirical Analysis in Social Sciences" at Hitotsubashi University is gratefully acknowledged. Views expressed in this paper are those of the 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 is known to change over time. More specifically, the variance, or the square root of the variance (volatility), tends to be large (small) following successive large (small) variances in previous periods. This phenomenon is known as "volatility clustering". A huge number of researchers have tried to estimate these changing variances because their values are crucially important for option pricing, risk management, optimal portfolio construction, etc. There are two popular classes of models for this sort of volatility dynamics, namely, generalized autoregressive conditional heteroskedastic (GARCH) models and stochastic volatility (SV) models. Based on GARCH or SV models with estimated model parameters, one can estimate the changing variances. See, for example, Bollerslev *et al.* (1994), Palm (1996) and Zivot (2008) for comprehensive surveys on GARCH models, Ghysels *et al.* (1996) for a review of some of the older papers on SV models and Shephard (2005) for a list of selected papers in the SV model literature.

Our main objective in this paper is to estimate the changing variances, or *integrated variance* (IV). The IV is a measure of the variability of financial asset returns over a specified period, for example, a day (a formal definition of IV will be given in Section 2). Recently, a new class of estimators for the IV has been developed by Barndorff-Nielsen and Shephard (2001), Barndorff-Nielsen and Shephard (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 high frequency financial time series data such as minute-by-minute return data or entire records of quote or transaction price data. The RV is a model-free estimator in the sense that we do not have to specify the volatility dynamics. Under moderate assumptions, the RV converges in probability to the IV, as the sampling frequency tends to be high.

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 called *microstructure noise* (MN) and emerges because of, for example, discreteness of prices, bid–ask bounce and infrequent trading, etc. When this assumption is violated, the RV is no longer a consistent estimator for the IV. It can be shown that, under the existence of MN, the RV diverges as the sampling frequency increases. Several alternative estimators of the RV, which are consistent even under the existence of MN, have been proposed by Zhou (1996), Zhang *et al.* (2005), Hansen and Lunde (2006), Bandi and Russell (2006) and Barndorff-Nielsen *et al.* (2008). See also Bandi and Russell (2008), who consider a mean-squared-error optimal sampling theory for reducing the effect of MN.

We call the RV calculated with observed log-prices contaminated by MN the *noise-contaminated RV* (NCRV) and refer to the component in the NCRV associated with the MN as the *MN component* (a formal definition of the NCRV and MN component is given in Section 2.3). We propose a state space approach to estimating the IV and MN components simultaneously. Our approach is an extension of the state space method proposed by Barndorff-Nielsen and Shephard (2002), who consider a situation with no MN. In this situation, Barndorff-Nielsen and Shephard (2002) show that the IV follows an ARMA process[1] for some specific continuous–time SV models. Barndorff-Nielsen and Shephard (2002) also show that the RV can be represented as a state space form, namely, the sum of the IV and an discretization error, which is a white noise uncorrelated with the IV. Thus, given the state space form parameters, one can apply the Kalman filter to filter out the discretization error. Simulation study by Barndorff-Nielsen and Shephard (2002) demonstrates that the estimates of IV series by Kalman smoother have much smaller mean squared error than the RV series itself. This ARMA representation result is further developed by Meddahi (2003). Meddahi (2003) shows that the IV follows an ARMA process for a general class of continuous–time SV models, which is called the *square root stochastic autoregressive variance* (SR–SARV) model (Andersen, 1994; Meddahi and Renault, 2004). Meddahi (2003) derives explicit relationships between the ARMA model parameters and the SV model parameters.

We develop the state space method by Barndorff-Nielsen and Shephard (2002) for dealing with the problem of MN. We assume that an observed log-price is the sum of the true log-price and an i.i.d. MN. We represent the NCRV by a state space form in that the NCRV is the sum of three unobserved components: the IV, which follows an ARMA process, a white noise (discretization error) and a MN component, which follows a MA(1) process. By applying the results of Granger and Morris (1976), we show that the sum of these three components, namely, the NCRV, follows

[1] We interchangeably use the term "ARMA process" and "ARMA model" in this paper.

an ARMA process. This ARMA process can be regarded as the (unique) reduced form of the state space form. The existence of MN component introduces many complexities in the identification of the state space form parameters. It is shown that the number of state space form parameters of the NCRV is more than the autocovariance structure of the NCRV can uniquely determine. In other words, the state space form parameters of the NCRV are not effectively identified in the sense that different sets of parameter values can give the same autocovariance structure. See Section 4 for more details.

We show that the state space form parameters can be expressed as functions of the unconditional mean and variance parameters of the underlying continuous–time SV model and parameters regarding MN (the variances of the MN and its square). Then, we prove that these parameters are uniquely identified. We illustrate how to estimate these identifiable parameters and the state space form parameters. With estimates of the state space form parameters, one can estimate the IV and MN components simultaneously by applying the Kalman filter to the state space form. One advantage of our method, compared with other existing methods, is that it can filter out not only the MN components but also the discretization errors. The proposed method is applied to yen/dollar spot exchange rate data. We find that the magnitude of the (daily) MN component is, on average, about 21% – 48% of the (daily) NCRV, depending on the sampling frequency.

The rest of the paper is organized as follows. In the next section, we introduce the class of SV models employed in this paper and define formally the RV, IV, MN and MN component. In Section 3, we briefly summarize the results in Meddahi (2003) on the ARMA representation of the IV. In Section 4, we explain our state space approach in detail. In Section 5, we conduct an empirical analysis applying our method to the yen/dollar spot exchange rate. The last section provides a summary and concluding remarks. Appendix A provides details on the derivations of the equations in the text. Some results are presented in Appendix B.

2 SR-SARV model, IV, RV and MN

2.1 Square root stochastic autoregressive variance (SR-SARV) model

Let $p(t)$ be the log of the (efficient) spot price at time t. Throughout the paper, we assume that $p(t)$ follows the SR–SARV model considered in Meddahi (2003), which is given by the following class of continuous–time SV models:

$$dp(t) = \sigma(t)dW_t, \quad \sigma^2(t) = \sigma^2 + \omega_1 P_1(f(t)) + \omega_2 P_2(f(t)), \tag{1}$$

where $f(t)$ is a state-variable process and the functions $P_1(\cdot)$ and $P_2(\cdot)$ are defined so that:

$$\begin{aligned} &E[P_1(f(t))] = E[P_2(f(t))] = 0, \quad \mathrm{var}[P_1(f(t))] = \mathrm{var}[P_2(f(t))] = 1, \\ &\qquad\qquad\qquad \mathrm{cov}[P_1(f(t)), P_2(f(t))] = 0, \\ &E[P_1(f(t+h))|f(s), p(s), s \le t] = \exp(-\lambda_1 h) P_1(f(t)), \\ &E[P_2(f(t+h))|f(s), p(s), s \le t] = \exp(-\lambda_2 h) P_2(f(t)), \quad \forall h > 0, \end{aligned} \tag{2}$$

where λ_1 and λ_2 are positive real numbers. The unconditional mean and variance of $\sigma^2(t)$ are $E[\sigma^2(s)] = \sigma^2$ and $\mathrm{var}[\sigma^2(s)] = \omega_1^2 + \omega_2^2$, respectively. Let $\kappa_1 = \exp(-\lambda_1)$ and $\kappa_2 = \exp(-\lambda_2)$. Hereafter, we work mainly with κ_1 and κ_2 instead of λ_1 and λ_2 because it is more convenient for describing our results. Thus, the model has a total of five free parameters: σ^2, ω_1^2, ω_2^2, κ_1 and κ_2.

The model given in (1) and (2) is called the "two–factor SR–SARV model". When $\omega_2 = 0$, the model is refereed to as the "one–factor SR–SARV model". The SR-SARV model includes many known models, such as constant elasticity of volatility processes, GARCH diffusion models (Nelson, 1990), eigenfunction stochastic volatility models (Meddahi, 2001) and positive Ornstein–Uhlenbeck Levy-driven models (Barndorff-Nielsen and Shephard, 2001). See Meddahi (2003) for more details.

2.2 Integrated and realized variances

Given the process of $\sigma^2(t)$, the IV is defined as

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

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

Under moderate assumptions, we can consistently estimate the IV by the estimator known as the RV, which is defined as

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

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 $m = 288$. In this case, $r_t^{(288)}$ denotes a five–minute return, because one day is 5×288 minutes. It is well known that, as $m \to \infty$, $RV_t^{(m)} \xrightarrow{p} IV_t$ (see, e.g., Barndorff-Nielsen and Shephard, 2002).

For the two–factor SR–SARV model, the variance and autocovariances of IV_t are expressed in terms of the SV model parameters as:

$$\begin{aligned} \mathrm{var}[IV_t] &= \frac{2\omega_1^2(\kappa_1 - \log\kappa_1 - 1)}{(\log\kappa_1)^2} + \frac{2\omega_2^2(\kappa_2 - \log\kappa_2 - 1)}{(\log\kappa_2)^2}, \\ \mathrm{cov}[IV_t, IV_{t-1}] &= \frac{\omega_1^2(1-\kappa_1)^2}{(\log\kappa_1)^2} + \frac{\omega_2^2(1-\kappa_2)^2}{(\log\kappa_2)^2}, \quad \text{and} \\ \mathrm{cov}[IV_t, IV_{t-2}] &= \frac{\omega_1^2\kappa_1(1-\kappa_1)^2}{(\log\kappa_1)^2} + \frac{\omega_2^2\kappa_2(1-\kappa_2)^2}{(\log\kappa_2)^2}. \end{aligned} \tag{3}$$

Let $d_t^{(m)} \equiv RV_t^{(m)} - IV_t$ and $\sigma_d^{2(m)} \equiv \mathrm{var}[d_t^{(m)}]$. For $m \geq 1$, we have:

$$\sigma_d^{2(m)} = \frac{2\sigma^4}{m} + \frac{4\omega_1^2 m}{(\log\kappa_1)^2}\left(\kappa_1^{\frac{1}{m}} - \log\kappa_1^{\frac{1}{m}} - 1\right) + \frac{4\omega_2^2 m}{(\log\kappa_2)^2}\left(\kappa_2^{\frac{1}{m}} - \log\kappa_2^{\frac{1}{m}} - 1\right). \tag{4}$$

It can be shown that $\sigma_d^{2(m)} \to 0$ as $m \to \infty$. See Meddahi (2003) for the above results.

2.3 MN component

Now assume that the observed log-price $p^*(t)$ is contaminated by a measurement error or MN so that:

$$p^*(t) = p(t) + \varepsilon(t).$$

We assume the following properties of MN $\varepsilon(t)$.

Assumption 1

(a) $\varepsilon(t) \sim i.i.d.(0, \sigma_\varepsilon^2)$ with $\omega_\varepsilon^2 \equiv \mathrm{var}[\varepsilon^2(t)] < \infty$.

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

We do not assume any specific distribution for $\varepsilon(t)$.

The observed return $r_t^{*(m)}$ is defined as:

$$r_t^{*(m)} \equiv p^*(t) - p^*\left(t - \tfrac{1}{m}\right) = r_t^{(m)} + e_t^{(m)}, \tag{5}$$

where $e_t^{(m)} \equiv \varepsilon(t) - \varepsilon(t - \frac{1}{m})$. It is easy to show that

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

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

$$RV_t^{*(m)} = \sum_{i=1}^{m} \left(r_{t-1+\frac{i}{m}}^{(m)} + e_{t-1+\frac{i}{m}}^{(m)} \right)^2 = RV_t^{(m)} + u_t^{(m)}, \tag{6}$$

where

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

Note that, unlike $RV_t^{*(m)}$, $u_t^{(m)}$ is not necessarily positive because the first term of $u_t^{(m)}$ may be negative. We call $u_t^{(m)}$ an *MN component*. We propose a way of estimating the MN component as well as the IV in a later section.

In Appendix A, we show that:

$$\begin{aligned} E\left[u_t^{(m)}\right] &= 2m\sigma_\varepsilon^2 \quad \text{and} \\ \text{cov}\left[u_t^{(m)}, u_s^{(m)}\right] &= \begin{cases} 8\sigma_\varepsilon^2\sigma^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4 & t = s, \\ \omega_\varepsilon^2 & t = s \pm 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{7}$$

Thus, $u_t^{(m)}$ has the autocovariance structure of a MA(1) process. Assume that the MA(1) process is expressed as:

$$u_t^{(m)} = c_u^{(m)} + \xi_t^{(m)} + \theta_u^{(m)} \xi_{t-1}^{(m)}, \quad \xi_t^{(m)} \sim WN(0, \sigma_\xi^{2(m)}), \tag{8}$$

where $WN(0, a)$ denotes a white noise process with variance a. The mean and autocovariances of $u_t^{(m)}$, in terms of $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$, are:

$$\begin{aligned} E\left[u_t^{(m)}\right] &= c_u^{(m)} \quad \text{and} \\ \text{cov}\left[u_t^{(m)}, u_s^{(m)}\right] &= \begin{cases} (1 + \theta_u^{2(m)})\sigma_\xi^{2(m)} & t = s, \\ \theta_u^{(m)} \sigma_\xi^{2(m)} & t = s \pm 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{9}$$

Later, we utilize these two different expressions of the moments of $u_t^{(m)}$ to derive the implicit relationships among the SV and MA(1) process parameters.

3 ARMA Representation of IV

In this section, we briefly summarize the results in Meddahi (2003) on an ARMA representation of IV for SR-SARV models.

3.1 One–factor case

Meddahi (2003, Proposition 3.1) shows that if the true process of $p(t)$ follows a one–factor SR–SARV model, then IV_t follows an ARMA(1, 1) process:

$$IV_t = c_{IV} + \kappa_1 IV_{t-1} + \eta_t + \theta_1 \eta_{t-1}, \tag{10}$$

where κ_1 is defined as in the statement below (2), η_t is a white noise process with $\text{var}(\eta_t) = \sigma_\eta^2$ and $\text{cov}(\eta_t, d_s^{(m)}) = 0$ for all t and s. Other ARMA(1, 1) model parameters c_{IV}, θ_1 and σ_η^2 are expressed in terms of the one–factor SR-SARV model parameters σ^2, ω_1^2 and κ_1 as:

$$c_{IV} = (1-\kappa_1)\sigma^2, \quad \theta_1 = \frac{1-\sqrt{1-4\rho^2}}{2\rho},$$
$$\sigma_\eta^2 = \frac{(1+\kappa_1^2)\mathrm{var}[IV_t] - 2\kappa_1 \mathrm{cov}[IV_t, IV_{t-1}]}{1+\theta_1^2}, \tag{11}$$

where

$$\rho \equiv \frac{-\kappa_1 + \mathrm{corr}[IV_t, IV_{t-1}]}{1+\kappa_1^2 - 2\kappa_1 \mathrm{corr}[IV_t, IV_{t-1}]}.$$

It can be shown that ρ is equal to $\theta_1/(1-\theta_1^2)$, i.e., the first order autocorrelation of the MA(1) process $\eta_t + \theta_1\eta_{t-1}$ in (10). The $\mathrm{corr}[IV_t, IV_{t-1}]$ is given by

$$\mathrm{corr}[IV_t, IV_{t-1}] = \frac{(1-\kappa_1)^2}{2(\kappa_1 - \log\kappa_1 - 1)}.$$

Note that $\mathrm{corr}[IV_t, IV_{t-1}]$ is a function of κ_1 and does not depend on other SV model parameters, which, in turn, implies that θ_1 is also a function of only κ_1. This is not true for the two–factor case. This substantially simplifies the identification problem of the state space form of the NCRV, as we will see in Section 4.

3.2 Two–factor case

Meddahi (2003, Proposition 3.3) shows that if the true process of $p(t)$ belongs to the two–factor SR–SARV model, then IV_t follows an ARMA(2, 2) process:

$$IV_t = c_{IV} + (\kappa_1+\kappa_2)IV_{t-1} - \kappa_1\kappa_2 IV_{t-2} + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}, \tag{12}$$

where κ_1 and κ_2 are defined as in the statement below Equation (2), η_t is a white noise process with $\mathrm{var}(\eta_t) = \sigma_\eta^2$ and $\mathrm{cov}(\eta_t, d_s^{(m)}) = 0$ for all t and s. Let $\phi_1 = \kappa_1 + \kappa_2$ and $\phi_2 = -\kappa_1\kappa_2$.[2] Other ARMA(2, 2) model parameters in (12) c_{IV}, θ_1, θ_2 and σ_η^2 are expressed in terms of the two–factor SR–SARV model parameters σ^2, ω_1^2, ω_2^2, κ_1 and κ_2 as:

$$c_{IV} = (1-\phi_1-\phi_2)\sigma^2, \quad \theta_1 = \frac{1-\sqrt{4s+1}}{2}\frac{\rho_1}{\rho_2}, \quad \theta_2 = \frac{\sqrt{4s+1}-2s-1}{2s},$$
$$\sigma_\eta^2 = \frac{\pi_1 \mathrm{var}[IV_t] - 2\pi_2 \mathrm{cov}[IV_t, IV_{t-1}] - 2\phi_2 \mathrm{cov}[IV_t, IV_{t-2}]}{1+\theta_1^2+\theta_2^2}, \tag{13}$$

where

$$\pi_1 = 1+\phi_1^2+\phi_2^2, \quad \pi_2 = \phi_1(1-\phi_2),$$

$$s \equiv -\frac{\rho_2^2}{\rho_1^2}\left[1+\frac{1}{2\rho_2} - \mathrm{sign}(\rho_2)\sqrt{\left(1+\frac{1}{2\rho_2}\right)^2 - \frac{\rho_1^2}{\rho_2^2}}\,\right],$$

$$\rho_1 \equiv \frac{-\phi_1(1-\phi_2) + (1+\phi_1^2-\phi_2)\mathrm{corr}[IV_t, IV_{t-1}] - \phi_1 \mathrm{corr}[IV_t, IV_{t-2}]}{(1+\phi_1^2+\phi_2^2) - 2\phi_1(1-\phi_2)\mathrm{corr}[IV_t, IV_{t-1}] - 2\phi_2 \mathrm{corr}[IV_t, IV_{t-2}]}, \tag{14}$$

$$\rho_2 \equiv \frac{-\phi_2 - \phi_1 \mathrm{corr}[IV_t, IV_{t-1}] + \mathrm{corr}[IV_t, IV_{t-2}]}{(1+\phi_2^2+\phi_1^2) - 2\phi_1(1-\phi_2)\mathrm{corr}[IV_t, IV_{t-1}] - 2\phi_2 \mathrm{corr}[IV_t, IV_{t-2}]},$$

[2] We can rewrite the ARMA(2, 2) form in (12) with a more familiar parameterization, i.e., $IV_t = c_{IV} + \phi_1 IV_{t-1} + \phi_2 IV_{t-2} + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}$. The expressions of κ_1 and κ_2 in terms of ϕ_1 and ϕ_2 are given as $\kappa_1 = \frac{\phi_1+\sqrt{\phi_1^2+4\phi_2}}{2}$ and $\kappa_2 = \frac{\phi_1-\sqrt{\phi_1^2+4\phi_2}}{2}$, respectively.

and sign(ρ_2) = 1 if $\rho_2 > 0$ and sign(ρ_2) = −1 if $\rho_2 < 0$. We assume that $\rho_2 \neq 0$, which implies that $\theta_2 \neq 0$. As in the one–factor case, we can show that $\rho_1 = (\theta_1 + \theta_1\theta_2)/(1 + \theta_1^2 + \theta_2^2)$ and $\rho_2 = \theta_2/(1 + \theta_1^2 + \theta_2^2)$, i.e., ρ_1 and ρ_2 are the first and second order autocorrelations of the MA(2) process $\eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}$ in (12), respectively. See Meddahi (2002) and Meddahi (2003) for more details.

4 State Space Approach

In this section, we explain our state space approach in detail. Our state space approach is in the same spirit as the state space method used in Barndorff-Nielsen and Shephard (2002), who consider the situation without MN. First, we give a state space form of the NCRV in Section 4.1. The existence of MN components requires additional efforts for checking the identification of the state space form. In Section 4.2, we show that the state space form parameters are not identifiable; however, they can be expressed as functions of fewer identifiable parameters. We illustrate how to estimate these identifiable parameters in Section 4.3.

In what follows, we assume that $\omega_2 = 0$ for ease of exposition. Corresponding results for the two–factor case can be derived in a similar manner and are summarized in the Appendix B.

4.1 State space form of the NCRV

Substituting $RV_t^{(m)} = IV_t + d_t^{(m)}$ into (6), we have:

$$RV_t^{*(m)} = IV_t + d_t^{(m)} + u_t^{(m)}. \tag{15}$$

Let η_t and $\xi_t^{(m)}$ be denoted by the state variables α_t and $\beta_t^{(m)}$, respectively. From (8), (10) and (15), we have the following state space form of $RV_t^{*(m)}$:

Observation equation

$$RV_t^{*(m)} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t^{(m)} \end{bmatrix} + d_t^{(m)}, \tag{16a}$$

State equation

$$\begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t^{(m)} \end{bmatrix} = \begin{bmatrix} c_{IV} \\ c_u^{(m)} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \kappa_1 & 0 & \theta_1 & 0 \\ 0 & 0 & 0 & \theta_u^{(m)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_{t-1} \\ u_{t-1}^{(m)} \\ \alpha_{t-1} \\ \beta_{t-1}^{(m)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t^{(m)} \end{bmatrix}, \tag{16b}$$

where

$$\begin{bmatrix} d_t^{(m)} \\ \eta_t \\ \xi_t^{(m)} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_d^{2(m)} & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\xi^{2(m)} \end{bmatrix} \right). \tag{16c}$$

Given the values of c_{IV}, κ_1, θ_1, 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 filter to the state space form.[3] One problem of the state space form is how to estimate those parameters. One may simply think that we could estimate them directly from the state space form by, for example, quasi-maximum likelihood (QML) estimation under Gaussian

[3] Note that here η_t and ξ_t do not follow a Gaussian distribution. In this case, the Kalman filter provides the best linear unbiased estimator (Anderson and Moore, 1979). See Durbin and Koopman (2001) for more details on the Kalaman filter.

noise assumption. We show; however, that this approach is not applicable for the state space form given in (16a) – (16c).

In general, parameters of a state space form are not necessarily identified (see, for example, Hamilton, 1994, p.388). More precisely, they are not identified in the sense that there are infinitely many combinations of the parameters that give the same autocovariance structure. Thus, we have to check whether state space form parameters are uniquely identified before proceeding to their estimation. We consider this problem in the next subsection. In fact, we show that the above parameters in the state space form cannot be uniquely identified.

4.2 Identification of model parameters

Because $RV_t^{*(m)}$ 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:

$$(1-\kappa_1 L)RV_t^{*(m)} = c_{RV}^{(m)} + (1+\delta_1^{(m)}L + \delta_2^{(m)}L^2)\tau_t^{(m)}, \quad \tau_t^{(m)} \sim WN(0, \sigma_\tau^{2(m)}). \tag{17}$$

Note that the AR coefficient κ_1 is the same as that of the IV_t in (10). The ARMA model representation of a state space form is commonly referred to as a reduced form or ARMA reduced form. Parameters of the ARMA reduced form are identifiable.

From (8), (10) and (15), we have

$$\begin{aligned}(1-\kappa_1 L)RV_t^{*(m)} &= (1-\kappa_1 L)IV_t + (1-\kappa_1 L)d_t^{(m)} + (1-\kappa_1 L)u_t^{(m)} \\ &= c_{IV} + \eta_t + \theta_1\eta_{t-1} + d_t^{(m)} - \kappa_1 d_{t-1}^{(m)} + \xi_t^{(m)} \\ &\quad +(1-\kappa_1)c_u^{(m)} + (\theta_u^{(m)} - \kappa_1)\xi_{t-1}^{(m)} - \kappa_1\theta_u^{(m)}\xi_{t-2}^{(m)}.\end{aligned} \tag{18}$$

The two expressions on the right-hand sides in (17) and (18) are of the same process and hence their means and autocovariances must be identical. The autocovariances of the MA process in (17) are given as

$$\begin{aligned}&\gamma_0^{(m)} = (1+\delta_1^{(m)2} + \delta_2^{(m)2})\sigma_\tau^{2(m)}, \quad \gamma_1^{(m)} = (\delta_1^{(m)} + \delta_1^{(m)}\delta_2^{(m)})\sigma_\tau^{2(m)}, \\ &\gamma_2^{(m)} = \delta_2^{(m)}\sigma_\tau^{2(m)} \quad \text{and} \quad \gamma_j = 0 \text{ for } j \geq 3.\end{aligned} \tag{19}$$

It is shown in the Appendix A that the autocovariances of the MA process in (18) are

$$\gamma_0^{(m)} = (1+\theta_1^2)\sigma_\eta^2 + (1+\kappa_1^2)\sigma_d^{2(m)} + [1 + (\theta_u^{(m)} - \kappa_1)^2 + \kappa_1^2\theta_u^{(m)2}]\sigma_\xi^{2(m)}, \tag{20a}$$

$$\gamma_1^{(m)} = \theta_1\sigma_\eta^2 - \kappa_1\sigma_d^{2(m)} + (\theta_u^{(m)} - \kappa_1 - \kappa_1\theta_u^{(m)2} + \kappa_1^2\theta_u^{(m)})\sigma_\xi^{2(m)}, \tag{20b}$$

$$\gamma_2^{(m)} = -\kappa_1\theta_u^{(m)}\sigma_\xi^{2(m)}, \tag{20c}$$

and $\gamma_j = 0$ for $j \geq 3$. By equating the means of the MA processes in (17) and (18), we have

$$c_{RV}^{(m)} = c_{IV} + (1-\kappa_1)c_u^{(m)}. \tag{20d}$$

Given the ARMA(1, 2) model parameters, $c_{RV}^{(m)}$, κ_1, δ_1, δ_2 and $\sigma_\tau^{2(m)}$, we can calculate $\gamma_j^{(m)}$, $j = 0, 1, 2$. Then, unknown parameters in the equations (20a)$\sim$(20d) are only the state space form parameters, c_{IV}, θ_1, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$. Observe that there are seven unknown parameters and only four equations. Hence, we cannot uniquely identify these parameters from these equations. In other words, for a given ARMA(1, 2) reduced form, there are infinitely many sets of values of c_{IV}, θ_1, 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) reduced form.

In view of (7) and (9), we obtain the following equations:

$$c_u^{(m)} = 2m\sigma_\varepsilon^2, \tag{21a}$$

$$(1+\theta_u^{(m)2})\sigma_\xi^{2(m)} = 8\sigma^2\sigma_\varepsilon^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4, \tag{21b}$$

$$\theta_u^{(m)}\sigma_\xi^{2(m)} = \omega_\varepsilon^2. \tag{21c}$$

Assuming that the MA parameter satisfies the invertibility condition, i.e., $|\theta_u^{(m)}| < 1$, we can solve the equations (21a) $\sim$ (21c) for $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$ as:

$$c_u^{(m)} = 2m\sigma_\varepsilon^2, \quad \sigma_\xi^{2(m)} = \frac{\omega_\varepsilon^2}{\theta_u^{(m)}} \quad \text{and} \quad \theta_u^{(m)} = A - \sqrt{A^2 - 1}, \tag{22}$$

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

From (3), (11) and (22), we see that c_{IV}, θ_1, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ are expressed as functions of κ_1, σ^2, ω_1^2, σ_ε^2 and ω_ε^2.[4] To emphasize these relationships, we denote them as:

$$\begin{aligned} &c_{IV}(\kappa_1, \sigma^2), \quad \theta_1(\kappa_1), \quad \sigma_\eta^2(\kappa_1, \omega_1^2), \quad c_u^{(m)}(\sigma_\varepsilon^2), \quad \theta_u^{(m)}(\sigma^2, \sigma_\varepsilon^2, \omega_\varepsilon^2), \\ &\sigma_d^{2(m)}(\kappa_1, \sigma^2, \omega_1^2) \quad \text{and} \quad \sigma_\xi^{2(m)}(\sigma^2, \sigma_\varepsilon^2, \omega_\varepsilon^2). \end{aligned} \tag{23}$$

Note that θ_1 is a function of only κ_1 and hence can be assumed to be known (because κ_1 is identified from the reduced form). Substituting the expressions in (23) into Equations (20a)$\sim$(20d), we have four equations for the four unknown parameters σ^2, ω_1^2, σ_ε^2 and ω_ε^2. Hence, the order condition for identification is satisfied. However, this result does not imply that one can *uniquely* identify σ^2, ω_1^2, σ_ε^2 and ω_ε^2.

To show the uniqueness of the identification, we explicitly derive the representations of σ^2, ω_1^2, σ_ε^2 and ω_ε^2 in terms of $c_{RV}^{(m)}$, κ_1, $\gamma_j^{(m)}$, $j = 0, ..., 2$. In Appendix A, we show that, given $c_{RV}^{(m)}$, κ_1, $\gamma_j^{(m)}$, $j = 0, ..., 2$ and (23), Equations (20a)$\sim$(20d) are uniquely[5] solved for σ^2, ω_1^2, σ_ε^2 and ω_ε^2 as:

$$\omega_\varepsilon^2 = -\frac{\gamma_2^{(m)}}{\kappa_1}, \quad \omega_1^2 = \frac{(\log\kappa_1)^2[\kappa_1\gamma_0^{(m)} + (1+\kappa_1^2)\gamma_1^{(m)} + \frac{1+\kappa_1^4}{\kappa_1}\gamma_2^{(m)}]}{(1-\kappa_1)^3(1+\kappa_1)}, \tag{24a}$$

$$\sigma_\varepsilon^2 = \sqrt{\frac{c_{RV}^{(m)2}}{2m^2(1-\kappa_1)^2} - \frac{(2m-1)\gamma_2^{(m)}}{2m\kappa_1} - \frac{\gamma_0^{(m)} - 2D\omega_1^2 - 2\gamma_2^{(m)}}{4m(1+\kappa_1^2)}}, \tag{24b}$$

and

$$\sigma^2 = \frac{c_{RV}^{(m)}}{1-\kappa_1} - 2m\sigma_\varepsilon^2, \quad \text{where} \quad D = B + m(1+\kappa_1^2)C, \tag{24c}$$

$$B \equiv \frac{\kappa_1^2 - 1 - (1+\kappa_1^2)\log\kappa_1}{(\log\kappa_1)^2} \quad \text{and} \quad C \equiv \frac{2\left(\kappa_1^{\frac{1}{m}} - 1 - \log\kappa_1^{\frac{1}{m}}\right)}{(\log\kappa_1)^2}. \tag{24d}$$

These results imply that the four parameters, σ^2, ω_1^2, σ_ε^2 and ω_ε^2 are uniquely identified from the ARMA(1, 2) reduced form in (17). Hence, in principle, we can estimate them. Again, it should be emphasized that these results do *not* imply that one can directly estimate the state space form parameters but rather that one can estimate the above four parameters by replacing the state space form parameters with the functions of the four parameters. The estimates of the state space form parameters are obtained by substituting the estimates of the four parameters into these functions.

4.3 Estimation of model parameters

We illustrate how to estimate the four parameters. There are two possible approaches: direct and indirect. Below, we illustrate first the indirect and then the direct approach. In both approaches, we apply QML estimation assuming Gaussian innovations.

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

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

We showed in (24) that these four parameters have explicit expressions in terms of the ARMA(1, 2) reduced form parameters. This suggests the following indirect approach for estimating these four parameters.

Summary of the indirect approach

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

Step 2 Estimate the unrestricted ARMA(1, 2) model in (17) by QML estimation assuming Gaussian innovations.

Step 3 Given the estimates of $c_{RV}^{(m)}$, κ_1, $\delta_1^{(m)}$, $\delta_2^{(m)}$ and σ_τ^2 obtained in Step 2, calculate the first three autocovariances of the MA process, namely, $\gamma_j^{(m)}$, $j = 0 \sim 2$ as in (19).

Step 4 Given the estimates of $c_{RV}^{(m)}$, κ_1 and $\gamma_j^{(m)}$, $j = 0 \sim 2$ obtained in Steps 2 and 3, estimate ω_ε^2, σ_ε^2, ω_1^2 and σ^2 applying the results in (24a) – (24d).

This approach is simple and easy to implement; however, it does not guarantee that the resulting parameter estimates are positive because of the inevitable uncertainty of the ARMA model estimation. For example, if $\gamma_2^{(m)} > 1$, then the estimate of ω_ε^2 by this approach is negative because $\kappa_1 > 0$ by assumption.

Alternatively, one can directly estimate these four parameters. In this approach, one calculates the log-likelihood directly from the four parameters and maximizes it with respect to the four parameters. Thus, we can easily impose the positivity of the four parameters. Below, we summarize how to obtain the QML estimates by this approach.

Summary of the direct approach

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

Step 2 Given κ_1, σ^2, ω_1^2, σ_ε^2 and ω_ε^2, calculate c_{IV}, θ_1, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ according to (3), (11) and (22).

Step 3 With the c_{IV}, θ_1, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ obtained in Step 2, calculate the Gaussian log-likelihood of the state space form given in (16a)–(16c) for RV_t^*.

Step 4 Maximize the log-likelihood obtained in Step 3 with respect to the five parameters κ_1, σ^2, ω_1^2, σ_ε^2 and ω_ε^2 to obtain the QML estimates.

This approach provides consistent estimators for the four parameters.

Before closing this section, it should be noted that if we can obtain estimates properly by the indirect approach, we do not need to proceed to the direct approach, because both approaches will give the identical estimates in this case.

5 Empirical Analysis

In this section, we conduct an empirical analysis with exchange rate data using the proposed state space method.

5.1 Data description

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

Price data are not available for each minute. When price data are missing we apply the previous tick method, i.e., we interpolate the most recent observed price. Furthermore, following Andersen, Bollerslev, Diebold and Labys (2001), we remove the data of inactive trading days. Whenever we do so, we always remove from 21:01 GMT on one night to 21:00 the next evening. For details on the motivation behind 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 criteria, which are similar to the criteria adapted in Beine *et al.* (2007). Specifically, we cut

(1) the days where there are more than 500 missing price observations,

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

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

By these criteria, we could remove all weekend data. However, the days such as US holidays that Andersen, Bollerslev, Diebold and Labys (2001) and Beine *et al.* (2007) remove are not necessarily removed by these criteria. This is because even when the US market is closed, 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 one–minute and five–minute returns.

With these returns, we calculate two series of daily NCRV, namely, one–minute NCRV ($m = 1440$) and five–minute NCRV ($m = 288$). Table 1 reports the descriptive statistics of these two series of NCRV, and Figure 2 plots them. The sample mean of the one–minute NCRV is greater than that of the five–minute NCRV. This is consistent with the existence of MN because the mean of the NCRV increases as the sampling frequency increases, or $m \to \infty$ under the existence of MN (see (20d) and (21a)). The first order autocorrelations of these two series of NCRV are somewhat lower than usually expected for variances of financial time series: they are 0.4794 for the one–minute NCRV and 0.4177 for the five–minute NCRV. This may be because of the existence of MN. In fact, in the next subsection, we show that estimates of the first order autocorrelation of the IV are significantly higher than these values.

5.2 Estimation of parameters, IV and MN component

For these two series of the NCRV, we estimate the parameters of the one– and two–factor SV models by the method described in Section 4.3 (and in Appendix B for the two–factor case). Note that, in general, the values of these two NCRV series are different although they both are estimates of the same IV series. Consequently, the estimates of the SV model parameters are different, depending on which NCRV series is used. We report only the results by the direct approach because the indirect approach does not provide positive variance estimates. Table 2 displays the estimates of the SV model parameters. Naturally, the estimated values of the SV model parameters for one–minute and five–minute NCRV series are very similar. In both the one– and two–factor cases, estimates with the five–minute NCRV series are slightly more efficient than those with the one–minute NCRV series according to the robust standard errors.[6] The estimates of the persistence parameters for two–factor SV model (i.e., $\widehat{\kappa}_1$ and $\widehat{\kappa}_2$) imply that there are two factors with significantly different levels of persistence. One of them is very persistent and the other is moderately persistent, although their unconditional variances are not significantly different. For the one–factor SV model, the persistence of these two factors must be captured by only one parameter, κ_1. As a result, the estimate of κ_1 in the one–factor case is somewhat lower than that in the two–factor case.

The estimates of state space form parameters in (16) (and in (40) for the two–factor case) are computed from the estimates of the SV model parameters. They are shown in Table 3. Again, the estimates of the common parameters, which do not depend on m, are very similar. We find that the estimates of the mean of the MN component, denoted by $\widehat{c}_u^{(m)}$, in one–minute NCRV series is

[6] To obtain the QML estimates of the SV model parameters, first, we calculate the QML estimates of the transformed ones, such as $\mu \equiv \log(\sigma^2)$, by applying an unconstrained maximization procedure. Then, the QML estimate of, for example, σ^2 is obtained by $\log(\widehat{\mu})$, where $\widehat{\mu}$ is the QML estimate of μ. The robust standard errors of the SV model parameter estimates are calculated as follows. First, generate samples from the asymptotic normal distribution of the estimators of the transformed parameters (such as $\widehat{\mu}$) with their robust asymptotic covariance matrix estimates (and the mean being set to the estimates). Next, for each sample, calculate the estimates of the SV model parameters. Lastly, calculate the sample standard deviations of these SV model parameter estimates, which are our robust standard errors.

greater than that in five–minute NCRV series, which implies that the one–minute NCRV series has a larger bias than the five–minute NCRV series. This is consistent with the theory. The magnitude of bias of the one–minute NCRV is about four times larger than that of the five–minute NCRV.

Table 4 reports the estimates of some important values including the autocorrelations of the IV. In both the one– and two–factor cases, the estimates of the first order autocorrelation of IV are significantly higher than those of the two NCRV series. This result suggests that the existence of MN lowers the autocorrelations of the NCRV series. The estimates of the ratio of the unconditional variance of the MN component to the unconditional variance of the NCRV imply that about half of the aggregate fluctuations of the NCRV series is because of the MN component.

We display the estimates of the IV series by Kalman smoothing for the five–minute and one–minute NCRV series in Figures 3(a) and (b), respectively. Figure 3(c) is the difference between them, or $\widehat{IV}_t^{(1440)} - \widehat{IV}_t^{(288)}$, where $\widehat{IV}_t^{(m)}$ is the estimate of IV_t with a given m. Note that these estimates are the estimates of the same IV series and thus are very similar. The IV estimates in the one–factor case seem smoother than those in the two–factor case. This is because of the result that the (estimated) autocorrelations of the IV series are lower for the two–factor case and thus they are relatively closer to white noise compared with the IV series obtained for the one–factor case. Figure 4 (a), (b) and (c) plot the smoothed estimates of the five–minute and one–minute MN component series and their differences, respectively. We can see that the MN component occasionally takes a large value. Figure 5 (a) and (b) display the estimates of the discretization errors by Kalman smoothing for five–minute and one–minute NCRV, respectively. The discretization error estimates for the one–minute NCRV series is quite small than those for the five–minute NCRV series, which is again consistent with the theoretical result. Corresponding figures for the two–factor case are given in Figures 6–8. They are very similar to those for the one–factor case.

Finally, we calculate the ratios of the MN component to the NCRV $\widehat{R}^{(m)}$. They are given by $\widehat{R}^{(m)} = \widehat{u}_t^{(m)}/RV_t^{*(m)}$, $t = 1, ..., 1809$, where $\widehat{u}_t^{(m)}$ is the estimate of $u_t^{(m)}$ by Kalman smoothing. The results are shown in Table 5. In the one–factor (two–factor) case, the maximum and minimum values of $\widehat{R}^{(m)}$ are, respectively, 0.8324 (0.6574) and −0.5804 (−3.7323) for the five–minute NCRV series and 0.8357 (1.0454) and −0.5804 (−0.4192) for the one–minute NCRV series. We also calculate the average magnitude of the MN component as the mean of $|R^{(m)}|$ (the average of $R^{(m)}$ is also reported in Table 5). In the one–factor (two–factor) case, the value of the mean is 0.4659 (0.2080) for the five–minute NCRV series and 0.4708 (0.4770) for the one–minute NCRV series. From these results, we conclude that the average magnitude of the MN component in the daily NCRV ranges from 21% to 48% of NCRV, depending on the sampling frequency.

6 Summary and Concluding Remarks

In this paper, we proposed a state space approach to estimating the IV and MN components simultaneously. Our method is based on the result in Meddahi (2003), who shows that when the true log-prices follow a general class of continuous–time SV models, the IV follows an ARMA process. We showed that under the existence of MN, the observed RV, or the NCRV, also follows an ARMA process. We represented the NCRV by a state space form and established the uniqueness of the identification of the state space form parameters. The proposed method was applied to yen/dollar exchange rate data, where we found that the NCRV calculated with five–minute returns is less biased than with one–minute returns. The two series of IV estimates by the proposed method with one–minute and five–minute returns are very similar. The method was also used for estimating the MN component.

In the estimation, we constructed the log-likelihood using only either the one–minute or five–minute NCRV series. It is more desirable to use both NCRV series for estimating the common parameters. It would be possible to obtain more efficient estimators by combining the one– and five–minute NCRV series. This is a subject for future research. It is also important to relax the assumption that there is no leverage effect in order to apply our method to stock return data.

Appendix A: Derivations of Equations

Hereafter, we suppress "(m)" in the notations $r_t^{(m)}$, $u_t^{(m)}$ and $e_t^{(m)}$, and let ε_t denote $\varepsilon(t)$ for notational simplicity.

Derivation of (7)

Because $\mathrm{var}(e_t) = 2\sigma_\varepsilon^2$ and r_t is independent of e_t by Assumption 1, we have:

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

To derive $\mathrm{var}[u_t]$ and $\mathrm{cov}[u_t, u_{t-1}]$, we calculate $\mathrm{cov}[r_s e_s, r_t e_t]$ and $\mathrm{cov}\left[e_t^2, e_s^2\right]$. When $t = s$, we have:

$$\begin{aligned} \mathrm{cov}\left[r_t e_t, r_t e_t\right] = &\ E\left[r_t^2 e_t^2\right] - (E\left[r_t e_t\right])^2 \\ = &\ E\left[e_t^2\right] E\left[r_t^2\right] - (E\left[r_t\right])^2 (E\left[e_t\right])^2 \\ = &\ 2\sigma_\varepsilon^2 E[(\textstyle\int_{t-1/m}^{t} \sigma(s) dW(s))^2] \\ = &\ 2\sigma_\varepsilon^2 E[\textstyle\int_{t-1/m}^{t} \sigma^2(s) ds] \\ = &\ \frac{2\sigma_\varepsilon^2 \sigma^2}{m}. \end{aligned} \tag{25}$$

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

$$\begin{aligned} \mathrm{cov}\left[r_s e_s, r_t e_t\right] = &\ E\left[r_s e_s r_t e_t\right] - E\left[r_s e_s\right] E\left[r_t e_t\right] \\ = &\ E\left[e_s e_t\right] E\left[r_s\right] E\left[r_t\right] - E\left[r_s\right] E\left[e_s\right] E\left[r_t\right] E\left[e_t\right] \\ = &\ 0. \end{aligned}$$

When $t = s$, we have:

$$\begin{aligned} \mathrm{cov}[e_t^2, e_t^2] = &\ \mathrm{var}[e_t^2] \\ = &\ E[e_t^4] - (E[e_t^2])^2 \\ = &\ E\left[\varepsilon_t^4 - 4\varepsilon_t^3 \varepsilon_{t-\frac{1}{m}} + 6\varepsilon_t^2 \varepsilon_{t-\frac{1}{m}}^2 - 4\varepsilon_t \varepsilon_{t-\frac{1}{m}}^3 + \varepsilon_{t-\frac{1}{m}}^4\right] - 4\sigma_\varepsilon^4 \\ = &\ 2E[\varepsilon_t^4] + 2\sigma_\varepsilon^4 \\ = &\ 2\omega_\varepsilon^2 + 4\sigma_\varepsilon^4. \end{aligned} \tag{26}$$

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

$$\begin{aligned} \mathrm{cov}\left[e_s^2, e_{s-\frac{1}{m}}^2\right] = &\ \mathrm{cov}\left[e_{s+\frac{1}{m}}^2, e_s^2\right] \\ = &\ \mathrm{cov}\left[\varepsilon_{s+\frac{1}{m}}^2 - 2\varepsilon_{s+\frac{1}{m}} \varepsilon_s + \varepsilon_s^2, \varepsilon_s^2 - 2\varepsilon_s \varepsilon_{s-\frac{1}{m}} + \varepsilon_{s-\frac{1}{m}}^2\right] \\ = &\ \mathrm{var}[\varepsilon_s^2] \\ = &\ \omega_\varepsilon^2. \end{aligned} \tag{27}$$

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

$$\begin{aligned} \mathrm{cov}[r_t e_t, e_s^2] = &\ E[r_t e_t e_s^2] - E[r_t e_t] E[e_s^2] \\ = &\ E[r_t] E[e_t e_s^2] - E[r_t] E[e_t] E[e_s^2] \\ = &\ 0. \end{aligned} \tag{28}$$

From (25) $\sim$ (28), we have:

$$
\begin{aligned}
\operatorname{var}[u_t] &= \operatorname{var}\left[2\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}} + \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
&= 4\operatorname{var}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}\right] + \operatorname{var}\left[\sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] + 4\operatorname{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] \\
&= 4\sum_{i=1}^{m}\sum_{j=1}^{m} \operatorname{cov}\left[r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, r_{t-1+\frac{j}{m}} e_{t-1+\frac{j}{m}}\right] + \sum_{i=1}^{m}\sum_{j=1}^{m} \operatorname{cov}\left[e^2_{t-1+\frac{i}{m}}, e^2_{t-1+\frac{j}{m}}\right] \\
&\quad +4\sum_{i=1}^{m}\sum_{j=1}^{m} \operatorname{cov}\left[r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, e^2_{t-1+\frac{j}{m}}\right] \\
&= 8\sigma^2_\varepsilon\sigma^2 + m(2\omega^2_\varepsilon + 4\sigma^4_\varepsilon) + 2(m-1)\omega^2_\varepsilon \\
&= 8\sigma^2_\varepsilon\sigma^2 + 2(2m-1)\omega^2_\varepsilon + 4m\sigma^4_\varepsilon,
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{cov}[u_t, u_{t+1}] &= \operatorname{cov}\left[2\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}} + \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}, 2\sum_{i=1}^{m} r_{t+\frac{i}{m}} e_{t+\frac{i}{m}} + \sum_{i=1}^{m} e^2_{t+\frac{i}{m}}\right] \\
&= 4\operatorname{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} r_{t+\frac{i}{m}} e_{t+\frac{i}{m}}\right] + 2\operatorname{cov}\left[\sum_{i=1}^{m} r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t+\frac{i}{m}}\right] \\
&\quad +2\operatorname{cov}\left[\sum_{i=1}^{m} r_{t+\frac{i}{m}} e_{t+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}\right] + \operatorname{cov}\left[\sum_{i=1}^{m} e^2_{t-1+\frac{i}{m}}, \sum_{i=1}^{m} e^2_{t+\frac{i}{m}}\right] \\
&= \operatorname{cov}\left[e^2_t, e^2_{t+\frac{1}{m}}\right] \\
&= \omega^2_\varepsilon.
\end{aligned}
$$

It is easy to check that $\operatorname{cov}[u_t, u_{t\pm i}] = 0$ for $i \geq 2$, and hence we have (7).

Derivation of (20a)$\sim$(20c)

Here, we derive the autocovariances of the MA process in (18). They are given by

$$
\begin{aligned}
\gamma_0 = \;& \operatorname{cov}\{\eta_t + \theta_1\eta_{t-1} + d_t - \kappa_1 d_{t-1} + \xi_t + (\theta_u - \kappa_1)\xi_{t-1} - \kappa_1\theta_u\xi_{t-2}, \\
& \eta_t + \theta_1\eta_{t-1} + d_t - \kappa_1 d_{t-1} + \xi_t + (\theta_u - \kappa_1)\xi_{t-1} - \kappa_1\theta_u\xi_{t-2}\} \\
= \;& \sigma^2_\eta + \theta^2_1\sigma^2_\eta + \sigma^2_d + \kappa^2_1\sigma^2_d + \sigma^2_\xi + (\theta_u - \kappa_1)^2\sigma^2_\xi + \kappa^2_1\theta^2_u\sigma^2_\xi \\
= \;& (1+\theta^2_1)\sigma^2_\eta + (1+\kappa^2_1)\sigma^2_d + [1 + (\theta_u - \kappa_1)^2 + \kappa^2_1\theta^2_u]\sigma^2_\xi,
\end{aligned}
$$

$$
\begin{aligned}
\gamma_1 = \;& \operatorname{cov}\{\eta_t + \theta_1\eta_{t-1} + d_t - \kappa_1 d_{t-1} + \xi_t + (\theta_u - \kappa_1)\xi_{t-1} - \kappa_1\theta_u\xi_{t-2}, \\
& \eta_{t-1} + \theta_1\eta_{t-2} + d_{t-1} - \kappa_1 d_{t-2} + \xi_{t-1} + (\theta_u - \kappa_1)\xi_{t-2} - \kappa_1\theta_u\xi_{t-3}\} \\
= \;& \theta_1\sigma^2_\eta - \kappa_1\sigma^2_d + (\theta_u - \kappa_1 - \kappa_1\theta^2_u + \kappa^2_1\theta_u)\sigma^2_\xi
\end{aligned}
$$

$$
\begin{aligned}
\gamma_2 = \;& \operatorname{cov}\{\eta_t + \theta_1\eta_{t-1} + d_t - \kappa_1 d_{t-1} + \xi_t + (\theta_u - \kappa_1)\xi_{t-1} - \kappa_1\theta_u\xi_{t-2}, \\
& \eta_{t-2} + \theta_1\eta_{t-3} + d_{t-2} - \kappa_1 d_{t-3} + \xi_{t-2} + (\theta_u - \kappa_1)\xi_{t-3} - \kappa_1\theta_u\xi_{t-4}\} \\
= \;& -\kappa_1\theta_u\sigma^2_\xi.
\end{aligned}
$$

Because it follows an MA(2) process, the autocovariances of the order greater than 2 is zero.

Derivation of (22)

From (21c), we have $\sigma^2_\xi = \omega^2_\varepsilon/\theta_u$. Substituting this into (21b), we have:

$$
(1+\theta^2_u)\frac{\omega^2_\varepsilon}{\theta_u} = 8\sigma^2\sigma^2_\varepsilon + 2(2m-1)\omega^2_\varepsilon + 4m\sigma^4_\varepsilon.
$$

Multiplying both sides by $\theta_u/\omega^2_\varepsilon$ and rearranging, we have:

$$
\theta^2_u - 2\left[4\frac{\sigma^2\sigma^2_\varepsilon}{\omega^2_\varepsilon} + 2m - 1 + 2m\frac{\sigma^4_\varepsilon}{\omega^2_\varepsilon}\right]\theta_u + 1 = 0.
$$

The two solutions of this quadratic equation for θ_u are given by

$$
\theta_u = A \pm \sqrt{A^2 - 1}, \quad \text{where} \quad A = 4\frac{\sigma^2\sigma^2_\varepsilon}{\omega^2_\varepsilon} + 2m - 1 + 2m\frac{\sigma^4_\varepsilon}{\omega^2_\varepsilon}.
$$

Because $A > 1$ 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 (22).

Derivation of (24a) and (24c)

From (20c) and (21c), we have $\omega_\varepsilon^2 = -\frac{\gamma_2}{\kappa_1}$, which is the first result in (24a). From (3), (4) and (11), we have

$$\sigma_\eta^2 = \frac{2B\omega_1^2}{1+\theta_1^2} \quad \text{and} \quad \sigma_d^2 = \frac{2\sigma^4}{m} + 2mC\omega_1^2, \tag{29}$$

where B and C are as given in (24d). From $\omega_\varepsilon^2 = \theta_u \sigma_\xi^2$ in (21c), we have:

$$\begin{aligned}(1+\theta_u^2 - 2\theta_u\kappa_1 + \kappa_1^2 + \kappa_1^2\theta_u^2)\sigma_\xi^2 &= \left(\tfrac{1}{\theta_u} + \theta_u - 2\kappa_1 + \tfrac{\kappa_1^2}{\theta_u} + \kappa_1^2\theta_u\right)\omega_\varepsilon^2 \\ &= \left[\left(\tfrac{1}{\theta_u} + \theta_u\right)(1+\kappa_1^2) - 2\kappa_1\right]\omega_\varepsilon^2,\end{aligned} \tag{30}$$

and

$$\begin{aligned}(\theta_u - \kappa_1 - \kappa_1\theta_u^2 + \kappa_1^2\theta_u)\sigma_\xi^2 &= \left(1 - \tfrac{\kappa_1}{\theta_u} - \kappa_1\theta_u + \kappa_1^2\right)\omega_\varepsilon^2 \\ &= \left[1+\kappa_1^2 - \left(\tfrac{1}{\theta_u} + \theta_u\right)\kappa_1\right]\omega_\varepsilon^2.\end{aligned} \tag{31}$$

Substituting (29), (30) and (31) into (20a) and (20b), we have:

$$\gamma_0 = 2D\omega_1^2 + 2\frac{1+\kappa_1^2}{m}\sigma^4 + \left[\left(\frac{1}{\theta_u} + \theta_u\right)(1+\kappa_1^2) - 2\kappa_1\right]\omega_\varepsilon^2, \tag{32a}$$

and

$$\gamma_1 = 2E\omega_1^2 - 2\frac{\kappa_1}{m}\sigma^4 - \left[\left(\frac{1}{\theta_u} + \theta_u\right)\kappa_1 - (1+\kappa_1^2)\right]\omega_\varepsilon^2, \tag{32b}$$

where $D = B + m(1+\kappa_1^2)C$, $E = \rho B - m\kappa_1 C$ and $\rho = \theta_1/(1-\theta_1^2)$. From (32), we have:

$$\begin{aligned}\kappa_1\gamma_0 + (1+\kappa_1^2)\gamma_1 &= 2\left[\kappa_1 D + (1+\kappa_1^2)E\right]\omega_1^2 + \left[(1+\kappa_1^2)^2 - 2\kappa_1^2\right]\omega_\varepsilon^2, \\ &= 2\left[\kappa_1 + (1+\kappa_1^2)\rho\right]B\omega_1^2 + (1+\kappa_1^4)\omega_\varepsilon^2, \\ &= \frac{(1-\kappa_1)^3(1+\kappa_1)}{(\log\kappa_1)^2}\omega_1^2 + (1+\kappa_1^4)\omega_\varepsilon^2,\end{aligned} \tag{33}$$

where, to obtain the third equality, we use the alternative expression of ρ explained below (11). From (33), we have:

$$\omega_1^2 = \frac{(\log\kappa_1)^2[\kappa_1\gamma_0 + (1+\kappa_1^2)\gamma_1 - (1+\kappa_1^4)\omega_\varepsilon^2]}{(1-\kappa_1)^3(1+\kappa_1)}.$$

Substituting $\omega_\varepsilon^2 = -\frac{\gamma_2}{\kappa_1}$, we obtain the second result in (24a). Next, note that from (22), we have:

$$\begin{aligned}\frac{1}{\theta_u} + \theta_u &= \frac{1+\theta_u^2}{\theta_u} \\ &= \frac{1+(A-\sqrt{A^2-1})^2}{A-\sqrt{A^2-1}} \\ &= \frac{A+\sqrt{A^2-1} + (A-\sqrt{A^2-1})^2(A+\sqrt{A^2-1})}{(A-\sqrt{A^2-1})(A+\sqrt{A^2-1})} \\ &= 2A.\end{aligned} \tag{34}$$

From (20d) and (21a), we have:

$$c_{RV} = (1-\kappa_1)\left(\sigma^2 + 2m\sigma_\varepsilon^2\right), \quad \text{or} \quad \sigma_\varepsilon^2 = \frac{c_{RV} - (1-\kappa_1)\sigma^2}{2(1-\kappa_1)m}. \tag{35}$$

Substituting σ_ε^2 in (35) into A in (22), we have:

$$
\begin{aligned}
2A = \;& 2\left[\tfrac{4\sigma^2}{\omega_\varepsilon^2}\left(\tfrac{c_{RV}-(1-\kappa_1)\sigma^2}{2(1-\kappa_1)m}\right)+2m-1+\tfrac{2m}{\omega_\varepsilon^2}\left(\tfrac{c_{RV}-(1-\kappa_1)\sigma^2}{2(1-\kappa_1)m}\right)^2\right] \\
= \;& 2\left[\tfrac{2c_{RV}\sigma^2}{(1-\kappa_1)m\omega_\varepsilon^2}-\tfrac{2\sigma^4}{m\omega_\varepsilon^2}+2m-1+\tfrac{c_{RV}^2-2(1-\kappa_1)c_{RV}\sigma^2+(1-\kappa_1)^2\sigma^4}{2(1-\kappa_1)^2m\omega_\varepsilon^2}\right] \\
= \;& \tfrac{4c_{RV}\sigma^2}{(1-\kappa_1)m\omega_\varepsilon^2}-\tfrac{4\sigma^4}{m\omega_\varepsilon^2}+2(2m-1)+\tfrac{c_{RV}^2}{(1-\kappa_1)^2m\omega_\varepsilon^2}-\tfrac{2c_{RV}\sigma^2}{(1-\kappa_1)m\omega_\varepsilon^2}+\tfrac{\sigma^4}{m\omega_\varepsilon^2} \\
= \;& \tfrac{2c_{RV}\sigma^2}{(1-\kappa_1)m\omega_\varepsilon^2}-\tfrac{3\sigma^4}{m\omega_\varepsilon^2}+2(2m-1)+\tfrac{c_{RV}^2}{(1-\kappa_1)^2m\omega_\varepsilon^2}.
\end{aligned}
\tag{36}
$$

From (32a), (34) and (36), we have:

$$
\begin{aligned}
\gamma_0 = \;& 2D\omega_1^2-\frac{(1+\kappa_1^2)}{m}\sigma^4+\frac{2(1+\kappa_1^2)c_{RV}}{(1-\kappa_1)m}\sigma^2 \\
& +2(2m-1)(1+\kappa_1^2)\omega_\varepsilon^2+\frac{(1+\kappa_1^2)c_{RV}^2}{(1-\kappa_1)^2m}-2\kappa_1\omega_\varepsilon^2.
\end{aligned}
\tag{37}
$$

Multiplying both sides in (37) by $m/(1+\kappa_1^2)$ and rearranging, we have:

$$
\sigma^4-\frac{2c_{RV}}{1-\kappa_1}\sigma^2-\frac{c_{RV}^2}{(1-\kappa_1)^2}+\frac{m\left(\gamma_0-2D\omega_1^2+2\kappa_1\omega_\varepsilon^2\right)}{1+\kappa_1^2}-2m(2m-1)\omega_\varepsilon^2=0.
$$

Solving this quadratic equation for σ^2, we have:

$$
\sigma^2=\frac{c_{RV}}{1-\kappa_1}\pm\sqrt{\frac{2c_{RV}^2}{(1-\kappa_1)^2}+2m(2m-1)\omega_\varepsilon^2-\frac{m\left(\gamma_0-2D\omega_1^2+2\kappa_1\omega_\varepsilon^2\right)}{(1+\kappa_1^2)}}. \tag{38}
$$

From $\sigma_\varepsilon^2>0$, $\kappa_1<1$ and (35), we must have $\frac{c_{RV}}{1-\kappa_1}>\sigma^2$. Hence, the sign of the second term in (38) is negative. From (35) and (38), we have:

$$
\sigma_\varepsilon^2=\frac{1}{2m}\sqrt{\frac{2c_{RV}^2}{(1-\kappa_1)^2}+2m(2m-1)\omega_\varepsilon^2-\frac{m\left(\gamma_0-2D\omega_1^2+2\kappa_1\omega_\varepsilon^2\right)}{(1+\kappa_1^2)}}. \tag{39}
$$

From (38) and (39), we obtain (24b) and (24c).

Appendix B: Results for the Two–factor Case

Let $\phi_1 = \kappa_1 + \kappa_2$, $\phi_2 = -\kappa_1\kappa_2$, $\pi_1 = 1 + \phi_1^2 + \phi_2^2$, $\pi_2 = \phi_1(1-\phi_2)$ and $\omega_2 \neq 0$ throughout Appendix B.

Let η_t and ξ_t be denoted by the state variables α_t and β_t, respectively. From (8), (12) and (15), we can express the NCRV in the following state space form:

Observation equation

$$RV_t^* = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_t \\ IV_{t-1} \\ u_t \\ \alpha_t \\ \alpha_{t-1} \\ \beta_t \end{bmatrix} + d_t, \tag{40a}$$

State equation

$$\begin{bmatrix} IV_t \\ IV_{t-1} \\ u_t \\ \alpha_t \\ \alpha_{t-1} \\ \beta_t \end{bmatrix} = \begin{bmatrix} c_{IV} \\ 0 \\ c_u \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 & 0 & \theta_1 & \theta_2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_u \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_{t-1} \\ IV_{t-2} \\ u_{t-1} \\ \alpha_{t-1} \\ \alpha_{t-2} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}, \tag{40b}$$

where the mean vector and variance matrix of $(d_t, \eta_t, \xi_t)'$ are as given in (16c).

Autocovariance functions

In the two–factor case, by applying the results in Granger and Morris (1976), we can show that the NCRV follows an ARMA(2, 3) process:

$$(1 - \phi_1 L - \phi_2 L^2) RV_t^* = c_{RV} + (1 + \delta_1 L + \delta_2 L^2 + \delta_3 L^3)\tau_t, \ \tau_t \sim WN(0, \sigma_\tau^2). \tag{41}$$

The same RV_t^* can alternatively be expressed as:

$$\begin{aligned} (1 - \phi_1 L - \phi_2 L^2) RV_t^* = \ & c_{IV} + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2} + d_t - \phi_1 d_{t-1} - \phi_2 d_{t-1} \\ & + (1 - \phi_1 - \phi_2)c_u + \xi_t + (\theta_u - \phi_1)\xi_{t-1} \\ & - (\phi_2 + \phi_1\theta_u)\xi_{t-2} - \phi_2\theta_u\xi_{t-3}, \end{aligned} \tag{42}$$

The autocovariance functions of the MA process in (41) are given as:

$$\begin{aligned} &\gamma_0 = (1 + \delta_1^2 + \delta_2^2 + \delta_3^2)\sigma_\tau^2, \quad \gamma_1 = (\delta_1 + \delta_1\delta_2 + \delta_2\delta_3)\sigma_\tau^2, \\ &\gamma_2 = (\delta_2 + \delta_1\delta_3)\sigma_\tau^2, \qquad\qquad\quad \gamma_3 = \delta_3\sigma_\tau^2, \end{aligned} \tag{43}$$

and $\gamma_j = 0$, for $j \geq 4$. Furthermore, some calculations lead us to the following autocovariance functions of the MA process in (42):

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_\eta^2 + \pi_1\sigma_d^2 + \left[\pi_1\left(\frac{1}{\theta_u} + \theta_u\right) - 2\pi_2\right]\theta_u\sigma_\xi^2, \tag{44a}$$

$$\gamma_1 = (\theta_1 + \theta_1\theta_2)\sigma_\eta^2 - \pi_2\sigma_d^2 + \left[\pi_1 - \phi_2 - \pi_2\left(\frac{1}{\theta_u} + \theta_u\right)\right]\theta_u\sigma_\xi^2, \tag{44b}$$

$$\gamma_2 = \theta_2\sigma_\eta^2 - \phi_2\sigma_d^2 - \left[\phi_2\left(\frac{1}{\theta_u} + \theta_u\right) + \pi_2\right]\theta_u\sigma_\xi^2, \tag{44c}$$

$$\gamma_3 = -\phi_2\theta_u\sigma_\xi^2, \tag{44d}$$

and $\gamma_j = 0$, for $j \geq 4$. By equating the means of the MA processes on the right hand sides in (41) and (42), we obtain:

$$c_{RV} = c_{IV} + (1 - \phi_1 - \phi_2)c_u. \tag{44e}$$

As in the one–factor case, the number of state space form parameters is greater than the number of ARMA reduced form parameters. This implies that the state space form in (40) is not identified. However, we show that the state space form parameters are expressed as functions of the underlying continuous SV model parameters σ^2, ω_1^2, ω_2^2, σ_ε^2 and ω_ε^2, which are uniquely identified from the ARMA reduced form in (41).

Identification of state space form parameters

Here we show that the parameters σ^2, ω_1^2, ω_2^2, σ_ε^2 and ω_ε^2 are uniquely identified from the reduced form parameters $c_{RV}, \kappa_1, \kappa_2$, and γ_j for $j = 0 \sim 3$.

As in the one–factor case, we can uniquely solve Equations (44a)∼(44e) with respect to σ^2, ω_1^2, ω_2^2, σ_ε^2 and ω_ε^2 as:

$$\omega_\varepsilon^2 = -\frac{\gamma_3}{\phi_2}, \quad \omega_1^2 = \frac{(\log\kappa_1)^2[(\alpha_1\beta_0 - \alpha_0\beta_1) - \kappa_2(\alpha_0\phi_1 + \alpha_2\beta_0)]}{(1-\kappa_1)^2(\kappa_1 - \kappa_2)(\alpha_1\phi_1 - \alpha_2\beta_1)}, \tag{45a}$$

$$\omega_2^2 = \frac{(\log\kappa_2)^2[(\alpha_1\beta_0 - \alpha_0\beta_1) - \kappa_1(\alpha_0\phi_1 + \alpha_2\beta_0)]}{(1-\kappa_2)^2(\kappa_2 - \kappa_1)(\alpha_1\phi_1 + \alpha_2\beta_1)}, \tag{45b}$$

$$\sigma_\varepsilon^2 = \frac{1}{2m}\sqrt{\frac{2c_{RV}^2}{(1-\phi_1-\phi_2)^2} + 2m(2m-1)\omega_\varepsilon^2 + H}, \tag{45c}$$

and

$$\sigma = \frac{c_{RV}}{1-\phi_1-\phi_2} - 2m\sigma_\varepsilon^2, \tag{45d}$$

where

$$\begin{aligned} &\alpha_0 = \pi_2\gamma_0 + \pi_1\gamma_1 - [\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2, \quad \alpha_1 = \pi_1(1 + \phi_1^2 - \phi_2) - 2\pi_2^2, \\ &\alpha_2 = 2\phi_2\pi_2 + \phi_1\pi_1, \\ &\beta_0 = \pi_2\gamma_2 - \phi_2\gamma_1 + (\phi_1^2 + \phi_2)(1 + \phi_2^2 - \phi_2)\omega_\varepsilon^2, \quad \beta_1 = \phi_2^2 - \phi_1^2 - \phi_2, \end{aligned} \tag{46}$$

$$H = \frac{m}{\phi_2}\left[\gamma_2 + \sum_{j=1}^{2}(2\phi_2 C_{1,j} + \phi_1 C_{2,j} - C_{3,j} + 2m\phi_2 C_{4,j})\omega_j^2 + \pi_2\omega_\varepsilon^2\right], \tag{47}$$

$$\begin{aligned} &C_{1,j} \equiv \frac{\kappa_j - \log\kappa_j - 1}{(\log\kappa_j)^2}, \quad C_{2,j} \equiv \frac{(1-\kappa_j)^2}{(\log\kappa_j)^2}, \quad C_{3,j} \equiv \frac{\kappa_j(1-\kappa_j)^2}{(\log\kappa_j)^2}, \quad \text{and} \\ &C_{4,j} = \frac{2(\kappa_j^{\frac{1}{m}} - \log\kappa_j^{\frac{1}{m}} - 1)}{(\log\kappa_j)^2} \quad \text{for} \quad j = 1, 2. \end{aligned} \tag{48}$$

In what follows, we derive the results in (45a) – (45c).

From $\theta_u\sigma_\xi^2 = \omega_\varepsilon^2$ in (21c) and $\gamma_3 = -\phi_2\theta_u\sigma_\xi^2$ in (44d), we have $\omega_\varepsilon^2 = -\frac{\gamma_3}{\phi_2}$, which is the first result in (45a). Furthermore, from (3), (4) and (13), after some calculations, it follows that:

$$\sigma_\eta^2 = \frac{2B_1\omega_1^2}{1+\theta_1^2+\theta_2^2} + \frac{2B_2\omega_2^2}{1+\theta_1^2+\theta_2^2} \quad \text{and} \quad \sigma_d^2 = \frac{2\sigma^4}{m} + 2mC_{4,1}\omega_1^2 + 2mC_{4,2}\omega_2^2, \tag{49}$$

where

$$B_j = \pi_1 C_{1,j} - \pi_2 C_{2,j} - \phi_2 C_{3,j} \quad \text{for} \quad j = 1, 2. \tag{50}$$

Substituting (49) into the autocovariance functions in (44) and rearranging, we have:

$$\gamma_0 = 2D_1\omega_1^2 + 2D_2\omega_2^2 + 2\frac{\pi_1}{m}\sigma^4 + \left[\pi_1\left(\frac{1}{\theta_u} + \theta_u\right) - 2\pi_2\right]\omega_\varepsilon^2, \tag{51a}$$

$$\gamma_1 = 2E_1\omega_1^2 + 2E_2\omega_2^2 - 2\frac{\pi_2}{m}\sigma^4 - \left[\pi_2\left(\frac{1}{\theta_u} + \theta_u\right) - (\pi_1 - \phi_2)\right]\omega_\varepsilon^2, \tag{51b}$$

$$\gamma_2 = 2F_1\omega_1^2 + 2F_2\omega_2^2 - 2\frac{\phi_2}{m}\sigma^4 - \left[\phi_2\left(\frac{1}{\theta_u} + \theta_u\right) + \pi_2\right]\omega_\varepsilon^2, \tag{51c}$$

where $D_j = B_j + m\pi_1 C_{4,j}$, $E_j = \rho_1 B_j - m\pi_2 C_{4,j}$, $F_j = \rho_2 B_j - m\phi_2 C_{4,j}$ for $j = 1, 2$, $\rho_1 = (\theta_1 + \theta_1\theta_2)/(1 + \theta_1^2 + \theta_2^2)$ and $\rho_2 = \theta_2/(1 + \theta_1^2 + \theta_2^2)$. Hence, we have

$$\pi_2\gamma_0 + \pi_1\gamma_1 = (\pi_2 + \rho_1\pi_1)(2B_1\omega_1^2 + 2B_2\omega_2^2) + [\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2, \tag{52a}$$

and

$$\pi_2\gamma_2 - \phi_2\gamma_1 = (\rho_2\pi_2 - \rho_1\phi_2)(2B_1\omega_1^2 + 2B_2\omega_2^2) - [\phi_2(\pi_1 - \phi_2) + \pi_2^2]\omega_\varepsilon^2. \tag{52b}$$

Noting that ρ_1 and ρ_2 can be expressed as in (14) (see the explanations below (14)), we have:

$$\begin{aligned}\rho_1 &= \frac{-\pi_2\text{var}[IV_t] + (1 + \phi_1^2 - \phi_2)\text{cov}[IV_t, IV_{t-1}] - \phi_1\text{cov}[IV_t, IV_{t-2}]}{(1 + \theta_1^2 + \theta_2^2)\sigma_\eta^2} \\ &= \frac{\sum_{j=1}^2[-2\pi_2 C_{1,j} + (1 + \phi_1^2 - \phi_2)C_{2,j} - \phi_1 C_{3,j}]\omega_j^2}{2B_1\omega_1^2 + 2B_2\omega_2^2},\end{aligned} \tag{53a}$$

and

$$\begin{aligned}\rho_2 &= \frac{-\phi_2\text{var}[IV_t] - \phi_1\text{cov}[IV_t, IV_{t-1}] + \text{cov}[IV_t, IV_{t-2}]}{(1 + \theta_1^2 + \theta_2^2)\sigma_\eta^2} \\ &= \frac{\sum_{j=1}^2(-2\phi_2 C_{1,j} - \phi_1 C_{2,j} + C_{3,j})\omega_j^2}{2B_1\omega_1^2 + 2B_2\omega_2^2}.\end{aligned} \tag{53b}$$

Substituting B_j in (50), ρ_1 and ρ_2 in (53) into (52), we have:

$$\begin{aligned}\pi_2\gamma_0 + \pi_1\gamma_1 = \quad & 2\pi_2\sum_{j=1}^{2}(\pi_1 C_{1,j} - \pi_2 C_{2,j} - \phi_2 C_{3,1})\omega_j^2 \\ & +\pi_1\sum_{j=1}^{2}[-2\pi_2 C_{1,j} + (1 + \phi_1^2 - \phi_2)C_{2,j} - \phi_1 C_{3,j}]\omega_j^2 \\ & +[\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2 \\ = \quad & \sum_{j=1}^{2}\{[-2\pi_2^2 + \pi_1(1 + \phi_1^2 - \phi_2)]C_{2,j} - (2\phi_2\pi_2 + \phi_1\pi_1)C_{3,j}\}\omega_j^2 \\ & +[\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2,\end{aligned} \tag{54}$$

and

$$\begin{aligned}\pi_2\gamma_2 - \phi_2\gamma_1 = \quad & \pi_2\sum_{j=1}^{2}(-2\phi_2 C_{1,j} - \phi_1 C_{2,j} + C_{3,j})\omega_j^2 \\ & -\phi_2\sum_{j=1}^{2}[-2\pi_2 C_{1,j} + (1 + \phi_1^2 - \phi_2)C_{2,j} - \phi_1 C_{3,j}]\omega_j^2 \\ & -[\phi_2(\pi_1 - \phi_2) + \pi_2^2]\omega_\varepsilon^2 \\ = \quad & \sum_{j=1}^{2}\{[-\phi_1\pi_2 - \phi_2(1 + \phi_1^2 - \phi_2)]C_{2,j} + (\pi_2 + \phi_1\phi_2)C_{3,j}\}\omega_j^2 \\ & -[\phi_2(\pi_1 - \phi_2) + \pi_2^2]\omega_\varepsilon^2 \\ = \quad & \sum_{j=1}^{2}\{(\phi_2^2 - \phi_1^2 - \phi_2)C_{2,j} + \phi_1 C_{3,j}\}\omega_j^2 \\ & -(\phi_1^2 + \phi_2)(\phi_2^2 - \phi_2 + 1)\omega_\varepsilon^2.\end{aligned} \tag{55}$$

We can regard (54) and (55) as the following system of two equations for ω_1^2 and ω_2^2:

$$\begin{aligned}\alpha_0 &= (\alpha_1 C_{2,1} - \alpha_2 C_{3,1})\omega_1^2 + (\alpha_1 C_{2,2} - \alpha_2 C_{3,2})\omega_2^2 \\ \beta_0 &= (\beta_1 C_{2,1} + \phi_1 C_{3,1})\omega_1^2 + (\beta_1 C_{2,2} + \phi_1 C_{3,2})\omega_2^2,\end{aligned} \tag{56}$$

where α_0, α_1, α_2, β_0 and β_1 are as given in (46). Solving (56), we have

$$\begin{aligned}\omega_1^2 = &\ \frac{(\alpha_1\beta_0 - \alpha_0\beta_1)C_{2,2} - (\alpha_2\beta_0 + \alpha_0\phi_1)C_{3,2}}{(\alpha_1\phi_1 + \alpha_2\beta_1)C_{2,2}C_{3,1} - (\alpha_1\phi_1 + \alpha_2\beta_1)C_{2,1}C_{3,2}} \\ = &\ \frac{(\log\kappa_1)^2[(\alpha_1\beta_0 - \alpha_0\beta_1) - \kappa_2(\alpha_2\beta_0 + \alpha_0\phi_1)]}{(1-\kappa_1)^2(\kappa_1-\kappa_2)(\alpha_1\phi_1 + \alpha_2\beta_1)},\end{aligned}$$

and

$$\begin{aligned}\omega_2^2 = &\ \frac{(\alpha_0\beta_1 - \alpha_1\beta_0)C_{2,1} + (\alpha_0\phi_1 + \alpha_2\beta_0)C_{3,1}}{(\alpha_1\phi_1 + \alpha_2\beta_1)C_{2,2}C_{3,1} - (\alpha_1\phi_1 + \alpha_2\beta_1)C_{3,2}C_{2,1}} \\ = &\ \frac{(\log\kappa_2)^2[(\alpha_0\beta_1 - \alpha_1\beta_0) + \kappa_1(\alpha_0\phi_1 + \alpha_2\beta_0)]}{(1-\kappa_2)^2(\kappa_1-\kappa_2)(\alpha_1\phi_1 + \alpha_2\beta_1)}.\end{aligned}$$

From (44e) and (21a), we have:

$$\sigma_\varepsilon^2 = \frac{c_{RV} - (1-\phi_1-\phi_2)\sigma^2}{2(1-\phi_1-\phi_2)m}. \tag{57}$$

Substituting σ_ε^2 in (57) into A in (22), we have:

$$2A = \frac{2c_{RV}\sigma^2}{(1-\phi_1-\phi_2)m\omega_\varepsilon^2} - \frac{3\sigma^4}{m\omega_\varepsilon^2} + 2(2m-1) + \frac{c_{RV}^2}{(1-\phi_1-\phi_2)^2 m\omega_\varepsilon^2}. \tag{58}$$

From (34), (51c) and (58), we have:

$$\begin{aligned}\gamma_2 = &\ 2\omega_1^2(\rho_2 B_1 - m\phi_2 C_{4,1}) + 2\omega_2^2(\rho_2 B_2 - m\phi_2 C_{4,2}) - 2\sigma^4\tfrac{\phi_2}{m} \\ &- \left\{\phi_2\left[\tfrac{2c_{RV}\sigma^2}{(1-\phi_1-\phi_2)m\omega_\varepsilon^2} - \tfrac{3\sigma^4}{m\omega_\varepsilon^2} + 2(2m-1) + \tfrac{c_{RV}^2}{(1-\phi_1-\phi_2)^2 m\omega_\varepsilon^2}\right] + \pi_2\right\}\omega_\varepsilon^2 \\ = &\ \rho_2(2B_1\omega_1^2 + 2B_2\omega_2^2) - \phi_2(2m\omega_1^2 C_{4,1} + 2m\omega_2^2 C_{4,2}) \\ &+\sigma^4\tfrac{\phi_2}{m} - \sigma^2\tfrac{2\phi_2 c_{RV}}{(1-\phi_1-\phi_2)m} - 2\phi_2\omega_\varepsilon^2(2m-1) - \tfrac{\phi_2 c_{RV}^2}{(1-\phi_1-\phi_2)^2 m} - \pi_2\omega_\varepsilon^2 \\ = &\ \sum_{j=1}^{2}(-2\phi_2 C_{1,j} - \phi_1 C_{2,j} + C_{3,j} - 2m\phi_2 C_{4,j})\omega_j^2 \\ &+\sigma^4\tfrac{\phi_2}{m} - \sigma^2\tfrac{2\phi_2 c_{RV}}{(1-\phi_1-\phi_2)m} - 2\phi_2\omega_\varepsilon^2(2m-1) - \tfrac{\phi_2 c_{RV}^2}{(1-\phi_1-\phi_2)^2 m} - \pi_2\omega_\varepsilon^2.\end{aligned} \tag{59}$$

Multiplying both sides in (59) by m/ϕ_2 and rearranging, we have:

$$\sigma^4 - \frac{2c_{RV}}{1-\phi_1-\phi_2}\sigma^2 - \frac{c_{RV}^2}{(1-\phi_1-\phi_2)^2} - 2\omega_\varepsilon^2 m(2m-1) - H,$$

where H is as given in (47). Solving the quadratic equation for σ^2, and by the same argument as used in (39), we have:

$$\sigma^2 = \frac{c_{RV}}{1-\phi_1-\phi_2} - \sqrt{\frac{2c_{RV}^2}{(1-\phi_1-\phi_2)^2} + 2m(2m-1)\omega_\varepsilon^2 + H}, \tag{60}$$

and

$$\sigma_\varepsilon^2 = \frac{1}{2m}\sqrt{\frac{2c_{RV}^2}{(1-\phi_1-\phi_2)^2} + 2m(2m-1)\omega_\varepsilon^2 + H}. \tag{61}$$

From (60) and (61), we have (45c) and (45d).

Finally, we summarize direct and indirect approaches for estimating the parameters in the two factor case.

Summary of the indirect approach

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

Step 2 Estimate the unrestricted ARMA(2, 3) model in (41) by QML estimation assuming Gaussian innovations.

Step 3 Given the estimates of $c_{RV}^{(m)}$, κ_1, κ_2,[7] $\delta_1^{(m)}$, $\delta_2^{(m)}$, $\delta_3^{(m)}$ and σ_τ^2 obtained in Step 2, calculate the first four autocovariances of the MA process, namely, $\gamma_j^{(m)}$, $j = 0 \sim 3$ as in (43).

Step 4 Given the estimates of $c_{RV}^{(m)}$, κ_1, κ_2 and $\gamma_j^{(m)}$, $j = 0 \sim 3$, obtained in Steps 2 and 3, estimate ω_ε^2, σ_ε^2, ω_1^2, ω_2^2 and σ^2, applying the results in (45a) – (45c).

Summary of the direct approach

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

Step 2 Given κ_1, κ_2, σ^2, ω_1^2, ω_2^2, σ_ε^2 and ω_ε^2, calculate c_{IV}, θ_1, θ_2, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ according to (3), (13) and (22).

Step 3 With the c_{IV}, θ_1, θ_2, σ_η^2, $c_u^{(m)}$, $\theta_u^{(m)}$, $\sigma_\xi^{2(m)}$ and $\sigma_d^{2(m)}$ obtained in Step 2, calculate the Gaussian log-likelihood of the state space form in (40a) – (40b), for RV_t^*.

Stpe 4 Maximize the log-likelihood obtained in Stpe 3 with respect to the seven parameters, κ_1, κ_2, σ^2, ω_1^2, ω_2^2, σ_ε^2 and ω_ε^2 to obtain the QML estimates.

[7] These can be obtained from the estimates of ϕ_1 and ϕ_2. See footnote 2.

References

Anderson BDO, Moore JB. 1979. *Optimal Filtering*. Prentice-Hall: Englewood Cliffs, NJ.

Andersen TG. 1994. Stochastic autoregressive volatility: a framework for volatility modeling. *Mathematical Finace* **4**: 75–102.

Andersen TG, Bollerslev T. 1998. Deutche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* **53**: 219–265.

Andersen TG, Bollerslev T, Diebold FX, Ebens H. 2001. The distribution of stock return volatility. *Journal of Financial Economics* **61**: 43–76.

Andersen TG, Bollerslev T, Diebold FX, Labys P. 2001. The distribution of exchange rate volatility, *Journal of the American Statistical Association* **92**: 42–55.

Bandi FM, Russell JR. 2006. Separating microstructure noise from volatility. *Journal of Financial Economics* **79**: 655–692.

Bandi FM, Russell JR. 2008. Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* **75**: 339–369.

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

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

Barndorff-Nielsen OE, Hansen PR, Lunde A, Shephard N. 2008. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica* **76**: 1481–1536.

Beine M, Lahaye J, Laurent S, Neely CJ, Palm FC. 2007. Central bank intervention and exchange rate volatility, its continuous and jump components. *International Journal of Finance and Economics* **12**: 201–223.

Bollerslev T, Domowitz I. 1993. Trading patterns and prices in the interbank foreign exchange market. *Journal of Finance* **48**: 1421–1443.

Bollerslev T, Engle RF, Nelson DB. 1994. ARCH models. 2959–3038. In *Handbook of Econometrics* **4**, Engle RF, McFadden DL(eds). Elsevier: NY.

Durbin J, Koopman SJ. 2001. *Time Series Analysis by State Space Methods*. Oxford University Press: NY.

Ghysels E, Harvey AC, Renault E. 1996. Stochastic volatility. 119–192. In *Handbook of Statistics* **14**, Maddala GS, Rao CR(eds). Elsevier: NY.

Granger CWJ, Morris MJ. 1976. Time series modelling and interpretation. *Journal of the Royal Statistical Society. Series A (General)* **139**: 246–257.

Hamilton DJ. 1994. *Time Series Analysis*. Princeton University Press: NJ.

Hansen PR, Lunde A. 2006. Realized variance and market microstructure noise. *Journal of Business and 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.

Meddahi N, Renault E. 2004. Temporal aggregation of volatility models. *Journal of Econometrics* **119**: 355–379.

Nelson DB. 1990. ARCH models as diffusion approximations. *Journal of Econometrics* 45: 7–39.

Palm FC. 1996. GARCH models of volatility. 209–240. In *Handbook of Statistics* **14** Maddala GS, Rao, CR(eds). Elsevier: NY.

Shephard N. 2005. *Stochastic Volatility, Selected Readings.* Oxford University Press: Oxford.

Zhang L, Mykland P, 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.

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

Zivot E. 2008. Practical issues in the analysis of GARCH models (forthcoming). In *Handbook of Financial Time Series*, Andersen TG, Davis RA, Kreiss J-P, Mikosch T(eds). Springer.

Table 1: Descriptive statistics of the NCRV

	One–minute NCRV	Five–minute NCRV
m	1440	288
Mean	0.5317	0.4039
Variance	0.0629	0.0620
SD	0.2507	0.2490
AC(1)	0.4794	0.4177
AC(2)	0.3628	0.3292
AC(3)	0.3261	0.2819
AC(4)	0.3294	0.2595
AC(5)	0.3246	0.2577

Note: The table reports the sample mean (Mean), sample variance (Variance) and sample standard deviation (SD) of the RV series calculated with different m, where m is the number of intervals for each NCRV series. AC(k) denotes the sample autocorrelation of order k.

Figure 1: Daily returns of the yen/dollar exchange rate

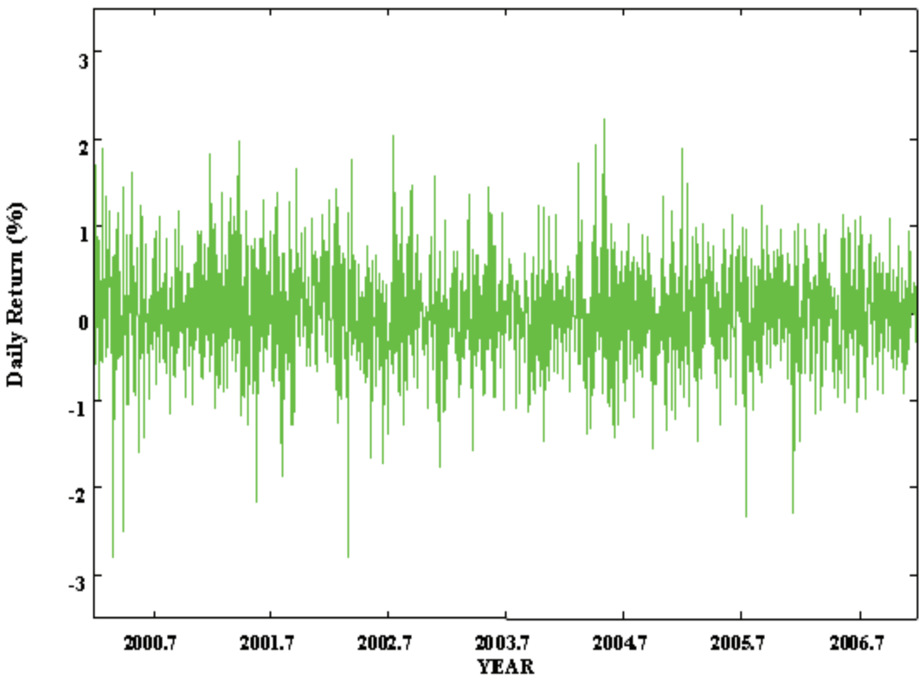


Table 2: Estimates of SV model parameters

	One–factor case		Two–factor case	
	One–minute	Five–minute	One–minute	Five–minute
$\widehat{\kappa}_1$	0.9301	0.8849	0.9825	0.9798
	(0.0415)	(0.0410)	(0.0187)	(0.0143)
$\widehat{\kappa}_2$	-	-	0.3241	0.6113
	-	-	(0.2321)	(0.1435)
$\widehat{\sigma}^2$	0.2857	0.3466	0.2960	0.3445
	(0.0247)	(0.0178)	(0.0417)	(0.0325)
$\widehat{\omega}_1^2$	0.0300	0.0279	0.0229	0.0145
	(0.0111)	(0.0083)	(0.0146)	(0.0064)
$\widehat{\omega}_2^2$	-	-	0.0271	0.0192
	-	-	(0.0213)	(0.0067)
$\widehat{\sigma}_\varepsilon^2$	0.0000861	0.0001002	0.0000839	0.0001043
	(0.0000100)	(0.0000029)	(0.0000157)	(0.0000062)
$\widehat{\omega}_\varepsilon^2$	0.0000059	0.0000296	0.0000039	0.0000263
	(0.0000009)	(0.0000043)	(0.0000019)	(0.0000048)
L	240.06713	181.18724	262.11225	193.84332

Note: L is the log-likelihood. The robust standard errors are in parentheses.

Table 3: Estimates of state space model parameters

	One–factor case		Two–factor case	
	One–min	Five–minute	One–minute	Five–minute
$\widehat{c}_{IV}$	0.0200	0.0399	0.0035	0.0027
$\widehat{\phi}_1$	0.9301	0.8849	1.3066	1.5911
$\widehat{\phi}_2$	-	-	-0.3184	-0.5989
$\widehat{\theta}_1$	0.2679	0.2677	-0.6280	-0.6512
$\widehat{\theta}_2$	-	-	-0.2196	-0.2421
$\widehat{\sigma}_\eta^2$	0.0025	0.0038	0.0159	0.0081
m	1440	288	1440	288
$\widehat{c}_u^{(m)}$	0.2479	0.0577	0.2417	0.0601
$\widehat{\theta}_u^{(m)}$	0.0002	0.0009	0.0002	0.0009
$\widehat{\sigma}_\xi^{2(m)}$	0.0340	0.0343	0.0228	0.0305
$\widehat{\sigma}_d^{2(m)}$	0.0002	0.0010	0.0002	0.0011

Table 4: Estimates of some important values

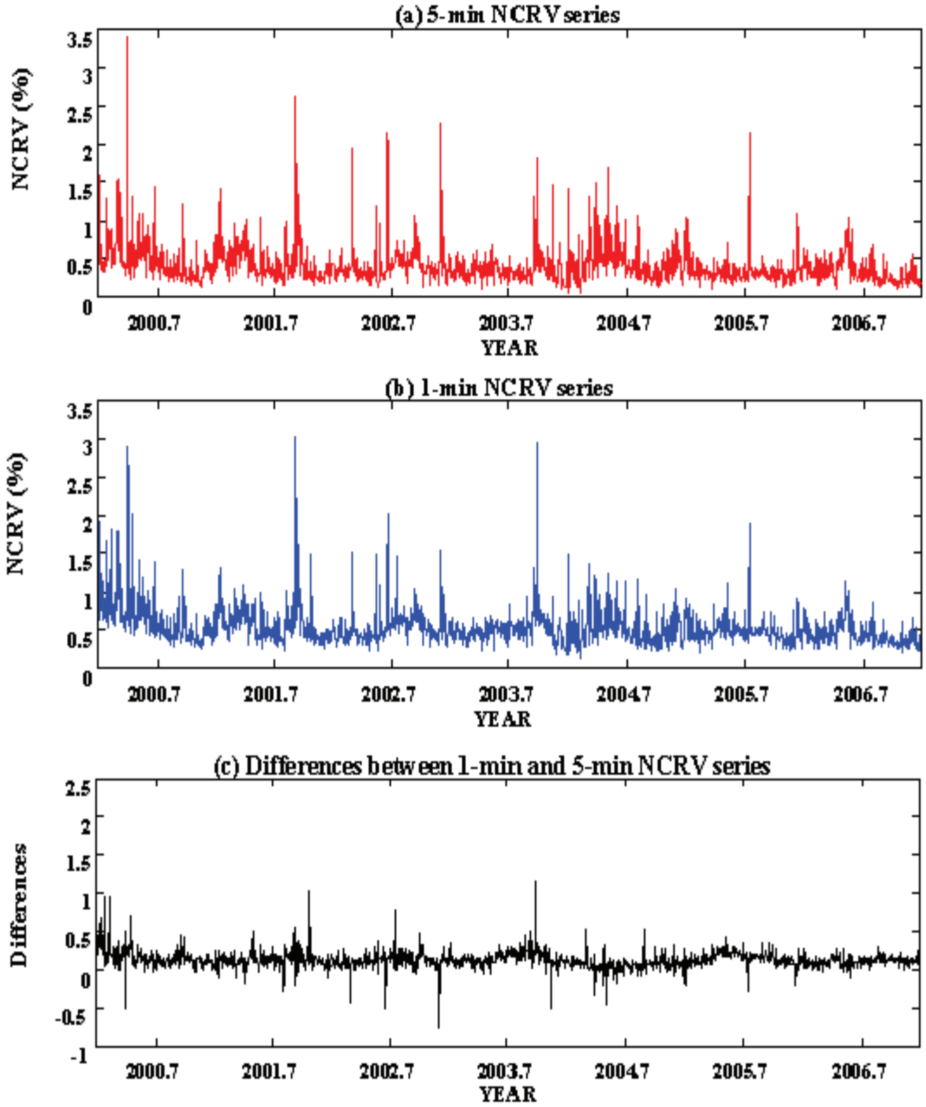
	One–factor case		Two–factor case	
	One–minute	Five–minute	One–minute	Five–minute
$\widehat{\text{var}[IV_t]}$	0.0293	0.0268	0.0420	0.0308
$\text{corr}\widehat{[IV_t, IV_{t-1}]}$	0.9531	0.9225	0.7675	0.8502
$\text{corr}\widehat{[IV_t, IV_{t-2}]}$	0.8865	0.8163	0.6011	0.6900
$\widehat{\text{var}[u_t^{(m)}]}$	0.0340	0.0343	0.0305	0.0305
$\widehat{\text{var}[IV_t]}/\text{var}\widehat{[RV_t^{*(m)}]}$	0.4618	0.4313	0.6467	0.4941
$\widehat{\text{var}[u_t^{(m)}]}/\text{var}\widehat{[RV_t^{*(m)}]}$	0.5358	0.5521	0.3504	0.4890
$\widehat{\sigma}_\eta^2/(\widehat{\sigma}_\eta^2 + \widehat{\sigma}_\xi^{(m)2} + \widehat{\sigma}_d^{(m)2})$	0.0686	0.0962	0.4097	0.2047
$\widehat{\sigma}_\xi^{(m)2}/(\widehat{\sigma}_\eta^2 + \widehat{\sigma}_\xi^{(m)2} + \widehat{\sigma}_d^{(m)2})$	0.9271	0.8775	0.5854	0.7686

Note: $\text{var}\widehat{[RV_t^{*(m)}]} = \widehat{\text{var}[IV_t]} + \widehat{\text{var}[u_t^{(m)}]} + \widehat{\text{var}[d_t^{(m)}]}$.

Table 5: Mean, max and min of the ratios of MN components to NCRV

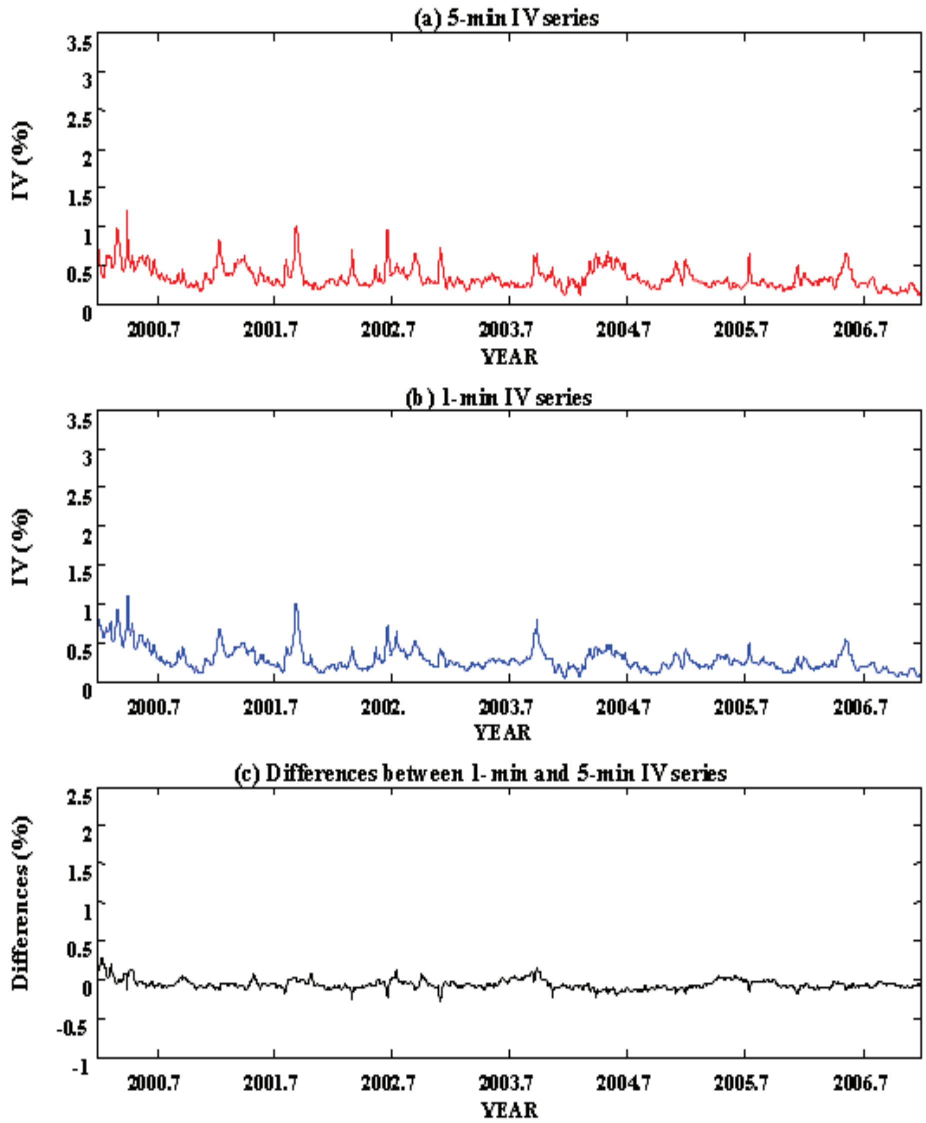
	One–factor case		Two–factor case	
	One–minute	Five–minute	One–minute	Five–minute
mean of $\widehat{R}_t^{(m)}$	0.4646	0.4594	0.4755	0.0753
mean of $\lvert\widehat{R}_t^{(m)}\rvert$	0.4708	0.4659	0.4770	0.2080
$\max\{\widehat{R}_t^{(m)}\}$	0.8357	0.8324	1.0454	0.6574
$\min\{\widehat{R}_t^{(m)}\}$	-0.5804	-0.5848	-0.4192	-3.7323
$\max\{\lvert\widehat{R}_t^{(m)}\rvert\}$	0.8357	0.8324	1.0454	3.7323
$\min\{\lvert\widehat{R}_t^{(m)}\rvert\}$	0.00398	0.00053	0.00773	0.00003

Figure 2: 1–minute and 5–minute NCRV series



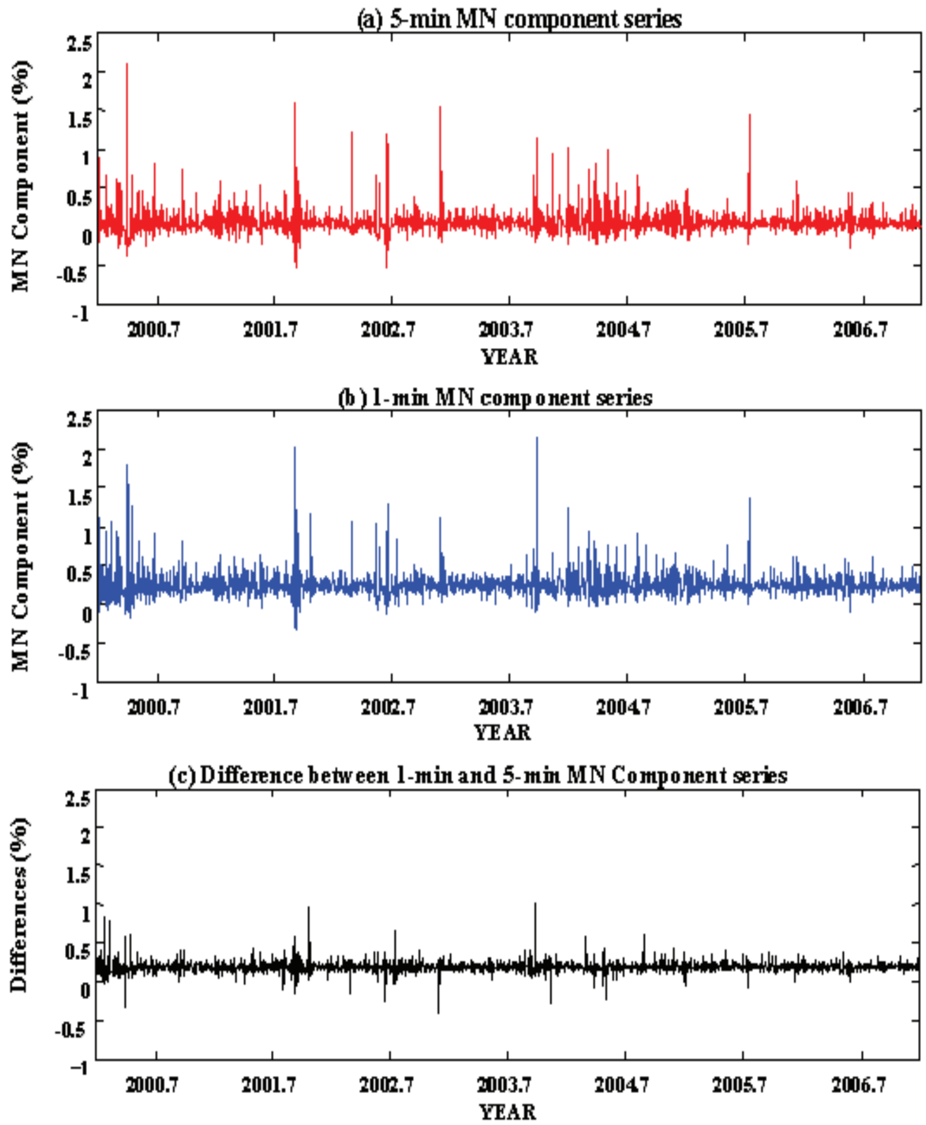
Note: Figure 2(c) displays 1–minute NCRV series minus 5–minute NCRV series.

Figure 3: Smoothed series of IV in the one–factor case



Note: Figure 3(c) displays 1–minute IV series minus 5–minute IV series.

Figure 4: Smoothed series of MN component in the one–factor case



Note: Figure 4(c) displays 1–minute MN component series minus 5–minute MN component series.

Figure 5: Smoothed series of discretization error in the one–factor case

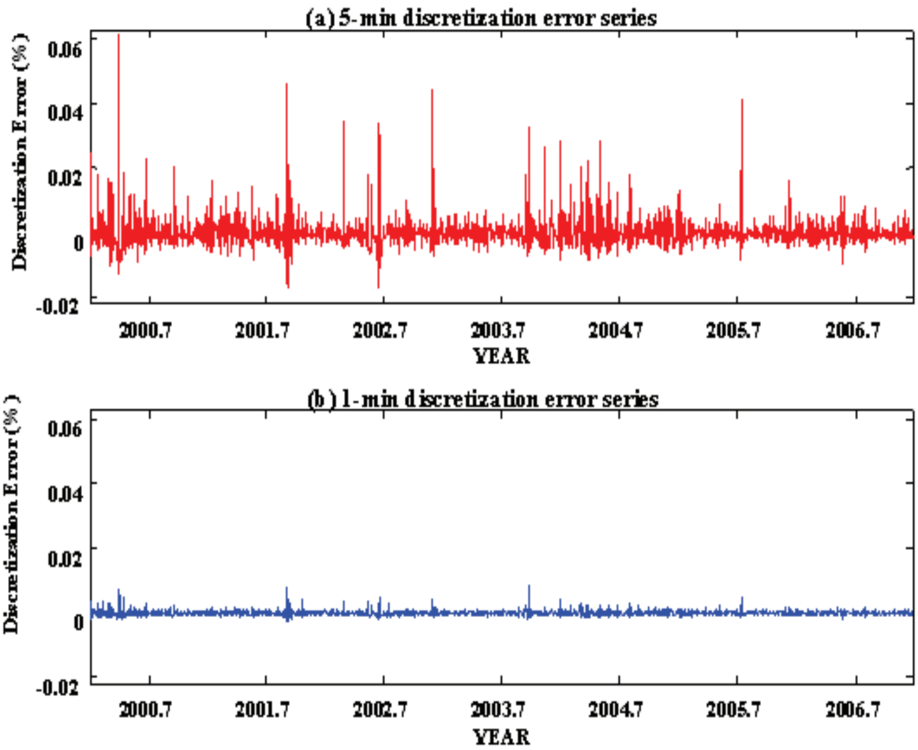
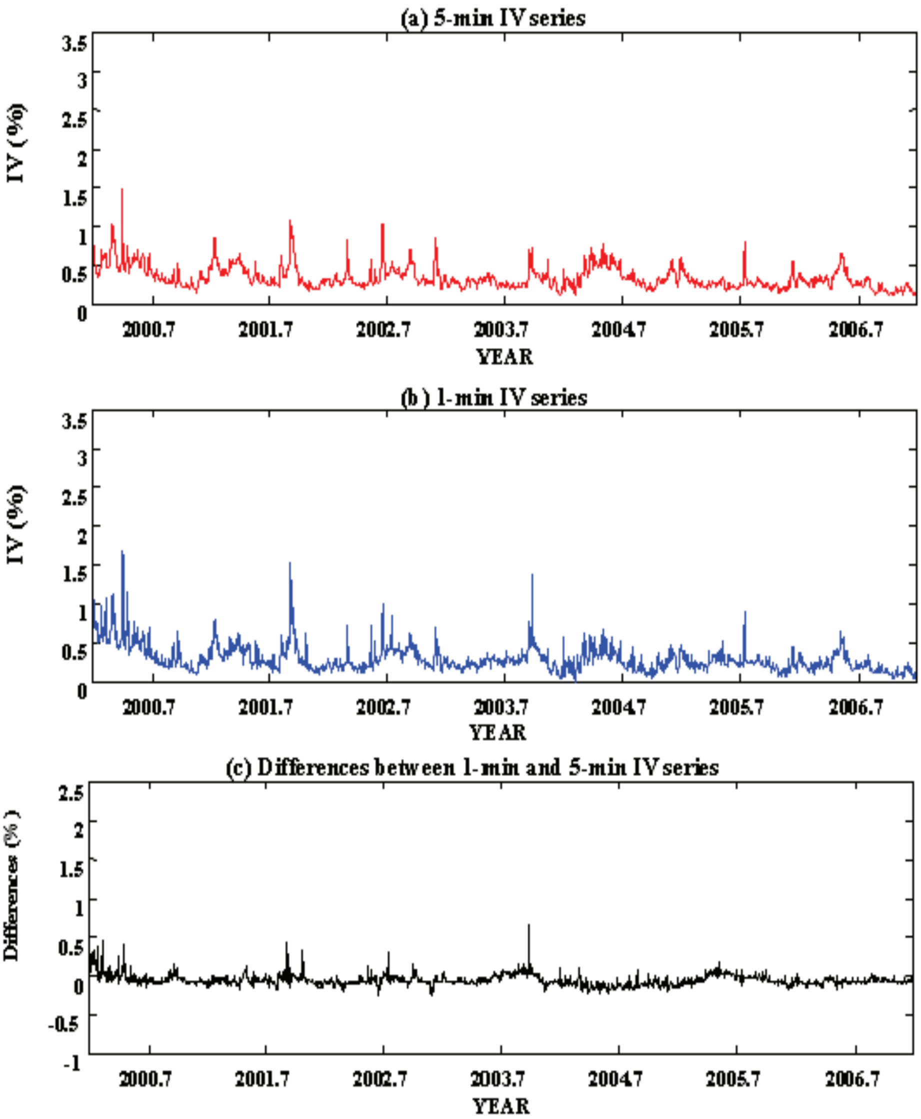
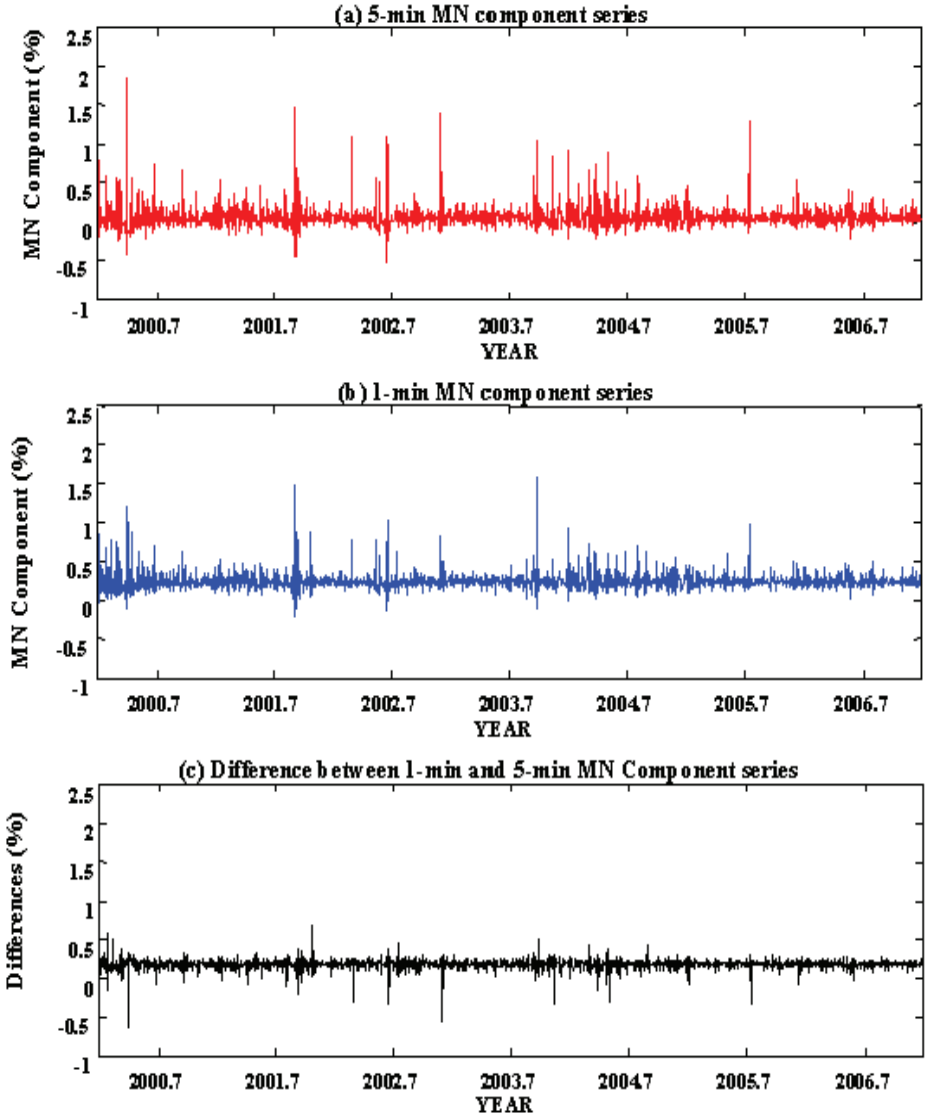


Figure 6: Smoothed series of IV in the two–factor case



Note: Figure 6(c) displays 1–minute IV series minus 5–minute IV series.

Figure 7: Smoothed series of MN component in the two–factor case



Note: Figure 7(c) displays 1–minute MN component series minus 5–minute MN component series.

Figure 8: Smoothed series of discretization error in the two–factor case