HITOTSUBASHI UNIVERSITY

Global COE Hi-Stat Discussion Paper Series 200

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

A State Space Approach to Estimating the Integrated Variance under the Existence of Market Microstructure Noise

Daisuke Nagakura
Toshiaki Watanabe

August 2011

Hi-Stat Discussion Paper

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

A State Space Approach to Estimating the Integrated Variance under the Existence of Market Microstructure Noise *

Daisuke Nagakura† and Toshiaki Watanabe‡

† Faculty of Economics,
Keio University,

E-mail: nagakura@z7.keio.jp,

‡ Institute of Economic Research,
Hitotsubashi University,

E-mail: watanabe@ier.hit-u.ac.jp

Abstract

We call the realized variance (RV), calculated with observed prices contaminated by (market) microstructure noises (MNs), the noise-contaminated RV (NCRV), and refer to the bias 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. 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 identifiable parameters. We illustrate how to estimate these parameters. We apply the proposed method to yen/dollar exchange rate data, where we find that most of the variation in NCRV is of the MN component. The proposed method also serves as a convenient way for estimating a general class of continuous-time stochastic volatility (SV) models under the existence of MN.

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

JEL code: C13; C22; C53

*This is a revised version of Global COE Hi-Stat Discussion Paper Series 115. The authors would like to thank Torben Andersen, Siem Jan Koopman, Rachidi Kotchoni, Peter Reinhard Hansen, David Hendry, Koichi Maekawa, Colin McKenzie, Cathy Ning, Masao Ogaki, Yasuhiro Omori, Kosuke Oya, Yushi Yoshida, and Eric Zivot as well as seminar and conference participants at Keio University, Hiroshima University of Economics, Tezukayama University, University of Washington, Waseda University, International Conference "High Frequency Data Analysis in Financial Markets", International Conference on Econometrics and World Economy, Conference "Recent Development in Finance and Econometrics", Canadian Economic Association 43rd Annual Meeting, and 2009 Far East and South Asia Meeting of Econometric Society for their useful comments. Financial supports 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 are gratefully acknowledged.

1 Introduction

In the finance literature of continuous-time models, the value of the *integrated variance* (IV) has played a crucial role for option pricing, risk management, optimal portfolio construction, etc. Roughly speaking, the IV is an integral of continuously changing instantaneous variances over a specified period, for example, a day. It has been used as a measure of the variability, or risk, of financial asset returns. Various methods have been proposed for estimating the IV. See McAleer and Medeiros (2008) for more details on those methods. Among those methods, we follow the line of the state space method proposed by Barndorff-Nielsen and Shephard (2002). Specifically, we extend their method to the case where the price observations contain measurement errors.

Recently, a new class of estimators for the IV, called the *realized variance* (RV), has received increasing popularity in the field of financial econometrics. It has been developed by Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev, Diebold and Ebens (2001), and Andersen, Bollerslev, Diebold and Labys (2001) among others. The RV is defined as a sum of squared intra-period returns. Under moderate assumptions, the RV converges in probability to the IV, as the sampling frequency tends to be high for a fixed interval such as a day. One of the key assumptions for the consistency of the RV is that there are no measurement errors, or (market) microstructure noises (MNs), in observed log-prices. When this assumption is violated, the RV is no longer a consistent estimator for the IV; it diverges under the existence of MN, as the sampling frequency increases for a fixed interval. The MN emerges because of market microstructure frictions such as discreteness of prices, bid-ask bounce and infrequent trading, etc. See, for example, Campbell, Lo and MacKinlay (1997) and Owens and Steigerwald (2006) for extensive discussions on the cause of MN.

Barndorff-Nielsen and Shephard (2002) consider a situation with no MN and propose a state space method for prediction, filtering, and smoothing the IV. They show that if the true price follows a specific continuous-time SV model, then the IV follows an ARMA process. They 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.[1] Thus, given the state space form parameters, one can apply the Kalman filter and smoother to filter out the discretization error. Barndorff-Nielsen and Shephard (2002) demonstrate that estimates of the IV by the Kalman smoother have much smaller mean squared error than the RV. 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, developed in Andersen (1994), Meddahi and Renault (1996, 2000, 2004). Meddahi (2003) derives explicit relationships between the ARMA model parameters and the SV model parameters.

In this paper, we develop the state space method by Barndorff-Nielsen and Shephard (2002), applying the results of Meddahi (2003), for dealing with the problem of MN. Throughout the paper, we assume that the true price follows the SR-SARV model and an observed log-price is the sum of the true log-price and an i.i.d. MN. We call the RV calculated with observed log-prices (that are assumed to be contaminated by MNs) the *noise-contaminated RV* (NCRV), referring to the bias component in the NCRV associated with the MN as the *MN component* (formal definitions of the NCRV and MN component are given in Section 2). We show that the MN component follows a MA(1) process. The main idea of our state space method is to 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), one can show that the sum of these three components, namely, the NCRV, follows an ARMA process. Our state space method can estimate the IV and MN component simultaneously. Note that we estimate not MN but MN component, although we can estimate the variances of MN and its square. One advantage of our method, compared to other existing methods, is that it can filter out not only the MN components but also the discretization errors. It is also worth emphasizing that our state space method can serve as a convenient way for estimating a general class of continuous-time SV models under the existence of MN.

We apply the proposed method to yen/dollar exchange rate data, where we find that most of the variation in NCRV is of the MN component. We also compare forecasting performances of our and the Barndorff-Nielsen, Shephard, and Meddahi (hereafter BSM) state space methods.[2] We find that our method provides much better forecasts than the BSM method when the sampling frequency is relatively high (1 or 5 minutes). When the sampling frequency is relatively low (more than 10 minutes), our method performs still better than the BSM method though the differences are not much pronounced. This is because when the sampling frequency is enough low, the MN effect is negligibly small and thus, the two methods work almost equally well.

The rest of the paper is organized as follows. In the next section, we introduce the class of SV models employed in this paper, namely, the SR-SARV model, and define formally the RV, IV, MN, and MN component. In Section 3, we explain our state space method. In Section 4, we apply our method to the yen/dollar spot exchange rate and compare forecasting performances of our and the BSM state space methods. The last section provides a summary and concluding remarks. Appendix A provides the proofs for Lemmas and Proposition in the text.

2 SR-SARV Model, IV, RV, and MN Component

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:

Assumption 1 (true price process)
The logarithm of spot price, $p(t)$, follows the SR-SARV model, 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 $W(t)$ is a standard Brownian motion and $f(t)$ is a state-variable process (possibly bivariate) independent of $W(t)$.[3] The functions $P_1(\cdot)$ and $P_2(\cdot)$ are defined so that:

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

where λ_1 and λ_2 are positive real numbers.

Assumption 1 implies that $E[\sigma^2(s)] = \sigma^2$ 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 state our results with κ_1 and κ_2 instead of λ_1 and λ_2 for notational convenience. 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, 2001b) 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 at time t 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 objective of research. For example, if the researcher is interested in changes in variances of daily (weekly) returns, t is interpreted as a day (week).

Under Assumption 1, we can consistently estimate the IV by the estimator known as the RV, denoted by $RV_t^{(m)}$, which is defined as:

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

where $r_t^{(m)} \equiv p(t) - p(t - \frac{1}{m})$, and m is a positive integer. The sampling frequency increases as m increases. Here, and hereafter, the superscript "(m)" implies that its value depends on m. If t denotes a day and we take observations every five minutes, then $m = 288$, 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).

2.3 Microstructure noise (MN) component

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

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

We assume the following properties for $\varepsilon(t)$:

Assumption 2 (properties of MN)

(a) $\varepsilon(t)$ is i.i.d. with $E[\varepsilon(t)] = 0$, $\text{var}[\varepsilon(t)] = \sigma_\varepsilon^2$, and $\text{var}[\varepsilon^2(t)] = \omega_\varepsilon^2 < \infty$ for all t.[4]

(b) $\varepsilon(t)$ is independent of $W(s)$ and $f(s)$ (hence $p(s)$ too) for any s and t.

Then an observed return $r_t^{*(m)}$ is expressed as:

$$r_t^{*(m)} \equiv p^*(t) - p^*\left(t - \tfrac{1}{m}\right) = r_t^{(m)} + e_t^{(m)}, \tag{3}$$

where $e_t^{(m)} \equiv \varepsilon(t) - \varepsilon(t - \frac{1}{m})$. Under Assumptions 1 and 2, it is easy to show that $E(r_t^{*(m)}) = E(e_t^{(m)}) = 0$, $\text{var}(r_t^{*(m)}) = \frac{\sigma^2}{m} + 2\sigma_\varepsilon^2$, $\text{var}(e_t^{(m)}) = 2\sigma_\varepsilon^2$, and

$$\text{cov}(r_t^{*(m)}, r_{t-\frac{i}{m}}^{*(m)}) = \text{cov}(e_t^{(m)}, e_{t-\frac{i}{m}}^{(m)}) = \begin{cases} -\sigma_\varepsilon^2 & i = 1, \\ 0 & i \geq 2. \end{cases} \tag{4}$$

Therefore, we have

$$\text{corr}(r_t^{*(m)}, r_{t-\frac{i}{m}}^{*(m)}) = \begin{cases} -\frac{m\sigma_\varepsilon^2}{\sigma^2 + 2m\sigma_\varepsilon^2}, & i = 1, \\ 0 & i \geq 2, \end{cases} \quad \text{and} \quad \text{corr}(e_t^{(m)}, e_{t-\frac{i}{m}}^{(m)}) = \begin{cases} -0.5, & i = 1, \\ 0 & i \geq 2. \end{cases} \tag{5}$$

Note that the result in (4) implies that the noise-contaminated observed return, $r_t^{*(m)}$, follows a zero mean MA(1) process. Note also that the first order autocorrelation of $r_t^{*(m)}$ decreases and converges to -0.5 as $m \to \infty$.

We define the NCRV, denoted by $RV_t^{*(m)}$, as $RV_t^{*(m)} \equiv \sum_{i=1}^m r_{t-1+\frac{i}{m}}^{*(m)2}$. Here, we formally define the "MN component".

Definition 1 (MN component)
The NCRV has the following representation:

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

where

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

We call $u_t^{(m)}$ an *MN component*.

Note that, unlike $RV_t^{*(m)}$, $u_t^{(m)}$ is not necessarily positive because the first term of $u_t^{(m)}$ can be negative and larger in absolute value than the second term. The following lemma is proved in Appendix:

Lemma 1 *Under Assumptions 1 and 2, the mean and autocovariances of MN component, $u_t^{(m)}$, are given as:*

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

Thus, $u_t^{(m)}$ has the autocovariance structure of a MA(1) process, and we can express the MA(1) process as:

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

where W.N.(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:

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

In the next section, we utilize these two different expressions of the moments of $u_t^{(m)}$ (i.e., (7) and (9)) to derive the implicit relationships among the SV and MA(1) process parameters.

3 State Space Approach

In this section, we explain our state space method. We state the results only on the case of $\omega_2 = 0$, i.e., the one-factor SR-SARV model, and omit the results for the two-factor SR-SARV model for the sake of brevity. See Appendix B in Nagakura and Watanabe (2011) on the results for the two-factor SR-SARV model.

3.1 State space from of NCRV

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:

$$IV_t = c_{IV} + \kappa_1 IV_{t-1} + \eta_t + \theta_1 \eta_{t-1}, \quad \eta_t \sim \text{W.N.}(\sigma_\eta^2), \tag{10}$$

where κ_1 is defined as stated below (2). Other ARMA(1, 1) model parameters c_{IV}, θ_1 and σ_η^2 are expressed as functions of the one-factor SR-SARV model parameters σ^2, ω_1^2 and κ_1. See Proposition 3.1 in Meddahi (2003) for more details on those functions.

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

$$RV_t^{*(m)} = IV_t + d_t^{(m)} + u_t^{(m)}. \tag{11}$$

The following lemma is proved in Appendix:

Lemma 2 *Under Assumptions 1 and 2,*

$$\text{cov}(d_t^{(m)},\, \eta_t) = \text{cov}(d_t^{(m)},\, \xi_t^{(m)}) = \text{cov}(\eta_t,\, \xi_t^{(m)}) = 0, \tag{12}$$

where $d_t^{(m)} = RV_t^{(m)} - IV_t$, and $\xi_t^{(m)}$ and η_t are given as in (8) and (10), respectively.

Let η_t and $\xi_t^{(m)}$ be denoted by state variables α_t and $\beta_t^{(m)}$, respectively. From (8), (10), (11), and (12), we have the following state space form of $RV_t^{*(m)}$:

Observation equation

$$RV_t^{*(m)} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t^{(m)} \end{bmatrix} + d_t^{(m)}, \tag{13a}$$

State equation

$$\begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t^{(m)} \end{bmatrix} = \begin{bmatrix} c_{IV} \\ c_u^{(m)} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \kappa_1 & 0 & \theta_1 & 0 \\ 0 & 0 & 0 & \theta_u^{(m)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_{t-1} \\ u_{t-1}^{(m)} \\ \alpha_{t-1} \\ \beta_{t-1}^{(m)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t^{(m)} \end{bmatrix}, \tag{13b}$$

where

$$\begin{bmatrix} d_t^{(m)} \\ \eta_t \\ \xi_t^{(m)} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \sigma_d^{2(m)} & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\xi^{2(m)} \end{bmatrix} \right). \tag{13c}$$

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.[5] 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, the quasi-maximum likelihood (QML) estimation with Gaussian error assumption. We show, however, that this approach is not applicable for the state space form given in (13a) – (13c).

In general, parameters of a state space form are not necessarily identified. More precisely, there are cases that those parameters are not identified from the autocovariance structure of the dependent variable in the sense that there are infinitely many sets of parameter values that give the same autocovariances. See for example, Hamilton (1994, p.388) and Harvey (1989, p.205) for more details. 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.[6]

3.2 Identification failure of state space form 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:

$$(1-\kappa_1 L)RV_t^{*(m)} = c_{RV}^{(m)} + (1+\delta_1^{(m)}L + \delta_2^{(m)}L^2)\tau_t^{(m)}, \quad \tau_t^{(m)} \sim \text{W.N.}(\sigma_\tau^{2(m)}). \tag{14}$$

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 (11), we have

$$\begin{aligned} (1-\kappa_1 L)RV_t^{*(m)} &= (1-\kappa_1 L)IV_t + (1-\kappa_1 L)d_t^{(m)} + (1-\kappa_1 L)u_t^{(m)} \\ &= c_{IV} + \eta_t + \theta_1\eta_{t-1} + d_t^{(m)} - \kappa_1 d_{t-1}^{(m)} + \xi_t^{(m)} \\ &\quad +(1-\kappa_1)c_u^{(m)} + (\theta_u^{(m)} - \kappa_1)\xi_{t-1}^{(m)} - \kappa_1\theta_u^{(m)}\xi_{t-2}^{(m)}. \end{aligned} \tag{15}$$

The two expressions on the right-hand sides in (14) and (15) are of the same process and hence their means and autocovariances must be identical. The autocovariances of the MA process in (14) are given as

$$\gamma_0^{(m)} = (1+\delta_1^{(m)2} + \delta_2^{(m)2})\sigma_\tau^{2(m)}, \quad \gamma_1^{(m)} = (\delta_1^{(m)} + \delta_1^{(m)}\delta_2^{(m)})\sigma_\tau^{2(m)}, \quad \gamma_2^{(m)} = \delta_2^{(m)}\sigma_\tau^{2(m)}, \tag{16}$$

and $\gamma_j = 0$ for $j \geq 3$. It is easy to show that the autocovariances of the MA process in (15) are

$$\gamma_0^{(m)} = (1+\theta_1^2)\sigma_\eta^2 + (1+\kappa_1^2)\sigma_d^{2(m)} + (1+\theta_u^{(m)2} - 2\theta_u^{(m)}\kappa_1 + \kappa_1^2 + \kappa_1^2\theta_u^{(m)2})\sigma_\xi^{2(m)}, \tag{17a}$$

$$\gamma_1^{(m)} = \theta_1\sigma_\eta^2 - \kappa_1\sigma_d^{2(m)} + (\theta_u^{(m)} - \kappa_1 - \kappa_1\theta_u^{(m)2} + \kappa_1^2\theta_u^{(m)})\sigma_\xi^{2(m)}, \tag{17b}$$

$$\gamma_2^{(m)} = -\kappa_1\theta_u^{(m)}\sigma_\xi^{2(m)}, \tag{17c}$$

and $\gamma_j = 0$ for $j \geq 3$. By equating the means of the MA processes in (14) and (15), we have

$$c_{RV}^{(m)} = c_{IV} + (1-\kappa_1)c_u^{(m)}. \tag{17d}$$

Given the ARMA(1, 2) model parameters, $c_{RV}^{(m)}$, κ_1, $\delta_1^{(m)}$, $\delta_2^{(m)}$ and $\sigma_\tau^{2(m)}$, we can calculate $\gamma_j^{(m)}$, $j = 0, 1, 2$. Then, unknown parameters in the equations (17a)$\sim$(17d) 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 only four equations for seven unknown parameters. 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.

3.3 Identifiable parameters

From (7) and (9), we obtain the following equations:

$$c_u^{(m)} = 2m\sigma_\varepsilon^2, \tag{18a}$$

$$(1+\theta_u^{(m)2})\sigma_\xi^{2(m)} = 8\sigma^2\sigma_\varepsilon^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4, \tag{18b}$$

$$\theta_u^{(m)}\sigma_\xi^{2(m)} = \omega_\varepsilon^2. \tag{18c}$$

Assuming that the MA parameter satisfies the invertibility condition, i.e., $|\theta_u^{(m)}| < 1$, we can solve the equations (18a) $\sim$ (18c) for $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$ as:

$$c_u^{(m)} = 2m\sigma_\varepsilon^2, \quad \sigma_\xi^{2(m)} = \frac{\omega_\varepsilon^2}{\theta_u^{(m)}} \quad \text{and} \quad \theta_u^{(m)} = A - \sqrt{A^2-1}, \tag{19}$$

where $A = 4\frac{\sigma^2\sigma_\varepsilon^2}{\omega_\varepsilon^2} + 2m - 1 + 2m\frac{\sigma_\varepsilon^4}{\omega_\varepsilon^2}$. Note that $0 < \theta_u^{(m)} < 1$ because $A > 1$.

From Proposition 3.1 in Meddahi (2003), and (19), 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. To make the functional relationship explicit, we may denote them as:

$$c_{IV}(\kappa_1,\sigma^2),\ \ \theta_1(\kappa_1),\ \ \sigma_\eta^2(\kappa_1,\omega_1^2),\ \ c_u^{(m)}(\sigma_\varepsilon^2),\ \ \theta_u^{(m)}(\sigma^2,\sigma_\varepsilon^2,\omega_\varepsilon^2),\ \ \sigma_d^{2(m)}(\kappa_1,\sigma^2,\omega_1^2),\ \text{and}\ \sigma_\xi^{2(m)}(\sigma^2,\sigma_\varepsilon^2,\omega_\varepsilon^2). \tag{20}$$

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 (20) into Equations (17a)$\sim$(17d), 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.

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, $\gamma_j^{(m)}$, $j = 0,...,2$. The following proposition assures the uniqueness of the identification:

Proposition 1 *Given $c_{RV}^{(m)}$, κ_1, $\gamma_j^{(m)}$, $j = 0,...,2$ and (20), under the condition $\sigma_\varepsilon^2 > 0$, Equations (17a)$\sim$(17d) are uniquely solved for σ^2, ω_1^2, σ_ε^2 and ω_ε^2 as:*

$$\omega_\varepsilon^2 = -\frac{\gamma_2^{(m)}}{\kappa_1}, \quad \omega_1^2 = \frac{(\log\kappa_1)^2[\kappa_1\gamma_0^{(m)} + (1+\kappa_1^2)\gamma_1^{(m)} + \frac{1+\kappa_1^4}{\kappa_1}\gamma_2^{(m)}]}{(1-\kappa_1)^3(1+\kappa_1)}, \tag{21a}$$

$$\sigma_\varepsilon^2 = \sqrt{\frac{c_{RV}^{(m)2}}{2m^2(1-\kappa_1)^2} - \frac{(2m-1)\gamma_2^{(m)}}{2m\kappa_1} - \frac{\gamma_0^{(m)} - 2D\omega_1^2 - 2\gamma_2^{(m)}}{4m(1+\kappa_1^2)}}, \tag{21b}$$

and

$$\sigma^2 = \frac{c_{RV}^{(m)}}{1-\kappa_1} - 2m\sigma_\varepsilon^2, \tag{21c}$$

where

$$D = B + m(1+\kappa_1^2)C, \quad B \equiv \frac{\kappa_1^2 - 1 - (1+\kappa_1^2)\log\kappa_1}{(\log\kappa_1)^2} \quad \textit{and} \quad C \equiv \frac{2(\kappa_1^{\frac{1}{m}} - 1 - \log\kappa_1^{\frac{1}{m}})}{(\log\kappa_1)^2}. \tag{21d}$$

Proposition 1 implies 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 (14). 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.

3.4 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.

We showed in (21) 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 (14) by QML estimation assuming Gaussian errors.[7]

Step 3 Given the estimates of $c_{RV}^{(m)}$, κ_1, $\delta_1^{(m)}$, $\delta_2^{(m)}$ and $\sigma_\tau^{2(m)}$ obtained in Step 2, calculate the first three autocovariances of the MA process, namely, $\gamma_j^{(m)}$, $j = 0, ..., 2$ as in (16).

Step 4 Given the estimates of $c_{RV}^{(m)}$, κ_1 and $\gamma_j^{(m)}$, $j = 0, ..., 2$ obtained in Steps 2 and 3, calculate ω_ε^2, σ_ε^2, ω_1^2 and σ^2 applying the results in (21a) – (21d).

This approach is simple and easy to implement, however, does not guarantee that the resulting parameter estimates are positive because of the intrinsic uncertainty of the ARMA model estimation. For example, if the estimate of $\gamma_2^{(m)}$ is positive, 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, one 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 Eq(3.7), Proposition 3.1 in Meddahi (2003) and (19).

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 (13a)–(13c) 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 (and κ_1). One can obtain estimates for the state space form parameters by substituting the estimates by either of the above two approaches into the functions in (20).

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.

4 Empirical Analysis

4.1 Data description

The yen/dollar spot exchange rate series we analyze 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.

[Figure 1 around here]

We apply the previous tick method, i.e., we use the most recent observed price, where price data are missing. There are trading days that display too many missing values or low trading activity. Following Andersen, Bollerslev, Diebold and Labys (2001), we remove the data of inactive trading days. Whenever we do so, we always remove the price data from 21:00 GMT on one night to 20:59 the next evening because we define one trading day as the 24 hours from 21:00 GMT on one night to 21:00 GMT 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):

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

(2) the days where there are more than 1000 minutes of zero returns

(3) the days where the price does not change for more than *consecutive* 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 returns with various m's, namely, one-minute ($m = 1440$), five-minute ($m = 288$), ten-minute ($m = 144$), fifteen-minute ($m = 96$), and thirty-minute ($m = 48$) returns.[8] Table 1 reports the sample means, sample variances, sample standard deviations, and sample autocorrelations of these returns. The autocorrelations beyond the first lag are close to zero, which indicates that these return data would be approximated as MA(1) processes (except for the cases of $m = 288$ and $m = 144$ in that the second order autocorrelations remain relatively high).

[Table 1 around here]

With these returns, we calculate five series of daily NCRV, namely, one-minute ($m = 1440$) five-minute ($m = 288$), ten-minute ($m = 144$), fifteen-minute ($m = 96$), and thirty-minute ($m = 48$) NCRV series. Table 2 reports the sample means, sample variances, sample standard deviations, and sample autocorrelations of these NCRV series. The sample means of these NCRV series increase as the sampling frequency tends to be high, or $m \to \infty$. This is consistent with the existence of MN (see (17d) and (18a)). The autocorrelations of these NCRV series are somewhat lower than usually expected for changing variances of financial time series. This may be because of the existence of MN. In fact, in the next subsection, we show that estimates of the autocorrelations of IV are significantly higher than these values.

[Table 2 around here]

4.2 Estimation of parameters, IV and MN component

For these NCRV series, we estimate the parameters of the one- and two-factor SR-SARV models (hereafter, simply one- and two- factor models) in (1) and (2), by the method described in Section 4.3 (and in Appendix B in Nagakura and Watanabe (2011) for the two-factor model). Note that, in general, the values of these NCRV series are different although they all are estimates of the same IV series. Consequently, the estimates of the SV model parameters for different NCRV series are different. We estimate the parameters by the direct approach. Table 3 displays the estimation results for these NCRV series. For both the one- and two-factor models, the estimates of parameters are very similar across these NCRV series, expect for $\widehat{\omega}_\varepsilon^2$, or estimates of the variance of the square of MN. Interestingly, it seems that $\widehat{\omega}_\varepsilon^2$ increases inversely proportional to m, while $\widehat{\sigma}_\varepsilon^2$ are stable across different sampling frequencies. This result implies that the fourth moment increases in proportion to the sampling interval. As we will discuss in Section 5.2, this may be due to the existence of jumps in returns. The estimates of the persistence parameters for the two-factor model (i.e., $\widehat{\kappa}_1$ and $\widehat{\kappa}_2$) imply that there are two factors with significantly different levels of persistence. One is very persistent, and the other is moderately persistent. For the one-factor model, the persistence of these two factors must be captured by only one parameter, κ_1. As a result, the estimates of κ_1 for the one-factor model are somewhat lower than those for the two-factor model.

[Tables 3 and 4 around here]

The estimates of the state space form parameters are computed from the estimates of the SV model parameters. The results are shown in Table 4. Again, the estimates of the parameters that do not depend on m are very similar across different m's, for both the one- and two-factor models. Also, for a fixed m, the estimates of the MN component parameters, that depend on m, for the one-factor model are very similar to those for the two-factor model. This implies that the number of factors employed does not affect the estimates of MN component parameters very much. The bias of the NCRV series is equal to the value of $c_u^{(m)}$. Its estimate, $\widehat{c}_u^{(m)}$, decreases as m decreases, and is almost negligible for $m \leq 288$, but then the estimate of the variance of the discretization error, $\widehat{\sigma}_d^{2(m)}$, increases, as expected from the bias-variance trade-off of NCRV that the theory implies. Our method is designed to remove the MN or bias components. Selecting an optimal number of m for doing this task most efficiently is beyond the scope of this paper. One possible approach for choosing a "better" (but not necessarily optimal) m is to select the m that gives the best forecasting performance among the candidate m's. See Section 4.3 for more details.

[Table 5 around here]

Table 5 reports the estimates of some important values including autocorrelations of the IV, unconditional variances of the IV and MN component, and their ratios to the unconditional variance of the NCRV. For both the one- and two-factor models, the estimates of the autocorrelations are significantly higher than the sample autocorrelations of the NCRV. This result suggests that apparent low autocorrelations of the NCRV do not reflect the correlations of the IV but is due to the existence of MN. The estimated ratio of the unconditional variance of the MN component to the unconditional variance of the NCRV implies that about half of the aggregate fluctuations of the NCRV series is due to the MN component. This result can be confirmed visually in Figures 2 and 3, which plot the estimates of the IV and MN component series by the Kalman smoothing (hereafter, we call them smoothed IV and MN component series, respectively), along with the corresponding NCRV series, for different m's and for the one- and two- factor models. Note that these estimates are of the same underlying IV series whose values do not depend on the value of m (on the other hand, the underlying MN component series differ for different m's). These smoothed IV series are very similar across different m's. From these figures, we can see that there are "spurious increases in the NCRV", namely, the NCRV occasionally takes a large value, however, it is mostly due not to the IV but to the MN component. In fact, the smoothed IV rarely takes the values more than 1.5. The smoothed IV series of the one-factor model seem smoother than those of the two-factor model. This is expected from the result that the (estimated) autocorrelations of the IV series of the two-factor model are lower, which implies that they are relatively closer to white noise compared with the smoothed IV series of the one-factor model.

[Figures 2 and 3]

4.3 Comparing forecasting performances

In this subsection, we compare forecasting performances of our and the BSM methods. We consider only one-day-ahead forecasting. Let $IV_{t+1|t}^{(NR)(m)}$ and $IV_{t+1|t}^{(BSM)(m)}$ denote one-day-ahead IV predictions by our (NR stands for "noise robust") and the BSM methods, respectively. Because the IV is not directly observed, we have to use a proxy for the IV in evaluating forecasting performances. Following Andersen, Bollerslev, Diebold, and Labys (2003), we use $RV_t^{*(48)}$, namely, 30-minute NCRV, as a proxy for the IV, which is supposed to be much less subject to the bias attributed to the MN, albeit a very noisy proxy.

The unknown model parameters are estimated by QMLE with $RV_t^{*(m)}$. Following Andersen, Bollerslev, Diebold, and Labys (2003), we run the so called Mincer-Zarnowitz style regressions. We also report estimates of expected values of the usual mean squared error (MSE) and QLIKE loss functions in Patton (2011):

$$\text{MSE} : T^{-1}\sum_{t=1}^{T-1}(RV_{t+1}^{*(48)} - IV_{t+1|t}^{X(m)})^2, \quad \text{and} \quad \text{QLIKE} : T^{-1}\sum_{t=1}^{T-1}(\log IV_{t+1|t}^{X(m)} + \frac{RV_{t+1}^{*(48)}}{IV_{t+1|t}^{X(m)}}),$$

The forecasting performances are evaluated by the values of R^2, MSE, and QLIKE.[9]

We run the following three regressions for both the one- and two-factor models:

$$\begin{aligned}
&\text{(I)} \quad RV_{t+1}^{*(48)} = a_0 + a_1 IV_{t+1|t}^{(NR)(m)} + \nu_t^{(a)},\\
&\text{(II)} \quad RV_{t+1}^{*(48)} = b_0 + b_1 IV_{t+1|t}^{(BSM)(m)} + \nu_t^{(b)},\\
&\text{(III)} \quad RV_{t+1}^{*(48)} = c_0 + c_1 IV_{t+1|t}^{(NR)(m)} + c_2 IV_{t+1|t}^{(BSM)(m)} + \nu_t^{(c)},
\end{aligned}$$

for $m = 96, 144$, 288, and 1440, which correspond to 15-, 10-, 5-, and 1-minute.

We conduct both in- and out-of-sample forecasting. To obtain the in-sample forecasts, we estimate unknown parameters using all available samples of $RV_s^{*(m)}$ $(s = 1, ..., 1809)$, and calculate $IV_{t+1|t}^{(X)(m)}$, $X = \{NR, BSM\}$, $t = 1, ..., 1808$, with the estimated parameter values and all the past samples up to time t. Then, we run the regressions (I) $\sim$ (III) for $t = 1, ..., 1808$, and calculate the associated R^2, MSE, and QLIKE values. To obtain the out-of-sample forecasts, starting with $t = 1200$, we estimate unknown parameters, using only the most recent 1200 samples before $t+1$ (i.e., $t, t-1, ..., t-1199$),[10] calculate $IV_{t+1|t}^{(X)(m)}$, $X = \{NR, BSM\}$, with the estimated parameter values and the most recent 1200 samples, and repeat this for $t = 1201, ...1808$. Then, again we run the three regressions for $t = 1200, ..., 1808$, and calculate the R^2, MSE, and QLIKE values.

[Table 6 around here]

Table 6 shows the results for the in-sample forecasting. First, we compare the results for the one- and two-factor models. For both our (the regression (I)) and the BSM (the regression (II)) methods and for any m, the two-factor model performs better than the one-factor model in terms of R^2 values. Specifically, for the regression (I), the R^2 in the two-factor case is greater than the R^2 in the one-factor case by 0.0082 at the largest when $m = 288$, and by 0.0056 at the smallest when $m = 1440$. The improvements in R^2 are more pronounced for the BSM method. The largest and smallest improvements in R^2 for the regression (II) are by 0.041 when $m = 288$, and 0.011 when $m = 144$, respectively. For the regression (III), the R^2 values are close to those of the regression (I). In terms of the MSE and QLIKE criterion, there is no significant difference in the forecasting performances between the one- and two- factor models for both our and the BSM methods.

Next, we focus on the comparison between our and the BSM methods. We first compare the results for the one-factor model and then for the two-factor model. For the one-factor model, the results in Table 6 shows that our method is superior to the BSM method in terms of the both criteria. The R^2 values in the regression (I) are better (larger) than those in the regression (II) for any m. In terms of MSE and QLIKE criterion, again, our method has smaller MSE and QLIKE values than the BSM method for any m. This can be confirmed visually in Figures 4 (a) and (b), that plot the predicted IV series by our and the BSM methods along with the 30-minute NCRV for $m = 1440$ and 288, respectively. These figures show that the predictions by the BSM method have large upward biases due to ignoring the MN effects, in particular for $m = 1440$. The results for the regression (III) also imply that our method works better than the BSM method. Adding the forecasts by the BSM method does not significantly improve the R^2 values compared with the regression (I), and the coefficients of the forecasts by the BSM method are not significantly different from zero for any m.

Our method seems to work equally well for any m, while the BSM method performs worse as m increases. This is because our method takes into account the MN effects, whereas the BSM method does not, and consequently, the forecasts by the BSM method are deteriorated by the MN effects, as m increases. Even when m is relatively small ($m = 144$ or 96), our method still works better than the BSM method, however, the differences in R^2, MSE, and QLIKE values between the two methods are small. In fact, the two methods provide very similar forecasts. This can also be checked in Figures 4 (c) and (d), that plot the predicted IV series by the two methods for $m = 144$ and 96, respectively, where the two predicted IV series overlap and hard to distinguish visually. This result is natural because the MN effects vanish as m gets small so that the BSM method works as well. Our empirical results confirm that the MN effects are almost negligible when $m \leq 144$, or the sampling frequency is more than 10 minutes, for the exchange rate data.

[Figures 4 and 5 around here]

We next turn to the results for the two-factor model. Interestingly, unlike the case of the one-factor model, even when $m = 1440$, the predicted IV series by the BSM method capture the "dynamics" of the 30-minute NCRV well, as seen in Figure 5 (a), though their "levels" are severely biased upwardly. This is the reason why the R^2 values in the regression (II) (the BSM method) are as good as those in the regression (I) (our method) even when $m = 1440$, whereas the MSE and QLIKE values of our method are significantly better than those of the BSM method. As m gets small, again the differences in the forecasting performances by the two methods become small for the same reason as in the case of the one-factor model.

[Table 7 around here]

Table 7 reports the results for the out-of-sample forecasting. Figures 6 and 7 plots the predicated IV series by our and the BSM methods for the one- and two-factor models, respectively, for the out-of-sample forecasting. The results are qualitatively very similar to those for the in-sample forecasting, and hence the same comments apply.

[Figures 6 and 7 around here]

5 Discussions on Some Issues

5.1 Range of autocorrelations of MA part of NCRV

In Section 3.2, we expressed autocovariances of MA part of NCRV in terms of state space form parameters. Because state space form parameters are functions of the underlying SV model parameters, it is possible to express those autocovariances in terms of the SV model parameters and see how the changes in values of the SV model parameters affect the values of the autocorrelations.

Let χ denote the kurtosis of $\varepsilon(t)$, i.e., $\chi = E[\varepsilon(t)^4]/\sigma_\varepsilon^4$. Using the results in Appendix, we can show that

$$\gamma_0^{(m)} = 2D\omega_1^2 + \frac{2(1+\kappa_1^2)\sigma^4}{m} + 8(1+\kappa_1^2)\sigma^2\sigma_\varepsilon^2 + 2(2m-1)(1+\kappa_1^2)\sigma_\varepsilon^4\chi + 2(1+\kappa_1^2)\sigma_\varepsilon^4 - 2\kappa_1\sigma_\varepsilon^4(\chi-1)n,$$

$$\gamma_1^{(m)} = 2E\omega_1^2 - \frac{2\kappa_1\sigma^4}{m} - 8\kappa_1\sigma^2\sigma_\varepsilon^2 - 2(2m-1)\kappa_1\sigma_\varepsilon^4\chi - 2\kappa_1\sigma_\varepsilon^4 + (1+\kappa_1^2)\sigma_\varepsilon^4(\chi-1),$$

and

$$\gamma_2^{(m)} = -\kappa_1\sigma_\varepsilon^2(\chi-1),$$

where D and E, both of which are functions of only κ_1 and m, are given as in (21d) and Appendix. The first and second order autocorrelations, $\rho_1^{(m)}$ and $\rho_2^{(m)}$, are defined as $\rho_i^{(m)} = \gamma_i^{(m)}/\gamma_0^{(m)}$, $i = 1, 2$. Interestingly, $\rho_1^{(m)}$ converges to $-\kappa_1/(1+\kappa_1^2)$ as $m \to \infty$ when other values are fixed, regardless of whether or not $\sigma_\varepsilon^2 = 0$. Note that, for an arbitrary MA(2) process, the minimum and maximum attainable values for $\rho_1^{(m)}$ are -0.5 and -0.5, respectively, whereas $-0.5 < -\kappa_1/(1+\kappa_1^2) < 0$ for any value in the parameter space of κ_1. On the other hand, when $\sigma_\varepsilon^2 = 0$, $\rho_2^{(m)} = 0$ regardless of m, whereas even when $\sigma_\varepsilon^2 \neq 0$, $\rho_2^{(m)}$ converges to 0.

Figure 8 plots the values of $\rho_1^{(m)}$ when one of σ^2, ω_1^2, σ_ε^2, χ, and m moves with other parameter values being fixed. We observe that as σ^2, σ_ε^2, and χ increase, $\rho_1^{(m)}$ decreases, whereas as ω_1^2 increases $\rho_1^{(m)}$ decreases. Interestingly, as m increases, initially $\rho_1^{(m)}$ increases and can even be positive values, then at some point, $\rho_1^{(m)}$ starts decreasing and converge to $-\kappa/(1+\kappa^2)$.

5.2 Effects of jumps in returns

Empirically, it is often observed that the financial asset returns exhibit abrupt changes or jumps. In this section, we briefly discuss the effects of jumps in returns to our state space approach. In particular, we consider the case in that the observed prices $p^*(t)$ is the sum of three components: $p^*(t) = p(t)+\varepsilon(t)+\zeta(t)$, where $\zeta(t)$ is a jump process that is assumed to satisfy: (a) $z_t^{(m)} \equiv \zeta(t) - \zeta(t - \frac{1}{m})$ are i.i.d. random variables; (b) $z_s^{(m)}$ is independent of $p(t)$ and $\varepsilon(t)$ for all t and s. For example, these assumptions are satisfied if $\zeta(t)$ follows a compound Poisson process so that $\zeta(t) = \sum_{j=1}^{N(t)} \tau_j$, where $N(t)$ is a Poisson process with intensity $\lambda > 0$, and jump sizes τ_j's are i.i.d. random variables independent of $N(t)$, $p(t)$ and $\varepsilon(t)$ for all j and t. In this case, the random variable $z_t^{(m)}$ is i.i.d. and its mean and variance are given as $E[z(t)] = \lambda\mu_1/m$ and $\text{var}[z(t)] = \lambda\mu_2/m$, where $\mu_r = E[\tau_j^r]$. In this case, the observed return $r_t^{*(m)}$ is given as $r_t^{*(m)} = r_t^{(m)} + e_t^{*(m)}$, where $e_t^{*(m)} = e_t^{(m)} + z_t^{(m)}$. For $e_t^{(m)}$, we show in Appendix that

$$E[e_t^{(m)}] = 0, \quad E[e_t^{(m)}] = \text{var}[e_t^{(m)}] = 2\sigma_\varepsilon^2, \quad \text{and} \quad \text{var}[e_t^{(m)2}] = 2\omega_\varepsilon^2 + 4\sigma_\varepsilon^4,$$

which do not depend on m. On the other hand, for $e_t^{*(m)}$, we can show that

$$E[e_t^{*(m)}] = \frac{\lambda\mu_1}{m}, \quad \text{var}[e_t^{*(m)}] = 2\sigma_\epsilon^2 + \frac{\lambda\mu_2}{m},$$

and

$$\begin{aligned}\mathrm{var}[e_t^{*(m)2}] &= 2\omega_\varepsilon^2 + 4\sigma_\varepsilon^4 + 8\sigma_\varepsilon^2\frac{\lambda}{m}\left(\mu_2 + \frac{\lambda\mu_1^2}{m}\right) + \frac{\lambda}{m}\left(\mu_4 + \frac{4\lambda\mu_3\mu_1}{m} + \frac{2\lambda\mu_2^2}{m} + \frac{4\lambda^2\mu_2\mu_1^2}{m^2}\right) \\ &= 2\omega_\varepsilon^2 + 4\sigma_\varepsilon^4 + \frac{\lambda}{m}(8\sigma_\varepsilon^2\mu_2 + \mu_4) + o(\tfrac{1}{m})\end{aligned}$$

Here, unlike $e_t^{(m)}$, they depend on the value of m. Thus, estimation of σ_ε^2 and ω_ε^2 would be affected or different for different m's. In fact, estimates of ω_ε^2 in our empirical analysis are significantly different for different m's, which may suggest that there are jumps in returns. It is known that a two factor model behaves similar to a diffusion-jump model (see Chernov, Gallant, Ghysels and Tauchen , 2003). To some extent, this may explain why MN component series in Figure 3 show a smaller variation than the ones in Figure 2.[11]

6 Summary and Concluding Remarks

In this paper, we extended the state space method proposed by Barndorff-Nielsen and Shephard (2002) to the situation in which there exist MNs. 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 SV model parameters. We applied the proposed method to yen/dollar exchange rate data, where we find that most of the variations in the NCRV are due to the MN component.

There are several issues that we do not consider in the present paper. For example, there is a trade-off between mean and variance of MN components. Our method is designed to remove the MN or bias components, and there would be an optimal m with which our method works most efficiently. We briefly discussed in Section 5.2 that if there exist jumps, our method would treat them as a part of MN, but then the parameters related to MN depends on m and other parameter estimates may also be influenced. We leave these issues as subjects of future research.

Appendix: Proofs

In this Appendix, we sketch the proofs for Lemmas 1, 2, and Proposition 1 in the text. See Nagakura and Watanabe (2011) for more detailed proofs. Hereafter, we suppress the superscript "(m)", and let ε_t denote $\varepsilon(t)$ for notational simplicity.

Proof of Lemma 1

Because $E(e_t^2) = 2\sigma_\varepsilon^2$ and r_t is independent of e_t by Assumption 2, we have $E(u_t) = 2m\sigma_\varepsilon^2$. To derive $\mathrm{var}(u_t)$ and $\mathrm{cov}(u_t, u_{t-1})$, we first calculate $\mathrm{cov}(r_s e_s, r_t e_t)$ and $\mathrm{cov}(e_t^2, e_s^2)$. For $\mathrm{cov}(r_s e_s, r_t e_t)$, when $t = s$, noting that $E(r_t) = E(e_t) = 0$, we have:

$$\mathrm{cov}(r_t e_t, r_t e_t) = E(e_t^2)E(r_t^2) = 2\sigma_\varepsilon^2 E\{[\textstyle\int_{t-\frac{1}{m}}^t \sigma(s)\mathrm{d}W(s)]^2\} = 2\sigma_\varepsilon^2 E[\int_{t-\frac{1}{m}}^t \sigma^2(s)\mathrm{d}s] = \frac{2\sigma_\varepsilon^2\sigma^2}{m}. \tag{22}$$

The third equality comes from the Ito isometry. When $t \neq s$, we have $\mathrm{cov}(r_s e_s, r_t e_t) = 0$. Next, for $\mathrm{cov}(e_t^2, e_s^2)$, after some calculations, we have:

$$\mathrm{cov}(e_t^2, e_s^2) = \begin{cases} 2\omega_\varepsilon^2 + 4\sigma_\varepsilon^4, & \text{for} \quad s = t, \\ \omega_\varepsilon^2 & \text{for} \quad s = t \pm \frac{1}{m}, \\ 0 & \text{for} \quad s = t \pm \frac{i}{m}, \quad i > 1. \end{cases} \tag{23}$$

Furthermore, we have $\mathrm{cov}(r_t e_t, e_s^2) = 0$ for any t and s. From the above results, we have:

$$\mathrm{var}(u_t) = \mathrm{var}\left(2\sum_{i=1}^m r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}} + \sum_{i=1}^m e_{t-1+\frac{i}{m}}^2\right) = 8\sigma_\varepsilon^2\sigma^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4,$$

and

$$\mathrm{cov}(u_t, u_{t+1}) = \mathrm{cov}\left(2\sum_{i=1}^m r_{t-1+\frac{i}{m}} e_{t-1+\frac{i}{m}} + \sum_{i=1}^m e_{t-1+\frac{i}{m}}^2, 2\sum_{i=1}^m r_{t+\frac{i}{m}} e_{t+\frac{i}{m}} + \sum_{i=1}^m e_{t+\frac{i}{m}}^2\right) = \omega_\varepsilon^2.$$

It is easy to check that $\mathrm{cov}(u_t, u_{t\pm i}) = 0$ for $i > 1$, and hence we have (7).

Proof of Lemma 2
From Assumption 2(b), it follows that, for all t and s and for any real numbers $\Delta > 0$ and $\Delta' > 0$,

$$\begin{aligned}
&\operatorname{cov}\left(\int_{t-\Delta}^{t}\sigma^2(x)\mathrm{d}x,\ e_s \times \int_{s-\Delta'}^{s}\sigma(x)\mathrm{d}W(x)\right)\\
&= E(e_s)E\left(\int_{t-\Delta}^{t}\sigma^2(x)^2\mathrm{d}x \times \int_{s-\Delta'}^{s}\sigma(x)\mathrm{d}W(x)\right) - E(e_s)E\left(\int_{t-\Delta}^{t}\sigma^2(x)^2\mathrm{d}x\right)E\left(\int_{s-\Delta'}^{s}\sigma(x)\mathrm{d}W(x)\right) = 0,
\end{aligned}$$

Similarly,

$$\operatorname{cov}\left(\int_{t-\Delta}^{t}\sigma^2(x)\mathrm{d}x,\ e_s^2\right) = 0, \quad \operatorname{cov}(r_t^2, r_s e_s) = 0, \quad \text{and} \quad \operatorname{cov}(r_t^2, e_s^2) = 0.$$

Hence, we have

$$\begin{aligned}
\operatorname{cov}(IV_t, u_s) = &\operatorname{cov}\left[\sum_{i=1}^{m}\int_{t-1+\frac{i-1}{m}}^{t-1+\frac{i}{m}}\sigma^2(x)\mathrm{d}x,\ 2\sum_{i=1}^{m}\left(e_{s-1+\frac{i}{m}} \times \int_{s-1+\frac{i-1}{m}}^{s-1+\frac{i}{m}}\sigma(x)\mathrm{d}W(x)\right) + \sum_{i=1}^{m}e^2_{s-1+\frac{i}{m}}\right]\\
= &\ 2\sum_{i=1}^{m}\sum_{j=1}^{m}\operatorname{cov}\left(\int_{t-1+\frac{i-1}{m}}^{t-1+\frac{i}{m}}\sigma^2(x)\mathrm{d}x,\ e_{s-1+\frac{j}{m}} \times \int_{s-1+\frac{j-1}{m}}^{s-1+\frac{j}{m}}\sigma(x)\mathrm{d}W(x)\right)\\
&+\sum_{i=1}^{m}\sum_{j=1}^{m}\operatorname{cov}\left(\int_{t-1+\frac{i-1}{m}}^{t-1+\frac{i}{m}}\sigma^2(x)\mathrm{d}x,\ e^2_{s-1+\frac{j}{m}}\right)\\
= &\ 0,
\end{aligned} \tag{24}$$

$$\begin{aligned}
\operatorname{cov}(RV_t, u_s) = &\operatorname{cov}\left(\sum_{i=1}^{m}r^2_{t-1+\frac{i}{m}},\ 2\sum_{i=1}^{m}r_{s-1+\frac{i}{m}}e_{s-1+\frac{i}{m}} + \sum_{i=1}^{m}e^2_{s-1+\frac{i}{m}}\right)\\
= &\ 2\sum_{j=1}^{m}\sum_{i=1}^{m}\operatorname{cov}\left(r^2_{t-1+\frac{j}{m}},\ r_{s-1+\frac{i}{m}}e_{s-1+\frac{i}{m}}\right) + \sum_{j=1}^{m}\sum_{i=1}^{m}\operatorname{cov}\left(r^2_{t-1+\frac{j}{m}},\ e^2_{s-1+\frac{i}{m}}\right)\\
= &\ 0,
\end{aligned}$$

and thus,

$$\operatorname{cov}(d_t, u_s) = \operatorname{cov}(RV_t - IV_t, u_s) = \operatorname{cov}(RV_t, u_s) - \operatorname{cov}(IV_t, u_s) = 0, \tag{25}$$

for all t and s. Meddahi (2002) shows that, under no leverage effects, $\operatorname{cov}[IV_t, d_t] = 0$. It is easy to extend his proof to show

$$\operatorname{cov}(IV_t, d_s) = 0, \tag{26}$$

for all t and s. From (24) and (26), and noting that η_t can be expressed as $\eta_t = \psi(L)IV_t$, where $\psi(L)$ is an appropriate lag polynomial, we have

$$\operatorname{cov}(\eta_t, u_s) = 0 \quad \text{and} \quad \operatorname{cov}(\eta_t, d_s) = 0, \tag{27}$$

for all t and s. From (25) and (27), and noting that $\xi_t = (1 - \theta_u L)^{-1}u_t$, we have $\operatorname{cov}(\eta_t, \xi_s) = 0$ and $\operatorname{cov}(d_t, \xi_s) = 0$, for all t and s, which completes the proof of Lemma 2.

Proof of Proposition 1
From (17c) and (18c), we have $\omega_\varepsilon^2 = -\frac{\gamma_2}{\kappa_1}$, which is the first result in (21a). From the results of Meddahi (2003), we have

$$\sigma_\eta^2 = \frac{2B\omega_1^2}{1+\theta_1^2} \quad \text{and} \quad \sigma_d^2 = \frac{2\sigma^4}{m} + 2mC\omega_1^2, \tag{28}$$

where B and C are given as in (21d). From $\omega_\varepsilon^2 = \theta_u\sigma_\xi^2$ in (18c), we have:

$$(1 + \theta_u^2 - 2\theta_u\kappa_1 + \kappa_1^2 + \kappa_1^2\theta_u^2)\sigma_\xi^2 = \left[\left(\frac{1}{\theta_u} + \theta_u\right)(1+\kappa_1^2) - 2\kappa_1\right]\omega_\varepsilon^2, \tag{29}$$

and

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

Substituting (28), (29) and (30) into (17a) and (17b), we have:

$$\gamma_0 = 2D\omega_1^2 + 2\frac{1+\kappa_1^2}{m}\sigma^4 + \left[\left(\frac{1}{\theta_u} + \theta_u\right)(1+\kappa_1^2) - 2\kappa_1\right]\omega_\varepsilon^2, \tag{31a}$$

and

$$\gamma_1 = 2E\omega_1^2 - 2\frac{\kappa_1}{m}\sigma^4 - \left[\left(\frac{1}{\theta_u} + \theta_u\right)\kappa_1 - (1+\kappa_1^2)\right]\omega_\varepsilon^2, \tag{31b}$$

where $E = \rho B - m\kappa_1 C$, $\rho = \theta_1/(1+\theta_1^2)$, and B, C, and D are given as in (21d). From (31), we have:

$$\begin{aligned}\kappa_1\gamma_0 + (1+\kappa_1^2)\gamma_1 &= 2\left[\kappa_1 D + (1+\kappa_1^2)E\right]\omega_1^2 + \left[(1+\kappa_1^2)^2 - 2\kappa_1^2\right]\omega_\varepsilon^2, \\ &= 2\left[\kappa_1 + (1+\kappa_1^2)\rho\right]B\omega_1^2 + (1+\kappa_1^4)\omega_\varepsilon^2, \\ &= \frac{(1-\kappa_1)^3(1+\kappa_1)}{(\log\kappa_1)^2}\omega_1^2 + (1+\kappa_1^4)\omega_\varepsilon^2,\end{aligned} \tag{32}$$

where, to obtain the third equality, we substitute an alternative expression of ρ in Eq(3.18) in Meddahi (2003). From (32), 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 (21a). Next, note that from (19), we have:

$$\frac{1}{\theta_u} + \theta_u = \frac{1+\theta_u^2}{\theta_u} = \frac{1 + (A - \sqrt{A^2-1})^2}{A - \sqrt{A^2-1}} = 2A. \tag{33}$$

From (17d) and (18a), we have:

$$c_{RV} = (1-\kappa_1)\left(\sigma^2 + 2m\sigma_\varepsilon^2\right), \quad \text{or} \quad \sigma_\varepsilon^2 = \frac{c_{RV} - (1-\kappa_1)\sigma^2}{2(1-\kappa_1)m}. \tag{34}$$

Substituting σ_ε^2 in (34) into A in (19), we have:

$$\begin{aligned}2A &= 2\left[\tfrac{4\sigma^2}{\omega_\varepsilon^2}\left(\tfrac{c_{RV}-(1-\kappa_1)\sigma^2}{2(1-\kappa_1)m}\right) + 2m - 1 + \tfrac{2m}{\omega_\varepsilon^2}\left(\tfrac{c_{RV}-(1-\kappa_1)\sigma^2}{2(1-\kappa_1)m}\right)^2\right] \\ &= \tfrac{2c_{RV}\sigma^2}{(1-\kappa_1)m\omega_\varepsilon^2} - \tfrac{3\sigma^4}{m\omega_\varepsilon^2} + 2(2m-1) + \tfrac{c_{RV}^2}{(1-\kappa_1)^2 m\omega_\varepsilon^2}.\end{aligned} \tag{35}$$

From (31a), (33) and (35), we have:

$$\gamma_0 = 2D\omega_1^2 - \frac{(1+\kappa_1^2)}{m}\sigma^4 + \frac{2(1+\kappa_1^2)c_{RV}}{(1-\kappa_1)m}\sigma^2 + 2(2m-1)(1+\kappa_1^2)\omega_\varepsilon^2 + \frac{(1+\kappa_1^2)c_{RV}^2}{(1-\kappa_1)^2 m} - 2\kappa_1\omega_\varepsilon^2. \tag{36}$$

Multiplying both sides in (36) 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:

$$\sigma^2 = \frac{c_{RV}}{1-\kappa_1} \pm \sqrt{\frac{2c_{RV}^2}{(1-\kappa_1)^2} + 2m(2m-1)\omega_\varepsilon^2 - \frac{m\left(\gamma_0 - 2D\omega_1^2 + 2\kappa_1\omega_\varepsilon^2\right)}{(1+\kappa_1^2)}}. \tag{37}$$

From $\sigma_\varepsilon^2 > 0$, $\kappa_1 < 1$ and (34), we must have $\frac{c_{RV}}{1-\kappa_1} > \sigma^2$. Hence, the sign of the second term in (37) is negative. From (34) and (37), we have:

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

From (37) and (38), we obtain (21b) and (21c).

References

Andersen, T.G. 1994. "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling." *Mathematical Finace* 4: 75–102.

Andersen, T.G., and T. Bollerslev. 1998. "Deutche Mark–Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies." *Journal of Finance* 53: 219–265.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens. 2001. "The Distribution of Stock 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.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2003. "Modeling and Forecasting Realized Volatility." *Econometrica* 71(2): 579-625.

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, Series B (Statistical Methodology)* 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 the Royal Statistical Society, Series B (Statistical Methodology)* 64: 253-280.

Bartlett, M. 1946. "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series." *Journal of the Royal Statistical Society*, Supplement 8: 27-41, 85-97, Corr.(1948) 10: 200.

Beine, M., J. Lahaye, S.C. Laurent, 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.

Brockwell, P.J., and R.A. Davis. 1991. *Time Series: Theory and Methods*, 2nd ed. NY: Springer-Verlag.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay. 1997. *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Chernov, M., A.R., Gallant, E. Ghysels, and G. Tauchen. 2003. "Alternative Models for Stock Price Dynamics." *Journal of Econometrics* 116(1-2): 225-257.

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

Hamilton, D.J. 1994. *Time Series Analysis*. NJ: Princeton University Press.

Harvey, A.C. 1989. *Forecasting, Structural Time Series Model and the Kalman Filter*. Cambridge: Cambridge University Press.

McAleer, M., and M.C. Medeiros. 2008. "Realized Volatility: A Review." *Econometrics Review* 27: 10-45.

Meddahi, N. 2001a. "A Theoretical Comparison Between Integrated and Realized Volatilities." Manuscript, Université de Montréal.

Meddahi, N. 2001b. "An Eigenfunction Approach for Volatility Modeling." Working Paper 2001s-70, CIRANO.

Meddahi, N. 2002. "A Theoretical Comparison Between Integrated and Realized Volatility." *Journal of Applied Econometrics* 17: 479-508.

Meddahi, N. 2003. "ARMA Representation of Integrated and Realized Variances." *Econometrics Journal* 6: 335–356.

Meddahi, N., and E. Renault. 1996. "Aggregation and Marginalization of GARCH and Stochastic Volatility Models." CREMAQ DP 96.30.433, Université de Toulouse and CRDE DP 3597, Université de Montréal.

Meddahi, N., and E. Renault. 2000. "Temporal Aggregation of Volatility Models." Working Paper 2000s-22, CIRANO.

Meddahi, N., and E. Renault. 2004. "Temporal Aggregation of Volatility Models." *Journal of Econometrics* 119: 355-379.

Nagakura, D., and T. Watanabe. 2009. "A State Space Approach to Estimating the Integrated Variance and Microstructure Noise Component." *IMES Discussion Paper Series*, 09-E-11, Institute for Monetary and Economic Studies, Bank of Japan.

Nagakura, D., and T. Watanabe. 2011. "A State Space Approach to Estimating the Integrated Variance under the Existence of Market Microstructure Noise Component." Available at SSRN: http://ssrn.com/abstract=1350210

Nelson, D.B. 1990. "ARCH Models as Diffusion Approximations." *Journal of Econometrics* 45: 7–39.

Owens, J.P., and D.G. Steigerwald. 2006. "Noise Reduced Realized Volatility: A Kalman Filter Approach." *Advances in Econometrics* 20: 211-227.

Patton, J.A. 2011. "Volatility Forecast Comparison Using Imperfect Volatility Proxies." *Journal of Econometrics* 160: 246-256.

Romano, P.J., and L.A. Thombs. 1996. "Inference For Autocorrelations Under Weak Assumptions." *Journal of the American Statistical Association* 91(434): 590-600.

Ubukata, M., and K. Oya. 2009. "Estimation and Testing for Dependence in Market Microstructure Noise." *Journal of Financial Econometrics* 7(2): 106-151.

Notes

[1] The discretization error is uncorrelated with the IV under the assumption of no "leverage effect". See Meddahi (2002) for more details.

[2] We call the method so because we apply the state space method of Barndorff-Nielsen and Shephard (2002) combined with the ARMA representation result of Meddahi (2003).

[3] The independence between $W(t)$ and $f(t)$ implies that there is no leverage effect in this price process.

[4]The i.i.d. assumption can be relaxed to some extent. See Nagakura and Watanabe (2011).

[5]Note that here η_t and $\xi_t^{(m)}$ do not follow a Gaussian distribution. In this case, the Kalman filter provides the best linear estimator (Hamilton, 1994, Chapter 13).

[6] The analysis in the subsequent subsections also shows that the identification restriction suggested in Barndorff-Nielsen and Shephard (2002) does not work to identify the state space form parameters (and is not necessary to identify the SV model parameters) in the current context.

[7]It is possible to apply the GMM estimation instead of the QML estimation in this step.

[8]These returns are calculated from the price data combined after the removals of the prices of inactive trading days. The way of adjusting the data is slightly different from the previous version of the paper (Nagakura and Watanabe, 2009), where we first calculate returns, and then remove the returns of inactive trading days according to the above criteria. The results obtained are very similar.

[9] A loss function is said to be "robust" if the ranking of IV forecasts by that loss function with an IV proxy is the same as the ranking done with the true value of the IV. The loss functions, R^2, MSE, and QLIKE are robust if the IV proxy is conditionally unbiased (Meddahi, 2001a; Patton, 2011).

[10]We use a numerical maximization procedure for maximizing the log-likelihood. When the numerical maximization procedure results in odd estimates, we re-run the procedure with different initial values. We repeat this till we get reasonable estimates.

[11]We thank an anonymous referee for pointing this.

Table 1: Descriptive Statistics of Returns

	$r(1)$	$r(5)$	$r(10)$	$r(15)$	$r(30)$
m	1440	288	144	96	48
Mean×1000	0.0059	0.0302	0.0604	0.0906	0.1812
Variance×10	0.0037	0.0143	0.0269	0.0390	0.0747
SD	0.0193	0.0377	0.0518	0.0625	0.0864
SAC(1)	−0.1501	−0.0589	−0.0520	−0.0472	−0.0217
	(0.0006)	(0.0014)	(0.0019)	(0.0024)	(0.0034)
SAC(2)	0.0038	−0.0151	−0.0100	0.0030	0.0037
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(3)	−0.0033	−0.0095	0.0010	0.0023	−0.0033
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(4)	−0.0034	−0.0042	0.0059	0.0019	0.0064
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(5)	−0.0051	0.0007	0.0000	−0.0008	−0.0069
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(6)	−0.0024	0.0007	0.0007	−0.0052	−0.0083
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(7)	−0.0028	−0.0006	0.0018	0.0074	0.0007
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(8)	−0.0019	0.0042	−0.0004	0.0036	−0.0006
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(9)	−0.0042	0.0014	−0.0040	−0.0055	0.0004
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)
SAC(10)	−0.0011	−0.0013	−0.0001	−0.0020	−0.0070
	(0.0006)	(0.0014)	(0.0020)	(0.0024)	(0.0034)

Note: the table reports the sample means (Mean), sample variances (Variance), sample standard deviations (SD), and k-th order sample autocorrelations (SAC(k)) for return series with different sampling frequencies. $r(k)$ denotes k-minute return and m is the number of intra-day returns for each return series. The asymptotic standard errors, which are estimated based on Bartlett's (1946) formula for MA(1) process, are in parentheses. These asymptotic standard errors are valid only if the time series follows an MA(1) process driven by an i.i.d. innovation with finite second moment, and may actually be very misleading otherwise (see Romano and Thombs, 1996). We just followed the custom and do not claim anything based on these standard errors.

Table 2: Descriptive Statistics of the NCRV

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)	NCRV(30)
m	1440	288	144	96	48
Mean	0.5382	0.4104	0.3870	0.3745	0.3583
Variance	0.0684	0.0682	0.0712	0.0696	0.0833
SD	0.2616	0.2611	0.2669	0.2638	0.2887
SAC(1)	0.4573	0.3975	0.3634	0.3402	0.2595
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(2)	0.3408	0.3021	0.2566	0.2452	0.1585
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(3)	0.3092	0.2590	0.2274	0.2005	0.1570
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(4)	0.3091	0.2365	0.2074	0.2047	0.1394
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(5)	0.2886	0.2239	0.2071	0.1943	0.1568
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(6)	0.2617	0.1795	0.1608	0.1699	0.1547
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(7)	0.2484	0.1754	0.1571	0.1444	0.0954
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(8)	0.2455	0.1591	0.1524	0.1466	0.0855
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(9)	0.2228	0.1531	0.1312	0.1327	0.0753
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)
SAC(10)	0.2521	0.1713	0.1537	0.1523	0.1023
	(0.0235)	(0.0235)	(0.0235)	(0.0235)	(0.0235)

Note: the table reports the sample mean (Mean), sample variance (Variance), sample standard deviation (SD), and k-th order sample autocorrelation (SAC(k)) for various NCRV series with different sampling frequencies. NCRV(k) denotes k-minute NCRV series and m is the number of intra-day returns used for calculating each NCRV series. The asymptotic standard errors, or $1/\sqrt{N}$, where N is the number of samples (see Brockwell and Davis , 1991, p.222), are in parentheses. These asymptotic standard errors are valid only if the time series is $i.i.d$ with finite second moment, and may actually be very misleading otherwise (see Romano and Thombs, 1996). We just followed the custom and do not claim anything based on these standard errors.

Table 3: Estimates of SV Model Parameters

(a) One-factor model

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)
m	1440	288	144	96
$\widehat{\kappa}_1$	0.9352	0.8783	0.8859	0.9075
	(0.0461)	(0.0513)	(0.0581)	(0.0624)
$\widehat{\sigma}^2$	0.2905	0.3523	0.3565	0.3549
	(0.0281)	(0.0180)	(0.0190)	(0.0204)
$\widehat{\omega}_1^2$	0.0308	0.0292	0.0265	0.0230
	(0.0121)	(0.0097)	(0.0090)	(0.0082)
$\widehat{\sigma}_\varepsilon^2 \times 100$	0.0087	0.0102	0.0107	0.0105
	(0.0010)	(0.0003)	(0.0010)	(0.0020)
$\widehat{\omega}_\varepsilon^2 \times 1000$	0.0067	0.0339	0.0756	0.1153
	(0.0010)	(0.0049)	(0.0112)	(0.0159)
L	138.257	70.158	-5.555	0.203

(b) Two-factor model

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)
$\widehat{\kappa}_1$	0.9814	0.9786	0.9757	0.9773
	(0.0181)	(0.0153)	(0.0146)	(0.0138)
$\widehat{\kappa}_2$	0.2890	0.5621	0.4418	0.4206
	(0.2428)	(0.1575)	(0.2068)	(0.1945)
$\widehat{\sigma}^2$	0.3323	0.3509	0.3577	0.3555
	(0.0541)	(0.0328)	(0.0326)	(0.0343)
$\widehat{\omega}_1^2$	0.0240	0.0152	0.0148	0.0140
	(0.0152)	(0.0063)	(0.0051)	(0.0051)
$\widehat{\omega}_2^2$	0.0308	0.0219	0.0249	0.0230
	(0.0293)	(0.0091)	(0.0160)	(0.0147)
$\widehat{\sigma}_\varepsilon^2 \times 100$	0.0074	0.0105	0.0105	0.0101
	(0.0014)	(0.0008)	(0.0022)	(0.0042)
$\widehat{\omega}_\varepsilon^2 \times 1000$	0.0044	0.0295	0.0608	0.0911
	(0.0023)	(0.0057)	(0.0156)	(0.0208)
L	159.2571	83.565	8.143	13.597

Note: the table shows the estimates of the SV model parameters in (1) and (2) by the method described in Section (3.4), NCRV(k) denotes k-minute NCRV, and L is the (quasi) log-likelihood. The robust standard errors are in parentheses. The robust standard errors are obtained as follows. First, by the QML, we estimate a reparameterized SV model such that $\mu \equiv \log(\sigma^2)$, so that we can apply unconstrained optimization procedures in maximizing the log-likelihood. Then, the QML estimate of, for example, σ^2 is obtained as $\exp(\widehat{\mu})$, where $\widehat{\mu}$ is the QML estimate of μ. Next, generate samples from the normal distribution with mean and covariance matrix being set to estimates of the reparameterized model parameters (such as $\widehat{\mu}$) and the robust estimate of their asymptotic covariance matrix, respectively. For each sample, calculate the corresponding SV model parameters. Finally, calculate the sample standard deviations of these (generated) SV model parameter, which are our (approximate) robust standard errors.

Table 4: Estimates of State Space Form Parameters

(a) One-factor model

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)
$\widehat{c}_{IV}$	0.0188	0.0429	0.0407	0.0328
$\widehat{\kappa}_1$	0.9352	0.8783	0.8859	0.9075
$\widehat{\theta}_1$	0.2679	0.2677	0.2677	0.2678
$\widehat{\sigma}^2_\eta$	0.0024	0.0041	0.0035	0.0025
m	1440	288	144	96
$\widehat{c}^{(m)}_u$	0.2450	0.0586	0.0308	0.0201
$\widehat{\theta}^{(m)}_u$	0.0002	0.0009	0.0017	0.0026
$\widehat{\sigma}^{2(m)}_\xi$	0.0391	0.0393	0.0437	0.0444
$\widehat{\sigma}^{2(m)}_d$	0.0002	0.0011	0.0023	0.0031

(b) Two-factor model

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)
$\widehat{c}_{IV}$	0.0044	0.0033	0.0049	0.0047
$\widehat{\phi}_1$	1.2704	1.5407	1.4175	1.3979
$\widehat{\phi}_2$	−0.2836	−0.5501	−0.4311	−0.4110
$\widehat{\theta}_1$	−0.6271	−0.6470	−0.6368	−0.6376
$\widehat{\theta}_2$	−0.2154	−0.2394	−0.2311	−0.2300
$\widehat{\sigma}^2_\eta$	0.0183	0.0101	0.0133	0.0125
m	1440	288	144	96
$\widehat{c}^{(m)}_u$	0.2121	0.0604	0.0293	0.0196
$\widehat{\theta}^{(m)}_u$	0.0002	0.0009	0.0017	0.0026
$\widehat{\sigma}^{2(m)}_\xi$	0.0255	0.0342	0.0352	0.0351
$\widehat{\sigma}^{2(m)}_d$	0.0002	0.0011	0.0023	0.0034

Note: ϕ_1 and ϕ_2 are defined as $\phi_1 = \kappa_1 + \kappa_2$ and $\phi_2 = -\kappa_1\kappa_2$. See Nagakura and Watanabe (2011) for more details on the state space from of the two factor SR-SARV model.

Table 5: Estimates of Some Important Values

(a) One-factor model

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)
$\widehat{V}_{IV}$	0.0301	0.0279	0.0254	0.0223
$\widehat{AC}_{IV}(1)$	0.9566	0.9180	0.9232	0.9378
$\widehat{AC}_{IV}(2)$	0.8946	0.8062	0.8178	0.8511
$\widehat{AC}_{IV}(3)$	0.8367	0.7081	0.7245	0.7724
$\widehat{AC}_{IV}(4)$	0.7825	0.6219	0.6419	0.7010
$\widehat{AC}_{IV}(5)$	0.7318	0.5462	0.5686	0.6361
$\widehat{V}_{u}^{(m)}$	0.0391	0.0393	0.0437	0.0444
$\widehat{V}_{IV}/\widehat{V}_{NCRV}^{(m)}$	0.4341	0.4089	0.3568	0.3195
$\widehat{V}_{u}^{(m)}/\widehat{V}_{NCRV}^{(m)}$	0.5637	0.5755	0.6133	0.6360

(a) Two-factor model

	NCRV(1)	NCRV(5)	NCRV(10)	NCRV(15)
$\widehat{V}_{IV}$	0.0450	0.0333	0.0340	0.0314
$\widehat{AC}_{IV}(1)$	0.7469	0.8259	0.7666	0.7616
$\widehat{AC}_{IV}(2)$	0.5776	0.6499	0.5650	0.5622
$\widehat{AC}_{IV}(3)$	0.5220	0.5470	0.4704	0.4728
$\widehat{AC}_{IV}(4)$	0.4993	0.4853	0.4232	0.4299
$\widehat{AC}_{IV}(5)$	0.4863	0.4467	0.3971	0.4066
$\widehat{V}_{u}^{(m)}$	0.0255	0.0342	0.0352	0.0351
$\widehat{V}_{IV}/\widehat{V}_{NCRV}^{(m)}$	0.6365	0.4851	0.4752	0.4492
$\widehat{V}_{u}^{(m)}/\widehat{V}_{NCRV}^{(m)}$	0.3603	0.4987	0.4922	0.5022

Note: $\widehat{AC}_{IV}(k)$ is the estimate of the k-th order autocorrelation of IV_t. $\widehat{V}_{IV}$, $\widehat{V}_{u}^{(m)}$, and $\widehat{V}_{NCRV}$ are the estimates of unconditional variances of IV_t, $u_t^{(m)}$, and $RV_t^{(m)}$, respectively. These estimates are obtained based on the estimated values of state space form parameters.

Table 6: Forecast Evaluation (In-sample Case)

(a) Results of Mincer - Zarnowitz style regression

	One-factor model				Two-factor model			
m	a_0	a_1		R^2	a_0	a_1		R^2
1440	0.1678	0.6581	–	0.0966	0.1442	0.6515	–	0.1022
	(0.0190)	(0.0698)	–		(0.0219)	(0.0705)	–	
288	0.0636	0.8362	–	0.1179	0.0633	0.8396	–	0.1261
	(0.0280)	(0.0826)	–		(0.0274)	(0.0807)	–	
144	0.0457	0.8763	–	0.1134	0.0458	0.8713	–	0.1204
	(0.0261)	(0.0763)	–		(0.0262)	(0.0759)	–	
96	0.0368	0.9055	–	0.1091	0.0369	0.9020	–	0.1169
	(0.0261)	(0.0762)	–		(0.0265)	(0.0769)	–	
m	b_0		b_1	R^2	b_0		b_1	R^2
1440	−0.0382	–	0.7365	0.0750	0.0056	–	0.6525	0.1015
	(0.0625)	–	(0.1191)	–	(0.0366)	–	(0.0708)	–
288	−0.0170	–	0.9138	0.0845	0.0087	–	0.8499	0.1255
	(0.0567)	–	(0.1414)	–	(0.0325)	–	(0.0819)	–
144	0.0909	–	0.6817	0.1094	0.0179	–	0.8777	0.1204
	(0.0206)	–	(0.0541)	–	(0.0283)	–	(0.0759)	–
96	0.0914	–	0.7020	0.1043	0.0177	–	0.9067	0.1162
	(0.0206)	–	(0.0551)	–	(0.0281)	–	(0.0772)	–
m	c_0	c_1	c_2	R^2	c_0	c_1	c_2	R^2
1440	0.0551	0.4931	0.3025	0.1032	0.1823	0.8280	−0.1778	0.1022
	(0.0671)	(0.0778)	(0.1494)	–	(0.1972)	(0.9185)	(0.9190)	–
288	0.0077	0.6856	0.2653	0.1212	0.0400	0.5123	0.3361	0.1265
	(0.0596)	(0.0981)	(0.1970)	–	(0.0447)	(0.5307)	(0.5355)	–
144	0.0531	0.5640	0.2652	0.1156	0.0311	0.4570	0.4209	0.1209
	(0.0274)	(0.1606)	(0.2232)	–	(0.0326)	(0.6195)	(0.6206)	–
96	0.0463	0.6234	0.2383	0.1105	0.0304	0.6421	0.2639	0.1171
	(0.0284)	(0.1687)	(0.2319)	–	(0.0308)	(0.6835)	(0.6866)	–

(b) Mean squared error (MSE)

	One-factor model		Two-factor model	
m	$\text{MSE}^{(NR)}$	$\text{MSE}^{(BSM)}$	$\text{MSE}^{(NR)}$	$\text{MSE}^{(BSM)}$
1440	0.0822	0.1103	0.0785	0.1104
288	0.0740	0.0791	0.0732	0.0760
144	0.0740	0.0774	0.0735	0.0743
96	0.0743	0.0766	0.0737	0.0766

(c) QLIKE

	One-factor model		Two-factor model	
m	$\text{QLIKE}^{(NR)}$	$\text{QLIKE}^{(BSM)}$	$\text{QLIKE}^{(NR)}$	$\text{QLIKE}^{(BSM)}$
1440	0.0182	0.0281	−0.0219	0.0192
288	−0.0604	−0.0395	−0.0599	−0.0531
144	−0.0585	−0.0534	−0.0614	−0.0589
96	−0.0596	−0.0544	−0.0610	−0.0544

Note: tables (a), (b), and (c) reports the results of the Mincer–Zarnowitz style regressions, the MSE, and QLIKE of one-day-ahead in-sample forecasts, respectively.

Table 7: Forecast Evaluation (Out-of-sample Case)

(a) Results of Mincer - Zarnowitz style regression

	One-factor model				Two-factor model			
m	a_0	a_1		R^2	a_0	a_1		R^2
1440	0.1003	0.8469	–	0.0707	0.1091	0.8301	–	0.0752
	(0.0286)	(0.1292)	–	–	(0.0268)	(0.1284)	–	–
288	−0.0059	1.0272	–	0.1032	−0.0001	0.9998	–	0.1050
	(0.0362)	(0.1239)	–	–	(0.0350)	(0.1195)	–	–
144	−0.00002	0.9571	–	0.1012	0.0038	0.9297	–	0.0990
	(0.0401)	(0.1323)	–	–	(0.0402)	(0.1308)	–	–
96	−0.0176	1.0143	–	0.1032	−0.0103	0.9990	–	0.1052
	(0.0383)	(0.1238)	–	–	(0.0367)	(0.1212)	–	–
m	b_0		b_1	R^2	b_0		b_1	R^2
1440	−0.1506	–	0.8879	0.0390	−0.1190	–	0.8754	0.0701
	(0.1333)	–	(0.2721)	–	(0.0601)	–	(0.1311)	–
288	0.0666	–	0.6813	0.0956	−0.0900	–	1.0759	0.1042
	(0.0299)	–	(0.0868)	–	(0.0453)	–	(0.1269)	–
144	0.0668	–	0.7185	0.0973	−0.0522	–	1.0290	0.0992
	(0.0298)	–	(0.0903)	–	(0.0427)	–	(0.1273)	–
96	0.0619	–	0.7540	0.0972	−0.0586	–	1.0802	0.1005
	(0.0322)	–	(0.0976)	–	(0.0440)	–	(0.1318)	–
m	c_0	c_1	c_2	R^2	c_0	c_1	c_2	R^2
1440	0.0137	0.7501	0.2153	0.0721	0.1446	0.9457	−0.1297	0.0753
	(0.1436)	(0.1723)	(0.3417)	–	(0.1624)	(0.5897)	(0.6039)	–
288	−0.0075	1.0565	−0.0209	0.1032	−0.0343	0.6468	0.3866	0.1054
	(0.0431)	(0.4799)	(0.3276)	–	(0.0754)	(0.7075)	(0.7554)	–
144	0.0153	0.6240	0.2742	0.1031	−0.0316	0.4663	0.5348	0.1012
	(0.0428)	(0.3870)	(0.2839)	–	(0.0456)	(0.4711)	(0.4911)	–
96	−0.0041	0.7758	0.1938	0.1039	−0.0149	0.9303	0.0781	0.1053
	(0.0417)	(0.4281)	(0.3371)	–	(0.0599)	(0.7111)	(0.7868)	–

(b) Mean squared error (MSE)

	One-factor model		Two-factor model	
m	$\text{MSE}^{(NR)}$	$\text{MSE}^{(BSM)}$	$\text{MSE}^{(NR)}$	$\text{MSE}^{(BSM)}$
1440	0.0530	0.0938	0.0537	0.0810
288	0.0472	0.0493	0.0471	0.0511
144	0.0475	0.0489	0.0478	0.0492
96	0.0474	0.0483	0.0472	0.0484

(c) QLIKE

	One-factor model		Two-factor model	
m	$\text{QLIKE}^{(NR)}$	$\text{QLIKE}^{(BSM)}$	$\text{QLIKE}^{(NR)}$	$\text{QLIKE}^{(BSM)}$
1440	−0.1688	−0.0899	−0.1530	−0.1195
288	−0.2251	−0.2184	−0.2254	−0.2059
144	−0.2246	−0.2208	−0.2227	−0.2153
96	−0.2240	−0.2208	−0.2251	−0.2184

Note: tables (a), (b), and (c) reports the results of the Mincer–Zarnowitz style regressions, the MSE, and QLIKE of one-day-ahead in-sample forecasts, respectively.

Figure 1: Daily Returns of the Yen/Dollar Exchange Rate

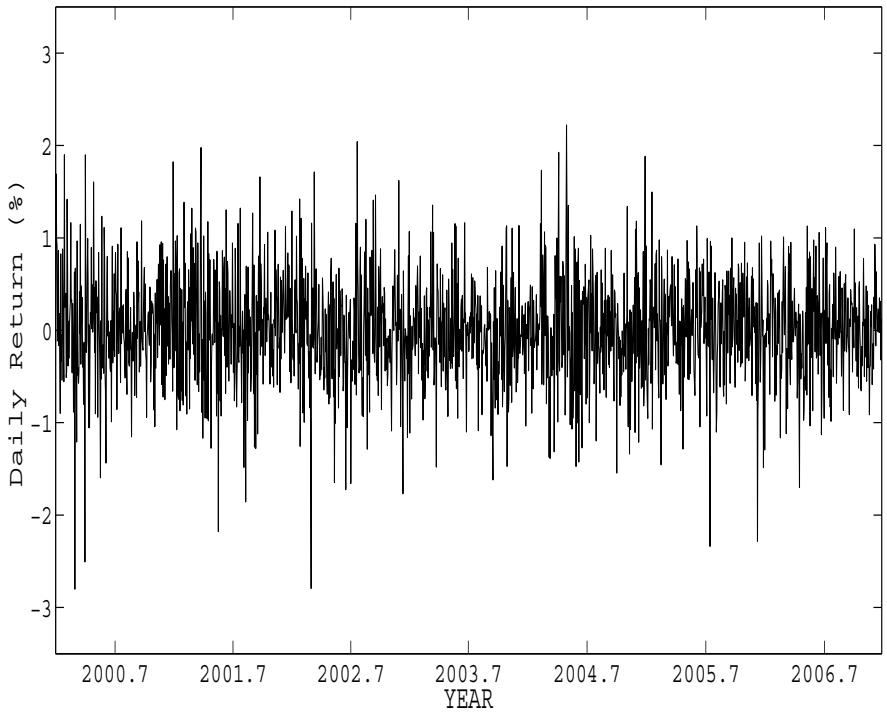


Figure 2: NCRV, Smoothed IV, and MN component Series (One-factor Model)

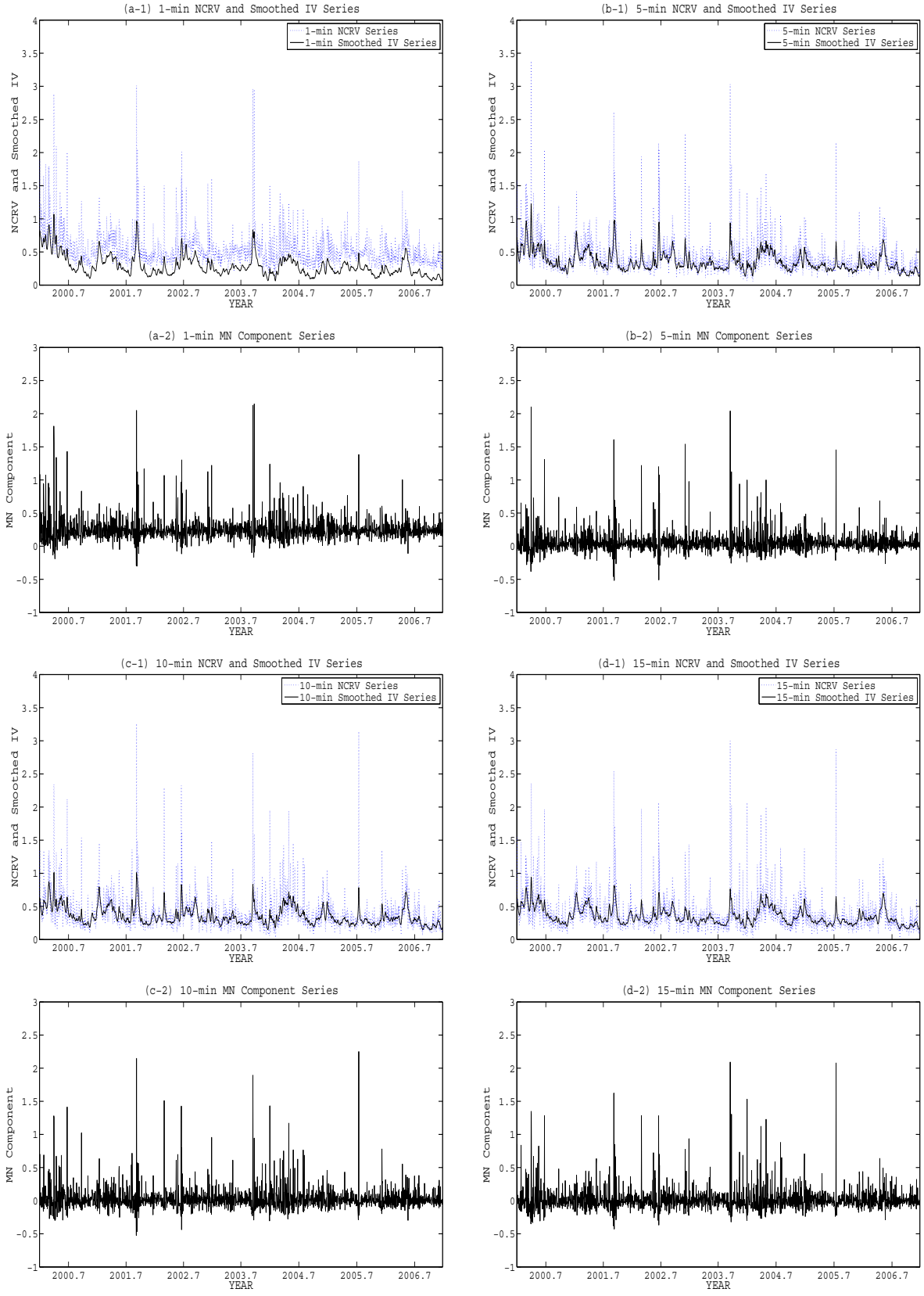


Figure 3: NCRV, Smoothed IV, and MN component Series (Two-factor Model)

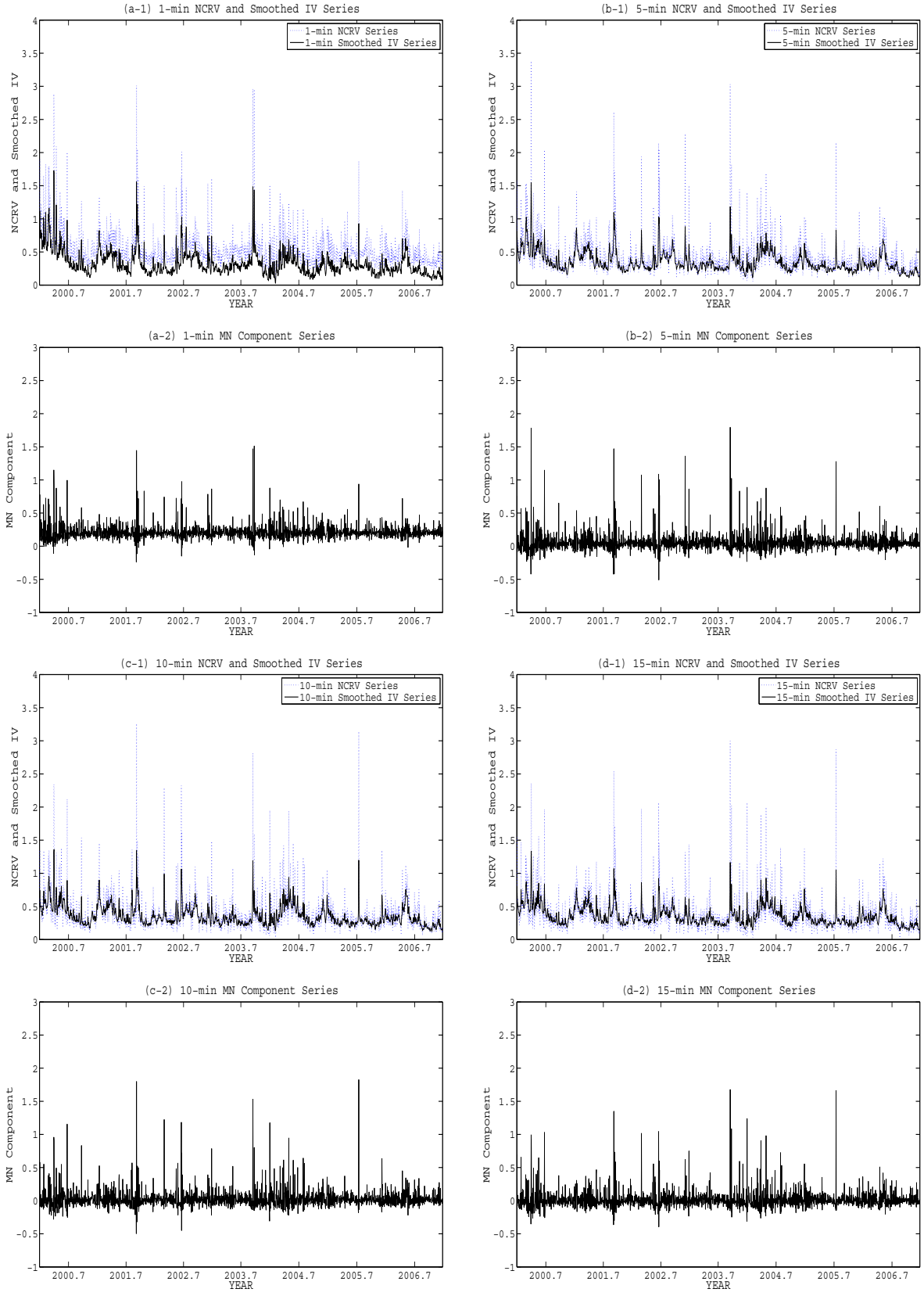


Figure 4: In Sample Forecasts (One-factor Model)

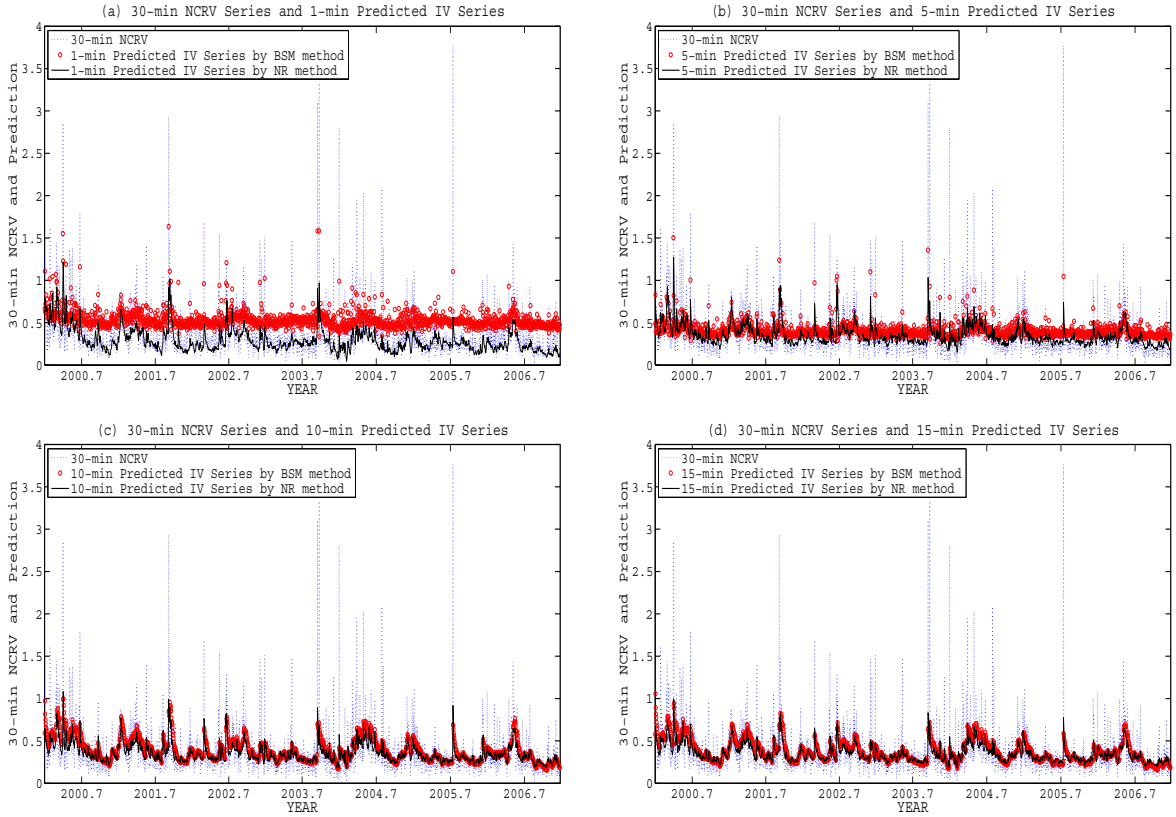


Figure 5: In Sample Forecasts (Two-factor Model)

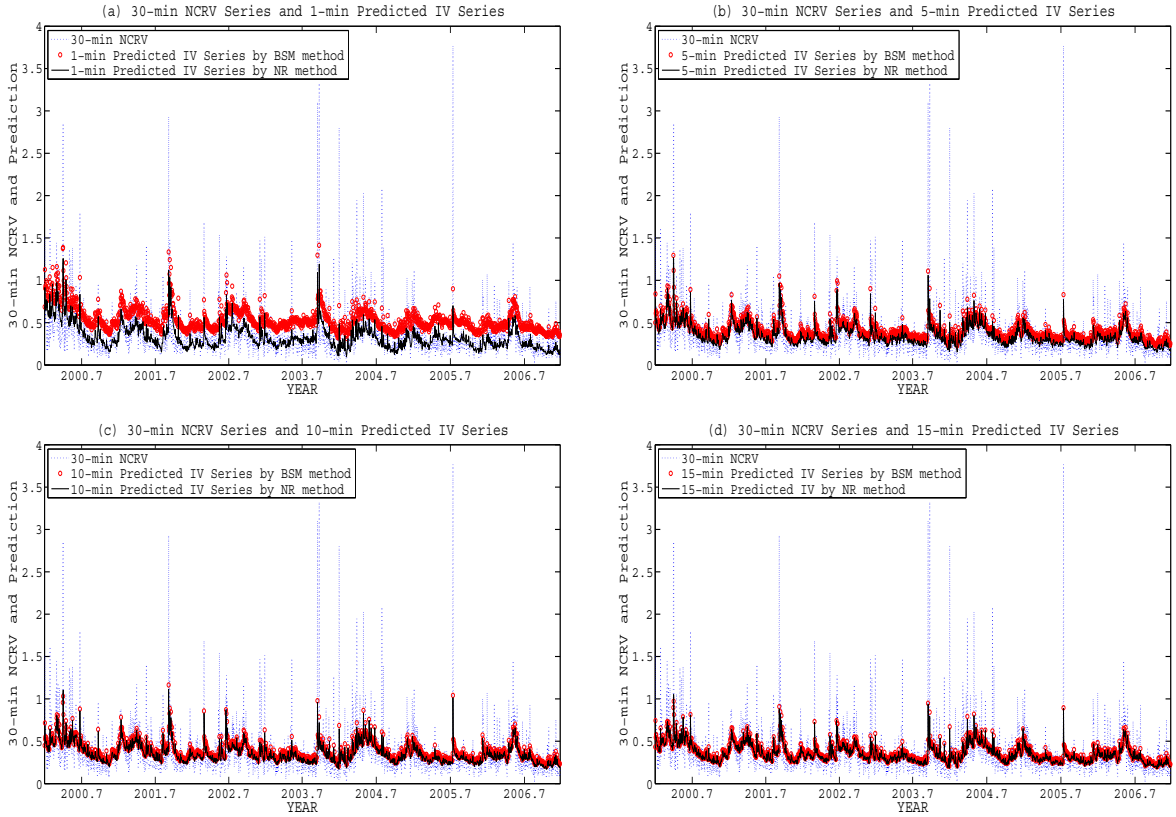


Figure 6: Out of Sample Forecasts (One-factor Model)

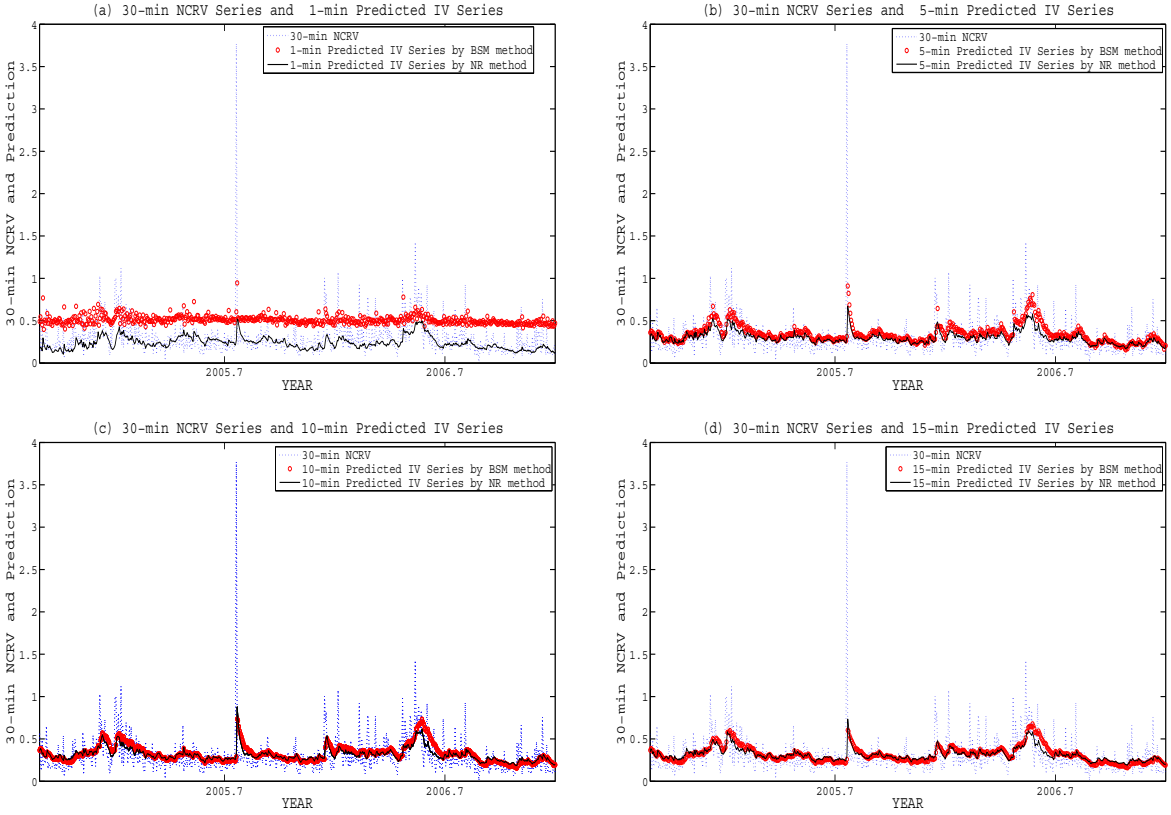


Figure 7: Out of Sample Forecasts (Two-factor Model)

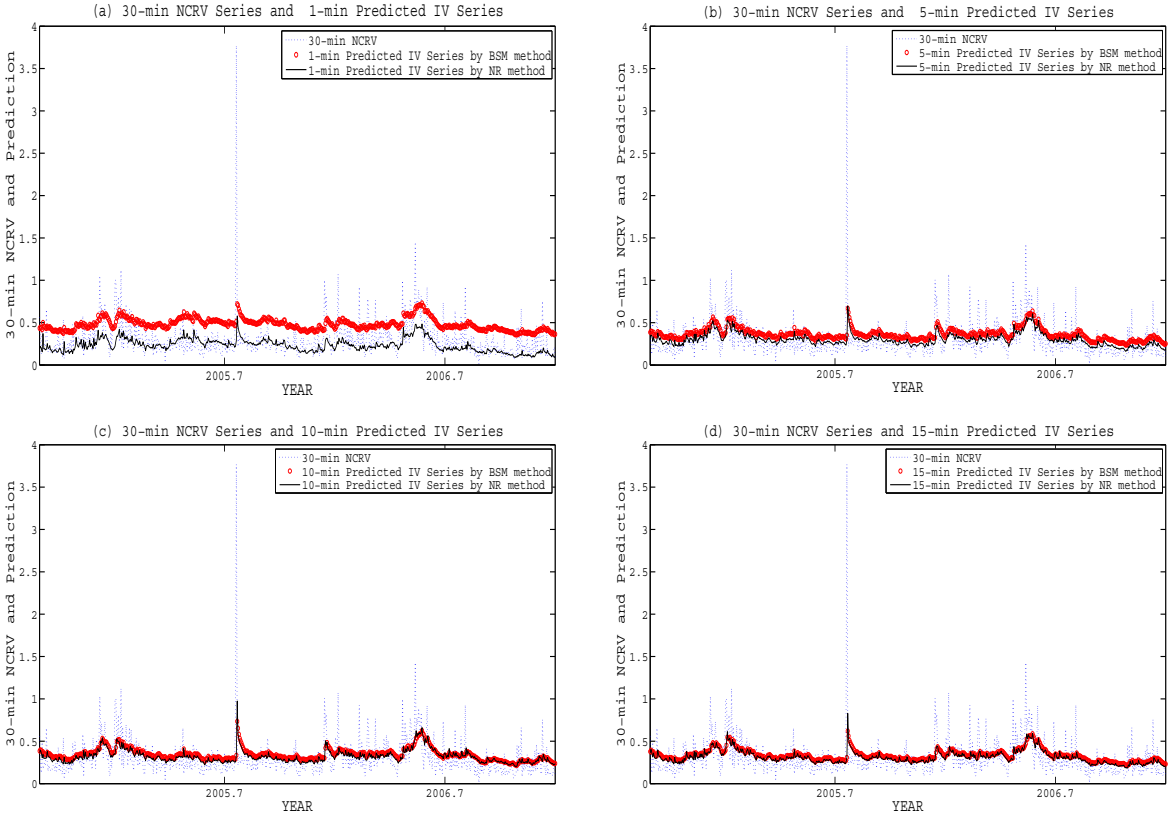
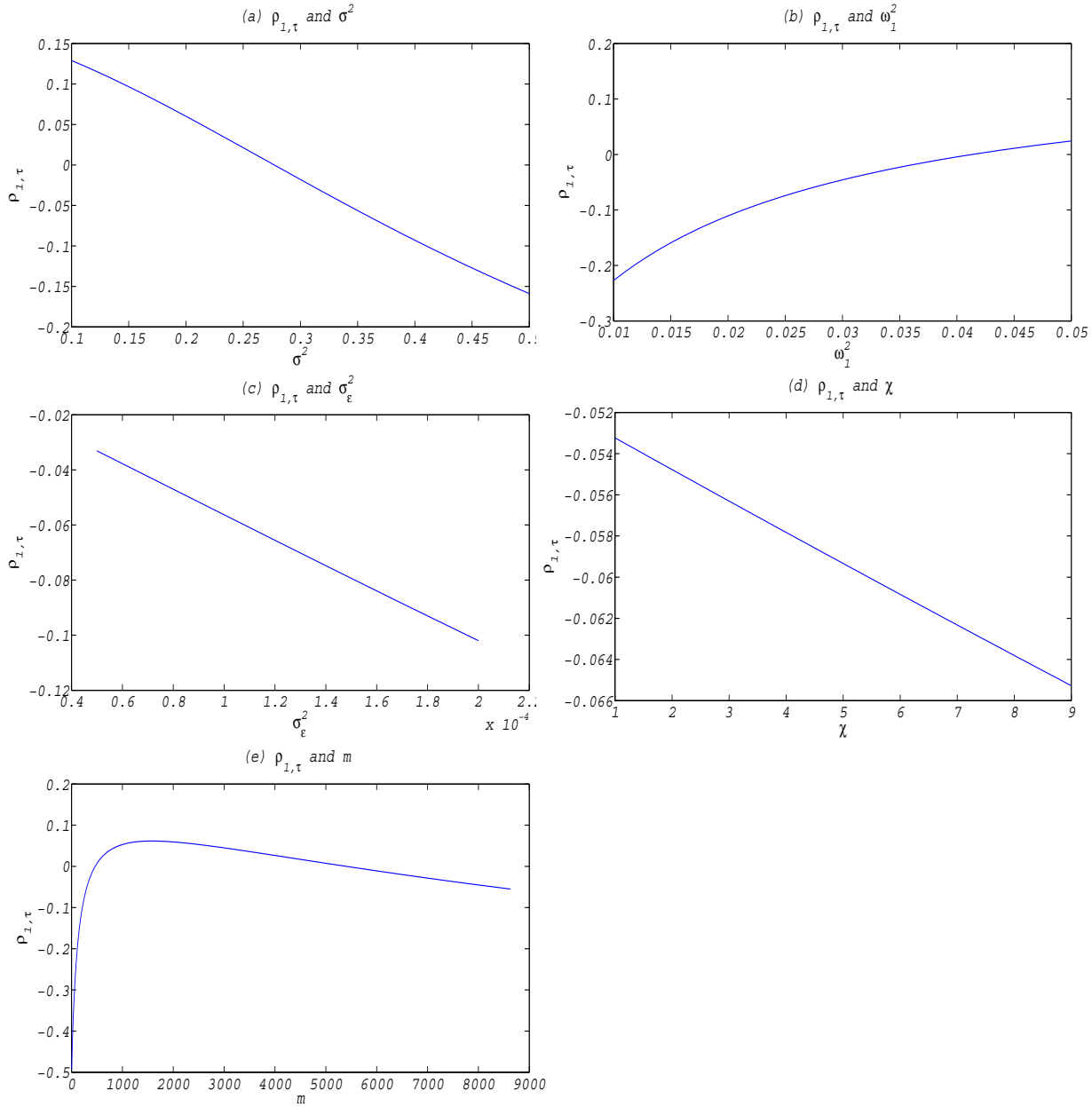


Figure 8: Scatter plot of $\rho_{1,\tau}$ and SV model parameters



Note: in drawing the scatter plots, SV model parameters are fixed at $\sigma^2 = 0.35$, $\omega_1^2 = 0.028$, $\sigma_\varepsilon^2 = 0.0001$, $\chi = 3$, and $m = 288$.