HITOTSUBASHI UNIVERSITY

Global COE Hi-Stat Discussion Paper Series 115

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

February 2010

Hi-Stat Discussion Paper

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

Daisuke Nagakura [†] and Toshiaki Watanabe [‡]

Abstract

We call the realized variance (RV) calculated with observed prices contaminated by (market) microstructure noises (MNs) the *noise-contaminated RV* (NCRV), referring 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. 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. 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.

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

1 Introduction

The variance of financial asset returns is known to change over time. More precisely, 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. 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, Engle and Nelson (1994), Palm (1996) and Zivot (2008) for comprehensive surveys on GARCH models, Ghysels, Harvey and Renault (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.

In the finance literature of continuous-time models, the value of the *integrated variance* (IV)[1] has played a crucial role for option pricing, risk management, optimal portfolio construction, etc. Roughly speaking, the IV is a cumulative sum (or integral) of continuously changing instantaneous variances over a specified period, for example, a day (a formal definition of IV will be given in Section 2). It has been used as a measure of the variability, or risk, of financial asset returns. Various methods, model-based and model-free, have been proposed for estimating the IV. We will briefly review a few of those recently developed later in this section. See McAleer and Medeiros (2008) for more details on those and other 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 (2001), Barndorff-Nielsen and Shephard (2002), Andersen, Bollerslev, Diebold, and Ebens (2001), and Andersen, Bollerslev, Diebold and Labys (2001) among others. 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 defined as a sum of squared intra-period returns and is a model-free estimator in the sense that we do not need a parametric specification for the volatility dynamics. 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. 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 causes of the MN. 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.

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.[2] 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.[3] 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 (Andersen, 1994; Meddahi and Renault, 2004). Meddahi (2003) derives explicit relationships between the ARMA model parameters and the SV model parameters.

---

[1]Instead of integrated variance, the term "integrated volatility" is often used in the literature. We do not use this term because integrated variance is a more appropriate term for what we estimate.

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

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

In this paper, we develop the state space method by Barndorff-Nielsen and Shephard (2002), using 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 serially uncorrelated (but not necessarily i.i.d.) MNs (see Assumptions 1 and 2 in the next section for more details). We call the RV calculated with observed log-prices (that are assumed to be contaminated by MN) 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 2.3). We show that the MN component follows a MA(1) 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. 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.

One of the difficulties of state space approaches is verifying identifiability of state space form parameters. In general, parameters of a state space form are not necessarily identified (see, for example, Hamilton (1994, p.388) and Harvey (1989, p.205)[4]). Hence, one has to check the identifiability before going to the estimation. We investigate the autocovariance structure of the NCRV, and show that the state space form parameters of the NCRV are *not* identified in the sense that different sets of parameter values can give the same autocovariance structure. See Section 4 for more details. This implies that the state space form parameters cannot be directly estimated from the second order moment properties of the NCRV. We solve the problem in this paper. We first 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 the variance parameters of the MN and its square. Then, we prove that these MN and SV model parameters are *uniquely* identified. We illustrate how to estimate these identifiable parameters and then the state space form parameters.[5] With estimates of the state space form parameters, one can estimate the IV and MN components simultaneously by applying the Kalman filter or smoother to the state space form. One advantage of our method, compared to other existing methods described below, is that it can filter out not only the MN components but also the discretization errors.

Our estimation strategy is model-based because we assume a certain general class of continuous-time SV models for the underlying true log price process. Several MN-robust and model-free IV estimators have developed. Zhang *et al.* (2005) propose the first consistent IV estimator under the existence of i.i.d. MN. This estimator is called the "two-time scale RV (TSRV)". They show that the TSRV is $n^{1/6}$ consistent. Zhang (2006) extends the TSRV, maintaining the i.i.d. MN assumption, and propose "multi-time scale RV (MSRV)", which achieves $n^{1/4}$ consistency. Barndorff-Nielsen *et al.* (2008) propose a unified approach that they call "realized kernel (RK)" and show that RK can have the same convergence rates and asymptotic distribution as TSRV or MSRV by properly choosing kernels.[6] See also Zhou (1996) Hansen and Lunde (2006), Bandi and Russell (2006), and Bandi and Russell (2008), for earlier works on model-free IV estimators under the existence of MN, and Andersen, Dobrev, and Schaumburg (2009), Hansen and Horel (2009), Jacod, *et al.* (2009), Ubukata and Oya (2009), and references therein for recent developments on some problems related to the MN.

An advantage of model-based methods is that it naturally provides a forecasting model unlike model-free IV estimators such as the ones mentioned above. Model-free IV estimators provide ex-post estimates of IV, and do not provide any forecasting model. Thus one has to elaborate a good forecasting model separately from the problem of estimating the IV (cf. Andersen, Bollerslev, Diebold, and Labys, 2003). Also, model-based estimation methods are generally more efficient than model-free estimation

---

[4]See also Morley, Nelson and Zivot (2003) on the identification problem for some state space models

[5]Alternatively, it might be possible to estimate the parameters using higher order moment properties of the NCRV although proving it seems intractable and would require more stringent assumptions on moments of the MN than the present paper does.

[6]Our state space, or filtering, method may be extended for filtering these model-free IV estimators, however, it is beyond the scope of the current paper and would be a subject of future research.

methods if the true process can be well-approximated by the supposed model. See Jungbacker and Koopman (2006, p.189) and Aït-Sahalia, Mykland and Zhang (2005) for discussions on the advantages of model-based methods. Considering the well-known difficulty for estimating continuous-time SV model parameters, 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 compare forecasting performances of our and the Barndorff-Nielsen, Shephard, and Meddahi (hereafter BSM) state space methods[7] in our empirical analysis. We find that, in terms of mean square error type forecasting criteria, 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. See Section 4 for more details.

We note two other recently proposed state space methods that are also designed to work under the existence of MN. Jungbacker and Koopman (2006) model a high-frequency return as a sum of an i.i.d. MN and a discrete SV model.[8] This discrete SV model is an approximation to a simple continuous-time SV model that is included as a special case of our more general class of continuous-time SV models. They take into account of a diurnal pattern of daily volatility dynamics and try to capture it by a flexible deterministic function. Owens and Steigerwald (2006) propose a state space method of reducing the bias in RV caused by MN. They represent the observed high frequency return as a state space form so that the return is the sum of the true return and a MA(1) noise. Then, they define their IV estimator as the sum of squares of smoothed returns plus bias correction terms, where the smoothed returns are obtained by running the Kalman smoother with estimated time-varying variances of the true returns. Owens and Steigerwald (2006) provide only a limited theoretical justification for their method, in particular, the method of estimating the time-varying variances of the true returns.

A fundamental difference between our and their state space methods is that their method is based on state space forms of observed (noise-contaminated) high frequency returns, while our method is based on a state space form of the NCRV. As a result, the estimation of our method is substantially easier to implement than Jungbacker and Koopman's (2006) method. Jungbacker and Koopman's (2006) modeling framework is essentially the same as that of usual discrete-time SV models for daily returns, in that they replace the daily returns simply with high frequency returns, and therefore the same difficulty for estimation (or rather increased difficulty due to the existence of MN) arises, which requires a sophisticated, computer-intensive, and time-consuming method for estimating the model parameters. By contrast, the estimation of our state space method is only as difficult as the estimation of ARMA or linear state space models.[9] Jungbakcer and Koopman's (2006) SV model is an approximation to a simple continuous-time SV model, and for the approximation to work, the sampling frequency needs to be high. On the other hand, apart from the assumptions on the MN, our state space form is not an approximation but exactly holds for a general class of SV models regardless of sampling frequencies. Also, unlike our method (and Barndorff-Nielsen and Shephard's (2002) method for no MN case), neither Jungbacker and Koopman's (2006) method nor Owen and Steigerwald's (2006) method possesses the advantage of being able to filter out the discretization errors.

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 detail, briefly summarizing the results in Meddahi (2003) on the ARMA representation of the IV. 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 details on the proofs that are omitted in the text. In the text, we mainly focus on the

---

[7]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).

[8]Jungbacker and Koopman (2006, p.191) assume that a return (rather than a log-price) is contaminated by an i.i.d. MN for simplifying the estimation.

[9]Another difference between our and Jungbacker and Koopman's (2006) methods is that their estimation method depends on the distributional assumption while ours is not.

one-factor SR-SARV model for ease of exposition. The results for the two factor model are presented in Appendix B.

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

In this section, we formally define the SR-SARV model, IV, RV, and MN component. We also show that the MN component follows a MA(1) process.

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 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 $W(t)$ is a standard Brownian motion and $f(t)$ is a state-variable process (possibly bivariate) independent of $W(t)$.[10] The functions $P_1(\cdot)$ and $P_2(\cdot)$ are defined so that:

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

where λ_1 and λ_2 are positive real numbers.

Assumption 1 implies that 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 describe 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, 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 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 research objective. For example, if the researcher is interested in changes in variances of daily (weekly) returns, t is interpreted as a day (week). It follows from Assumption 1, that

$$r_t|IV_t \sim N(0, IV_t), \quad \text{and hence} \quad \frac{r_t}{\sqrt{IV_t}} \sim N(0, 1),$$

where $r_t \equiv p(t) - p(t-1)$.

---

[10] As a consequence of independence between $W(t)$ and $f(t)$, there is no leverage effect in this price process.

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. Here, and hereafter, the superscript "(m)" implies that its value depends on m (note that the sampling frequency increases as m increases). 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).

Meddahi (2003) shows that, for the two-factor SR-SARV model, the variance and autocovariances of IV_t are expressed in terms of the SV model parameters as:

$$\text{var}[IV_t] = \frac{2\omega_1^2(\kappa_1 - \log \kappa_1 - 1)}{(\log \kappa_1)^2} + \frac{2\omega_2^2(\kappa_2 - \log \kappa_2 - 1)}{(\log \kappa_2)^2}, \tag{3}$$

and

$$\text{cov}[IV_t, IV_{t-1}] = \frac{\omega_1^2(1 - \kappa_1)^2}{(\log \kappa_1)^2} + \frac{\omega_2^2(1 - \kappa_2)^2}{(\log \kappa_2)^2}. \tag{4}$$

The discretization error, $d_t^{(m)}$, is defined as $d_t^{(m)} \equiv RV_t^{(m)} - IV_t$. Let $\sigma_d^{2(m)} \equiv \text{var}[d_t^{(m)}]$. For $m \geq 1$, we have:

$$\sigma_d^{2(m)} = \frac{2\sigma^4}{m} + \frac{4\omega_1^2 m}{(\log \kappa_1)^2} \left(\kappa_1^{\frac{1}{m}} - \log \kappa_1^{\frac{1}{m}} - 1 \right) + \frac{4\omega_2^2 m}{(\log \kappa_2)^2} \left(\kappa_2^{\frac{1}{m}} - \log \kappa_2^{\frac{1}{m}} - 1 \right). \tag{5}$$

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

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) $E[\varepsilon(t)] = 0$, $\text{var}[\varepsilon(t)] = \sigma_\varepsilon^2$, $\text{var}[\varepsilon^2(t)] = \omega_\varepsilon^2 < \infty$ for all t,
$\text{cov}[\varepsilon^k(t), \varepsilon(s)] = \text{cov}[\varepsilon^2(t), \varepsilon^2(s)] = 0$ for $k = 1, 2, 3$ and $t \neq s$, and
$E[\varepsilon^2(t)\varepsilon(s)\varepsilon(u)] = E[\varepsilon(t)\varepsilon(s)\varepsilon(u)\varepsilon(v)] = 0$ for $t \neq s \neq u \neq v$.

(b) $\varepsilon(t)$ is independent of $W(s)$ and $f(s)$ (hence $p(s)$ too) for all s and t.[11]

We do not assume any specific distribution for $\varepsilon(t)$. These assumptions on MN seem reasonable for exchange rate return data that we analyze in the empirical section, where the return data clearly indicates its MA(1) structure, which is deduced from Assumption 2 as stated in (7).[12]

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

$$r_t^{*(m)} \equiv p^*(t) - p^* \left(t - \frac{1}{m} \right) = r_t^{(m)} + e_t^{(m)}, \tag{6}$$

---

[11] This assumption can be relaxed to some extent, albeit leading to rather complicated conditions difficult to interpret, instead.

[12] See Ubukata and Oya (2009) on estimating and testing for serial dependence in MN.

where $e_t^{(m)} \equiv \varepsilon(t) - \varepsilon(t - \frac{1}{m})$. Under Assumptions 1 and 2, it is easy to show that

$$E\left[r_t^{*(m)}\right] = E\left[e_t^{(m)}\right] = 0, \quad \mathrm{var}\left[r_t^{*(m)}\right] = \frac{\sigma^2}{m} + 2\sigma_\varepsilon^2, \quad \mathrm{var}\left[e_t^{(m)}\right] = 2\sigma_\varepsilon^2,$$

and

$$\mathrm{cov}\left[r_t^{*(m)}, r_{t-\frac{i}{m}}^{*(m)}\right] = \mathrm{cov}\left[e_t^{(m)}, e_{t-\frac{i}{m}}^{(m)}\right] = \begin{cases} -\sigma_\varepsilon^2 & i = 1, \\ 0 & i \geq 2. \end{cases} \tag{7}$$

Therefore, we have

$$\mathrm{corr}\left[r_t^{*(m)}, r_{t-\frac{i}{m}}^{*(m)}\right] = \begin{cases} -\frac{m\sigma_\varepsilon^2}{\sigma^2 + 2m\sigma_\varepsilon^2}, & i = 1, \\ 0 & i \geq 2, \end{cases} \quad \text{and} \quad \mathrm{corr}\left[e_t^{(m)}, e_{t-\frac{i}{m}}^{(m)}\right] = \begin{cases} -0.5, & i = 1, \\ 0 & i \geq 2. \end{cases} \tag{8}$$

Note that the result in (7) 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 $r_t^{*(m)}$ increases 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:

$$RV_t^{*(m)} = \sum_{i=1}^m \left(r_{t-1+\frac{i}{m}}^{(m)} + e_{t-1+\frac{i}{m}}^{(m)}\right)^2 = RV_t^{(m)} + u_t^{(m)}, \tag{9}$$

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. We propose a way of estimating the IV as well as the MN component in the next section. The following lemma is proved in Appendix A:

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

$$E\left[u_t^{(m)}\right] = 2m\sigma_\varepsilon^2 \quad and \quad cov\left[u_t^{(m)}, u_s^{(m)}\right] = \begin{cases} 8\sigma_\varepsilon^2\sigma^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4 & t = s, \\ \omega_\varepsilon^2 & t = s \pm 1, \\ 0 & otherwise. \end{cases} \tag{10}$$

Thus, $u_t^{(m)}$ has the autocovariance structure of a $MA(1)$ process. Without loss of generality, we can express the $MA(1)$ process as:

$$u_t^{(m)} = c_u^{(m)} + \xi_t^{(m)} + \theta_u^{(m)}\xi_{t-1}^{(m)}, \quad \xi_t^{(m)} \sim WN(0, \sigma_\xi^{2(m)}), \tag{11}$$

where $WN(0, a)$ denotes a white noise process with variance a.[13] The mean and autocovariances of $u_t^{(m)}$, in terms of $c_u^{(m)}$, $\theta_u^{(m)}$ and $\sigma_\xi^{2(m)}$, are:

$$E\left[u_t^{(m)}\right] = c_u^{(m)} \quad and \quad cov\left[u_t^{(m)}, u_s^{(m)}\right] = \begin{cases} (1 + \theta_u^{2(m)})\sigma_\xi^{2(m)} & t = s, \\ \theta_u^{(m)}\sigma_\xi^{2(m)} & t = s \pm 1, \\ 0 & otherwise. \end{cases} \tag{12}$$

---

[13] The proof of Lemma 1 in Appendix A implies that $\xi_t^{(m)}$ is conditionally heteroskedastic for a fixed m, but this does not matter for our results because the assumption that the variance process is stationary implies that the unconditional variance of $\xi_t^{(m)}$ is constant for a fixed m, which is all what we need for our results in this paper.

In the next section, we utilize these two different expressions of the moments of $u_t^{(m)}$ (i.e., (10) and (12)) 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 in detail. First, we briefly summarize the result of Meddahi (2003) for the one-factor SR-SARV model in Section 3.1. Then, we give a state space form of the NCRV in Section 3.2. In Sections 3.3 and 3.4, we show that the state space form parameters are not identifiable, however, they can be expressed as functions of identifiable parameters. In Section 3.5, we illustrate how to estimate these identifiable parameters.

In what follows, we focus on the case of $\omega_2 = 0$, i.e., the one-factor SR-SARV model for ease of exposition. Corresponding results for the two-factor SV model can be derived in a similar manner although relevant calculations are more involved. The results on the two-factor SV model are summarized in Appendix B.

3.1   ARMA representation for the one-factor SR-SARV model

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:

$$IV_t = c_{IV} + \kappa_1 IV_{t-1} + \eta_t + \theta_1 \eta_{t-1}, \tag{13}$$

where κ_1 is defined as in the statement below (2), η_t is a white noise process with $\mathrm{var}(\eta_t) = \sigma_\eta^2$.[14] 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:

$$c_{IV} = (1 - \kappa_1)\sigma^2, \quad \theta_1 = \frac{1 - \sqrt{1 - 4\rho^2}}{2\rho} \tag{14}$$

and

$$\sigma_\eta^2 = \frac{(1 + \kappa_1^2)\mathrm{var}[IV_t] - 2\kappa_1 \mathrm{cov}[IV_t, IV_{t-1}]}{1 + \theta_1^2}, \tag{15}$$

where

$$\rho \equiv \frac{-\kappa_1 + \mathrm{corr}[IV_t, IV_{t-1}]}{1 + \kappa_1^2 - 2\kappa_1 \mathrm{corr}[IV_t, IV_{t-1}]}. \tag{16}$$

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 (13). See Meddahi (2003) for more details on these results.

3.2   State space form of the NCRV

Substituting $RV_t^{(m)} = IV_t + d_t^{(m)}$ into (9), we have:

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

The following lemma is proved in Appendix A:

Lemma 2 *Under Assumptions 1 and 2,*

$$\mathrm{cov}[d_t^{(m)}, \ \eta_t] = \mathrm{cov}[d_t^{(m)}, \ \xi_t^{(m)}] = \mathrm{cov}[\eta_t, \ \xi_t^{(m)}] = 0, \tag{18}$$

where $d_t^{(m)} = RV_t^{(m)} - IV_t$, and $\xi_t^{(m)}$ and η_t are given as in (11) and (13), respectively.

---

[14]Meddahi (2003) shows that η_t is conditionally heteroskedastic, but because of the stationarity assumption of the variance process, it is unconditionally homoskedastic, which is all what we need for our results in this paper.

Let η_t and $\xi_t^{(m)}$ be denoted by state variables α_t and $\beta_t^{(m)}$, respectively. From (11), (13), (17), and (18), we have the following state space form of $RV_t^{*(m)}$:

Observation equation

$$RV_t^{*(m)} = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t^{(m)} \end{bmatrix} + d_t^{(m)}, \tag{19a}$$

State equation

$$\begin{bmatrix} IV_t \\ u_t^{(m)} \\ \alpha_t \\ \beta_t^{(m)} \end{bmatrix} = \begin{bmatrix} c_{IV} \\ c_u^{(m)} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \kappa_1 & 0 & \theta_1 & 0 \\ 0 & 0 & 0 & \theta_u^{(m)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_{t-1} \\ u_{t-1}^{(m)} \\ \alpha_{t-1} \\ \beta_{t-1}^{(m)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t^{(m)} \end{bmatrix}, \tag{19b}$$

where

$$\begin{bmatrix} d_t^{(m)} \\ \eta_t \\ \xi_t^{(m)} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_d^{2(m)} & 0 & 0 \\ 0 & \sigma_\eta^2 & 0 \\ 0 & 0 & \sigma_\xi^{2(m)} \end{bmatrix} \right). \tag{19c}$$

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.[15] 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 innovation assumption. We show, however, that this approach is not applicable for the state space form given in $(19a) - (19c)$.

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.[16]

3.3   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:

$$(1 - \kappa_1 L) RV_t^{*(m)} = c_{RV}^{(m)} + (1 + \delta_1^{(m)} L + \delta_2^{(m)} L^2) \tau_t^{(m)}, \quad \tau_t^{(m)} \sim WN(0, \sigma_\tau^{2(m)}). \tag{20}$$

Note that the AR coefficient κ_1 is the same as that of the IV_t in (13). 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.

---

[15]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 (Anderson and Moore, 1979; Hamilton, 1994, Chapter 13). See Durbin and Koopman (2001) for more details on the Kalman filter.

[16]The analysis in the subsequent subsections also show 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.

From (11), (13) and (17), we have

$$
\begin{aligned}
(1 - \kappa_1 L) RV_t^{*(m)} =\ & (1 - \kappa_1 L) IV_t + (1 - \kappa_1 L) d_t^{(m)} + (1 - \kappa_1 L) u_t^{(m)} \\
=\ & c_{IV} + \eta_t + \theta_1 \eta_{t-1} + d_t^{(m)} - \kappa_1 d_{t-1}^{(m)} + \xi_t^{(m)} \\
& + (1 - \kappa_1) c_u^{(m)} + (\theta_u^{(m)} - \kappa_1) \xi_{t-1}^{(m)} - \kappa_1 \theta_u^{(m)} \xi_{t-2}^{(m)}.
\end{aligned}
\tag{21}
$$

The two expressions on the right-hand sides in (20) and (21) are of the same process and hence their means and autocovariances must be identical. The autocovariances of the MA process in (20) are given as

$$
\gamma_0^{(m)} = (1 + \delta_1^{(m)2} + \delta_2^{(m)2}) \sigma_\tau^{2(m)},
$$

$$
\gamma_1^{(m)} = (\delta_1^{(m)} + \delta_1^{(m)} \delta_2^{(m)}) \sigma_\tau^{2(m)},
\tag{22}
$$

$$
\gamma_2^{(m)} = \delta_2^{(m)} \sigma_\tau^{2(m)},
$$

and $\gamma_j = 0$ for $j \geq 3$. It is easy to show that the autocovariances of the MA process in (21) are

$$
\gamma_0^{(m)} = (1 + \theta_1^2) \sigma_\eta^2 + (1 + \kappa_1^2) \sigma_d^{2(m)} + (1 + \theta_u^{(m)2} - 2\theta_u^{(m)} \kappa_1 + \kappa_1^2 + \kappa_1^2 \theta_u^{(m)2}) \sigma_\xi^{2(m)},
\tag{23a}
$$

$$
\gamma_1^{(m)} = \theta_1 \sigma_\eta^2 - \kappa_1 \sigma_d^{2(m)} + (\theta_u^{(m)} - \kappa_1 - \kappa_1 \theta_u^{(m)2} + \kappa_1^2 \theta_u^{(m)}) \sigma_\xi^{2(m)},
\tag{23b}
$$

$$
\gamma_2^{(m)} = -\kappa_1 \theta_u^{(m)} \sigma_\xi^{2(m)},
\tag{23c}
$$

and $\gamma_j = 0$ for $j \geq 3$. By equating the means of the MA processes in (20) and (21), we have

$$
c_{RV}^{(m)} = c_{IV} + (1 - \kappa_1) c_u^{(m)}.
\tag{23d}
$$

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 (23a)$\sim$(23d) 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.

3.4   Identifiable parameters

From (10) and (12), we obtain the following equations:

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

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

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

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

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

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

10

From (3), (4), (15), and (25), 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.[17] To make the functional relationship explicit, we may denote them as:

$$c_{IV}(\kappa_1, \sigma^2), \quad \theta_1(\kappa_1), \quad \sigma_\eta^2(\kappa_1, \omega_1^2), \quad c_u^{(m)}(\sigma_\varepsilon^2), \quad \theta_u^{(m)}(\sigma^2, \sigma_\varepsilon^2, \omega_\varepsilon^2),$$

$$\sigma_d^{2(m)}(\kappa_1, \sigma^2, \omega_1^2), \quad \text{and} \quad \sigma_\xi^{2(m)}(\sigma^2, \sigma_\varepsilon^2, \omega_\varepsilon^2). \tag{26}$$

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 (26) into Equations (23a)∼(23d), 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, $\gamma_j^{(m)}$, $j = 0, ..., 2$. The following proposition assures the uniqueness of the identification:

Proposition 1 *Given $c_{RV}^{(m)}$, κ_1, $\gamma_j^{(m)}$, $j = 0, ..., 2$ and (26), under the condition $\sigma_\varepsilon^2 > 0$, Equations (23a)∼(23d) are uniquely solved for σ^2, ω_1^2, σ_ε^2 and ω_ε^2 as:*

$$\omega_\varepsilon^2 = -\frac{\gamma_2^{(m)}}{\kappa_1}, \quad \omega_1^2 = \frac{(\log \kappa_1)^2 [\kappa_1 \gamma_0^{(m)} + (1 + \kappa_1^2)\gamma_1^{(m)} + \frac{1+\kappa_1^4}{\kappa_1}\gamma_2^{(m)}]}{(1 - \kappa_1)^3(1 + \kappa_1)}, \tag{27a}$$

$$\sigma_\varepsilon^2 = \sqrt{\frac{c_{RV}^{(m)2}}{2m^2(1 - \kappa_1)^2} - \frac{(2m - 1)\gamma_2^{(m)}}{2m\kappa_1} - \frac{\gamma_0^{(m)} - 2D\omega_1^2 - 2\gamma_2^{(m)}}{4m(1 + \kappa_1^2)}}, \tag{27b}$$

and

$$\sigma^2 = \frac{c_{RV}^{(m)}}{1 - \kappa_1} - 2m\sigma_\varepsilon^2, \tag{27c}$$

where

$$D = B + m(1 + \kappa_1^2)C, \quad B \equiv \frac{\kappa_1^2 - 1 - (1 + \kappa_1^2)\log \kappa_1}{(\log \kappa_1)^2} \quad \text{and} \quad C \equiv \frac{2\left(\kappa_1^{\frac{1}{m}} - 1 - \log \kappa_1^{\frac{1}{m}}\right)}{(\log \kappa_1)^2}. \tag{27d}$$

Proof See Appendix A.

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 (20). 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. Corresponding results for the two-factor model are given in Proposition 2 in Appendix B.

3.5   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 (27) 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

---

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

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

Step 2 Estimate the unrestricted ARMA$(1, 2)$ model in (20) by QML estimation assuming Gaussian innovations.[18]

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 (22).

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, calculate ω_ε^2, σ_ε^2, ω_1^2 and σ^2 applying the results in (27a) $-$ (27d).

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 (3), (4), (15), and (25).

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 (19a)–(19c) 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 (26).

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

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

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]

---

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

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.[19] 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. Figure 2 plots these NRCV 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 (23d) and (24a)). 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 and Figure 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 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. This result implies that the fourth moment increases in proportion to the

---

[19]These returns are calculated from price data combined after the removals of the prices of inactive trading days according to the above criteria. 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 not significantly different.

sampling interval although the variances of MNs are constant across different sampling frequencies. 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 in (19) (and in (50) for the two-factor model) 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.

[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 3 and 4, 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"; 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 3 and 4 and Table 6 around here]

To see the similarities among the smoothed IV series with different m's, we calculate the mean absolute differences (MADs) and correlations among these smoothed IV series. Table 6 presents them. For the one-factor model, the MAD takes the largest value of 0.0913 with the combination of $m = 1440$ and 96, and takes the smallest value of 0.0165 with $m = 144$ and 96. On the other hand, the correlation is 0.9843 at the largest with $m = 144$ and 96, and is 0.8700 at the smallest with $m = 1440$ and 96. It is observed that the closer the values of m's are, the smaller are the differences between corresponding smoothed IV series. In any case, the differences among the smoothed IV series are very small. The results for the two-factor model are qualitatively very similar and thus the same comment applies although the corresponding values are slightly different. All these results imply that our method works well regardless of m.

14

[Table 7 around here]

One informal way of assessing the quality of IV estimates is to see whether the daily return normalized by the square root of these IV estimates are enough close to the standard normal as the theory suggests. In this regards, we calculate the Jarque-Bera normality test (hereafter, JB test) for the return normalized by the square root of the NCRV and smoothed IV, which are the estimates of $\sqrt{IV_t}$, or the *integrated volatility*. To see the point of this argument, let r_t and z_t denote the daily return and the return normalized by the integrated volatility, i.e., $z_t \equiv r_t/\sqrt{IV_t}$, respectively. For an IV estimator, $\widehat{IV}_t$, such as the NCRV or smoothed IV, the daily return normalized by the square root of the IV estimate, $\widehat{z}_t$, is written as $\widehat{z}_t = z_t \exp(v_t)$, where $v_t = -0.5(\log \widehat{IV}_t - \log IV_t)$. Thus, strictly speaking, even if z_t follows the standard normal distribution, $\widehat{z}_t$ does not, unless $v_t = 0$. However, it is likely that if z_t follows the standard normal and $\widehat{IV}_t$ is a good estimator so that v_t is enough close to zero, the JB test does not reject the null of normality.[20] In other words, if the JB test rejects the null hypothesis, it implies either (or both) z_t does not follow a normal distribution or (and) $\widehat{IV}_t$ is not a good estimator of IV_t. Hence, though not a formal procedure, it is still helpful for roughly checking the quality of IV estimates to see whether the JB test rejects the null of normality for the return normalized by the square root of the IV estimates.

Table 7 shows the results of the JB test along with the sample mean, sample standard deviation, sample skewness, and sample kurtosis of the normalized return with different m. For the daily return without normalization, the normality is clearly rejected. In particular, the sample standard deviation and sample kurtosis are 0.5802 and 4.3475, respectively, which are far from the theoretical values of 1.0 and 3.0 for the standard normal distribution. For the daily return normalized by the square root of NCRV, the normality is rejected for $m \leq 288$. The sample kurtosis of the normalized daily return with $m \leq 288$ is platykurtic, or less than 3.0. In the case of $m = 1440$, the sample kurtosis is 3.0101 and the JB test does not reject the normality, however, the sample standard deviation is 0.7586, which is far below from 1.0. In sum, the series of the square root of the NCRV do not seem to work well as integrated volatility estimates that are consistent with the theory. On the other hand, the results for the square root of the smoothed IV are closely consistent with the theory, in particular, for the two-factor model. For the one-factor model, the sample kurtoses of the normalized returns tend to be leptokurtic, namely, greater than 3, but the sample standard deviations are reasonably close to 1 (ranging from 0.9445 to 1.1065), and the normality is not rejected for the cases of $m = 288$ and $m = 144$. For the two-factor model, the results are amazingly good. The sample standard deviations and kurtosis are reasonably close to 1 and 3, respectively, for any m, and the normality is not rejected for any m.

4.3 Comparing forecasting performances

In this subsection, we compare forecasting performances of our and the Barndorff-Neilsen, Shephard, and Meddahi (hereafter BSM) methods. We call the latter method so because we apply the state space method of Barndorff-Nielsen and Shephard (2002) combined with the ARMA representation result of Meddahi (2003).[21] Following Andersen, Bollerslev, Diebold, and Labys (2003), we run the so called Mincer-Zarnowitz style regressions. We also calculate the conventional mean absolute errors (MAEs). The forecasting performances are evaluated by the values of R^2 and MAE.

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, where unknown model parameters are estimated from $RV_t^{*(m)}$. Because the IV is not 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.[22]

---

[20]Note that the JB test is a test not for the standard normal but for normality. Thus, if v_t is close to a constant (not necessarily zero), then the JB test will not reject the null in this case.

[21]Using the result in Meddahi (2003) makes it easier to implement the state space method in Barndorff-Nielsen and Shephard (2002).

[22]Using a RV as a proxy for the IV leads to a downward bias in the R^2 of the Mincer-Zarnowitz regression. Andersen,

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

(i): $RV_{t+1}^{*(48)} = a_0 + a_1 IV_{t+1|t}^{(NR)(m)} + \nu_t^{(a)}$,

(ii): $RV_{t+1}^{*(48)} = b_0 + b_1 IV_{t+1|t}^{(BSM)(m)} + \nu_t^{(b)}$,

(iii): $RV_{t+1}^{*(48)} = c_0 + c_1 IV_{t+1|t}^{(NR)(m)} + c_2 IV_{t+1|t}^{(BSM)(m)} + \nu_t^{(c)}$

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. For in-sample forecasting, first, we estimate unknown parameters with all available samples of $RV_t^{*(m)}$ ($t = 1, ..., 1809$), and calculate $IV_{t+1|t}^{(X)(m)}$, $X = \{NR, BSM\}$, for $t = 1, ..., 1808$, with the estimated parameter values. Then, we run the regressions (i) $\sim$ (iii) for $t = 1, ..., 1808$, and calculate the associated R^2 and MAE values. For out-of-sample forecasting, we estimate unknown parameters using only most recent 1200 samples previous to the time point for which we wish to forecast, and calculate $IV_{t+1|t}^{(X)(m)}$, $X = \{NR, BSM\}$, for $t = 1200, ...1808$ with the estimated values. Then, again we run the three regressions for $t = 1200, ..., 1808$, and calculate the R^2 and MAE values. We also calculate the mean absolute differences (MADs) and correlations between the predicated IV series by our and the BSM methods to see the similarities among the forecasts by the two methods.

[Table 8 around here]

Table 8 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), R^2 increases 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 increases 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), similar comments as the case of the regression (i) apply, where R^2 values are close to those of the regression (i). In terms of the MAE 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 8 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 MAE criterion, again, our method has smaller MAE values than the BSM method for any m. This can be confirmed visually in Figures 5 (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 (a), 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 98), our method still works better than the BSM method, however, the differences in R^2 and MAE values between the two methods are small. In fact, the two methods provide very similar forecasts, as implied by the values of the MAD and correlation between the two forecasts reported at the bottom of Table 8 (for example, the correlation

---

Bollerslev and Meddahi (2005) propose a method for correcting the bias, however, because their method multiplies R^2 by the same correction term regardless of the forecasting method employed, it does not alter the order of R^2 values, and hence is not necessary for our purpose of comparison of different forecasting methods.

between the forecasts by the two methods is 0.9383 when $m = 98$). This can also be checked in Figures 5 (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 5 and 6 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 6 (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 MAE 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 9 around here]

Table 9 reports the results for the out-of-sample forecasting. Figures 7 and 8 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 7 and 8 around here]

5   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 state space form parameters. The proposed method was applied to yen/dollar exchange rate data, where we find that most of the variations in the NCRV are due to the MN component.

We used only one NCRV series for the estimation of unknown model parameters. It would be possible to combine several NCRV series to obtain more accurate estimates. This is a subject for the future research. It is also important to relax the assumption of no leverage effect (see footnote 10) in order to apply our method to stock return data.

Appendix A: Proofs

Hereafter, we suppress the superscript "(m)", and let ε_t denote $\varepsilon(t)$ for notational simplicity.

Proof of (25)
From (24c), we have $\sigma_\xi^2 = \omega_\varepsilon^2/\theta_u$. Substituting this into (24b), we have:

$$(1+\theta_u^2)\frac{\omega_\varepsilon^2}{\theta_u} = 8\sigma^2\sigma_\varepsilon^2 + 2(2m-1)\omega_\varepsilon^2 + 4m\sigma_\varepsilon^4.$$

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

$$\theta_u^2 - 2\left[4\frac{\sigma^2\sigma_\varepsilon^2}{\omega_\varepsilon^2} + 2m - 1 + 2m\frac{\sigma_\varepsilon^4}{\omega_\varepsilon^2}\right]\theta_u + 1 = 0.$$

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

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

Because $A > 1$ for $m \geq 1$, we have $A - \sqrt{A^2-1} < 1 < A + \sqrt{A^2-1}$. Assuming that θ_u satisfies the invertibility condition, we obtain θ_u in (25).

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:

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

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 $t = s$, we have:

$$
\begin{aligned}
\mathrm{cov}\left[r_t e_t, r_t e_t\right] =\ & E\left[r_t^2 e_t^2\right] - (E\left[r_t e_t\right])^2 \\
=\ & E\left[e_t^2\right] E\left[r_t^2\right] - (E\left[r_t\right])^2 (E\left[e_t\right])^2 \\
=\ & 2\sigma_\varepsilon^2 E[(\int_{t-1/m}^t \sigma(s)dW(s))^2] \\
=\ & 2\sigma_\varepsilon^2 E[\int_{t-1/m}^t \sigma^2(s)ds] \\
=\ & \frac{2\sigma_\varepsilon^2 \sigma^2}{m}.
\end{aligned}
\tag{28}
$$

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

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

Next, for $\mathrm{cov}\left[e_t^2, e_s^2\right]$, when $t = s$, we have from Assumption 2(a):

$$
\begin{aligned}
\mathrm{cov}[e_t^2, e_t^2] =\ & E[e_t^4] - (E[e_t^2])^2 \\
=\ & E\left[\varepsilon_t^4 - 4\varepsilon_t^3 \varepsilon_{t-\frac{1}{m}} + 6\varepsilon_t^2 \varepsilon_{t-\frac{1}{m}}^2 - 4\varepsilon_t \varepsilon_{t-\frac{1}{m}}^3 + \varepsilon_{t-\frac{1}{m}}^4\right] - 4\sigma_\varepsilon^4 \\
=\ & 2E[\varepsilon_t^4] + 2\sigma_\varepsilon^4 \\
=\ & 2\omega_\varepsilon^2 + 4\sigma_\varepsilon^4.
\end{aligned}
\tag{29}
$$

When $t = s \pm \frac{1}{m}$, we have from Assumption 2(a):

$$
\begin{aligned}
\text{cov}\left[e_s^2, e_{s-\frac{1}{m}}^2\right] &= \text{cov}\left[e_{s+\frac{1}{m}}^2, e_s^2\right] \\
&= \text{cov}\left[\varepsilon_{s+\frac{1}{m}}^2 - 2\varepsilon_{s+\frac{1}{m}}\varepsilon_s + \varepsilon_s^2, \varepsilon_s^2 - 2\varepsilon_s\varepsilon_{s-\frac{1}{m}} + \varepsilon_{s-\frac{1}{m}}^2\right] \\
&= \text{var}[\varepsilon_s^2] \\
&= \omega_\varepsilon^2.
\end{aligned}
\tag{30}
$$

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

$$
\begin{aligned}
\text{cov}[r_t e_t, e_s^2] &= E[r_t e_t e_s^2] - E[r_t e_t] E[e_s^2] \\
&= E[r_t] E[e_t e_s^2] - E[r_t] E[e_t] E[e_s^2] \\
&= 0.
\end{aligned}
\tag{31}
$$

From (28) $\sim$ (31), we have:

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

and

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

It is easy to check that $\text{cov}[u_t, u_{t\pm i}] = 0$ for $i \geq 2$, and hence we have (10).

Proof of Lemma 2

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

$$
\text{cov}\left[\int_t^{t+\Delta} \sigma^2(u)\mathrm{d}u, \ e_s \times \int_s^{s+\Delta'} \sigma(u)\mathrm{d}W(u)\right]
$$

$$
= E\left[e_s\right] E\left[\int_t^{t+\Delta} \sigma^2(u)\mathrm{d}u \times \int_s^{s+\Delta'} \sigma(u)\mathrm{d}W(u)\right] - E\left[e_s\right] E\left[\int_t^{t+\Delta} \sigma^2(u)\mathrm{d}u\right] E\left[\int_s^{s+\Delta'} \sigma(u)\mathrm{d}u\right]
$$

$$
= 0,
$$

$$
\text{cov}\left[\int_t^{t+\Delta} \sigma^2(u)\mathrm{d}u, \ e_s^2\right] = 0, \quad \text{cov}\left[r_t^2, r_s e_s\right] = 0, \quad \text{and} \quad \text{cov}\left[r_t^2, e_s^2\right] = E[r_t^2 e_s^2] - E[r_t^2]E[e_s^2] = 0.
$$

Hence, we have

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

$$(32)$$

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

and thus,

$$
\operatorname{cov}\left[d_t, u_s\right] = \operatorname{cov}\left[RV_t - IV_t, u_s\right] = \operatorname{cov}\left[RV_t, u_s\right] - \operatorname{cov}\left[IV_t, u_s\right] = 0, \tag{33}
$$

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

$$
\operatorname{cov}\left[IV_t, d_s\right] = 0, \tag{34}
$$

for all t and s. From (32) and (34), 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

$$
\operatorname{cov}\left[\eta_t, u_s\right] = 0 \quad \text{and} \quad \operatorname{cov}\left[\eta_t, d_s\right] = 0, \tag{35}
$$

for all t and s. From (33) and (35), and noting that $\xi_t = (1 - \theta_u L)^{-1} u_t$,[23] we have

$$
\operatorname{cov}\left[\eta_t, \xi_s\right] = 0 \quad \text{and} \quad \operatorname{cov}\left[d_t, \xi_s\right] = 0,
$$

for all t and s, which completes the proof of Lemma 2.

Proof of Proposition 1

From (23c) and (24c), we have $\omega_\varepsilon^2 = -\frac{\gamma_2}{\kappa_1}$, which is the first result in (27a). From (3), (4), (5), and (15), we have

$$
\sigma_\eta^2 = \frac{2B\omega_1^2}{1 + \theta_1^2} \quad \text{and} \quad \sigma_d^2 = \frac{2\sigma^4}{m} + 2mC\omega_1^2, \tag{36}
$$

---

[23]θ_u is invertible. See the statement below (25).

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

$$(1 + \theta_u^2 - 2\theta_u \kappa_1 + \kappa_1^2 + \kappa_1^2 \theta_u^2)\sigma_\xi^2 = \left[\left(\tfrac{1}{\theta_u} + \theta_u\right)(1 + \kappa_1^2) - 2\kappa_1\right]\omega_\varepsilon^2, \tag{37}$$

and

$$(\theta_u - \kappa_1 - \kappa_1 \theta_u^2 + \kappa_1^2 \theta_u)\sigma_\xi^2 = \left[1 + \kappa_1^2 - \left(\tfrac{1}{\theta_u} + \theta_u\right)\kappa_1\right]\omega_\varepsilon^2. \tag{38}$$

Substituting (36), (37) and (38) into (23a) and (23b), we have:

$$\gamma_0 = 2D\omega_1^2 + 2\frac{1 + \kappa_1^2}{m}\sigma^4 + \left[\left(\frac{1}{\theta_u} + \theta_u\right)(1 + \kappa_1^2) - 2\kappa_1\right]\omega_\varepsilon^2, \tag{39a}$$

and

$$\gamma_1 = 2E\omega_1^2 - 2\frac{\kappa_1}{m}\sigma^4 - \left[\left(\frac{1}{\theta_u} + \theta_u\right)\kappa_1 - (1 + \kappa_1^2)\right]\omega_\varepsilon^2, \tag{39b}$$

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 (39), we have:

$$\begin{aligned}
\kappa_1 \gamma_0 + (1 + \kappa_1^2)\gamma_1 &= 2\left[\kappa_1 D + (1 + \kappa_1^2)E\right]\omega_1^2 + \left[(1 + \kappa_1^2)^2 - 2\kappa_1^2\right]\omega_\varepsilon^2, \\
&= 2\left[\kappa_1 + (1 + \kappa_1^2)\rho\right]B\omega_1^2 + (1 + \kappa_1^4)\omega_\varepsilon^2, \\
&= \frac{(1 - \kappa_1)^3(1 + \kappa_1)}{(\log \kappa_1)^2}\omega_1^2 + (1 + \kappa_1^4)\omega_\varepsilon^2,
\end{aligned} \tag{40}$$

where, to obtain the third equality, we use the alternative expression of ρ in (16). From (40), 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 (27a). Next, note that from (25), we have:

$$\begin{aligned}
\frac{1}{\theta_u} + \theta_u &= \frac{1 + \theta_u^2}{\theta_u} \\
&= \frac{1 + (A - \sqrt{A^2 - 1})^2}{A - \sqrt{A^2 - 1}} \\
&= 2A.
\end{aligned} \tag{41}$$

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

$$c_{RV} = (1 - \kappa_1)\left(\sigma^2 + 2m\sigma_\varepsilon^2\right), \quad \text{or} \quad \sigma_\varepsilon^2 = \frac{c_{RV} - (1 - \kappa_1)\sigma^2}{2(1 - \kappa_1)m}. \tag{42}$$

Substituting σ_ε^2 in (42) into A in (25), we have:

$$\begin{aligned}
2A &= 2\left[\frac{4\sigma^2}{\omega_\varepsilon^2}\left(\frac{c_{RV} - (1 - \kappa_1)\sigma^2}{2(1 - \kappa_1)m}\right) + 2m - 1 + \frac{2m}{\omega_\varepsilon^2}\left(\frac{c_{RV} - (1 - \kappa_1)\sigma^2}{2(1 - \kappa_1)m}\right)^2\right] \\
&= 2\left[\frac{2c_{RV}\sigma^2}{(1 - \kappa_1)m\omega_\varepsilon^2} - \frac{2\sigma^4}{m\omega_\varepsilon^2} + 2m - 1 + \frac{c_{RV}^2 - 2(1 - \kappa_1)c_{RV}\sigma^2 + (1 - \kappa_1)^2\sigma^4}{2(1 - \kappa_1)^2 m\omega_\varepsilon^2}\right] \\
&= \frac{4c_{RV}\sigma^2}{(1 - \kappa_1)m\omega_\varepsilon^2} - \frac{4\sigma^4}{m\omega_\varepsilon^2} + 2(2m - 1) + \frac{c_{RV}^2}{(1 - \kappa_1)^2 m\omega_\varepsilon^2} - \frac{2c_{RV}\sigma^2}{(1 - \kappa_1)m\omega_\varepsilon^2} + \frac{\sigma^4}{m\omega_\varepsilon^2} \\
&= \frac{2c_{RV}\sigma^2}{(1 - \kappa_1)m\omega_\varepsilon^2} - \frac{3\sigma^4}{m\omega_\varepsilon^2} + 2(2m - 1) + \frac{c_{RV}^2}{(1 - \kappa_1)^2 m\omega_\varepsilon^2}.
\end{aligned} \tag{43}$$

From (39a), (41) and (43), we have:

$$\gamma_0 = 2D\omega_1^2 - \frac{(1 + \kappa_1^2)}{m}\sigma^4 + \frac{2(1 + \kappa_1^2)c_{RV}}{(1 - \kappa_1)m}\sigma^2 + 2(2m - 1)(1 + \kappa_1^2)\omega_\varepsilon^2 + \frac{(1 + \kappa_1^2)c_{RV}^2}{(1 - \kappa_1)^2 m} - 2\kappa_1 \omega_\varepsilon^2. \tag{44}$$

21

Multiplying both sides in (44) 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:

$$\sigma^2 = \frac{c_{RV}}{1 - \kappa_1} \pm \sqrt{\frac{2c_{RV}^2}{(1 - \kappa_1)^2} + 2m(2m - 1)\omega_\varepsilon^2 - \frac{m\left(\gamma_0 - 2D\omega_1^2 + 2\kappa_1\omega_\varepsilon^2\right)}{(1 + \kappa_1^2)}}. \tag{45}$$

From $\sigma_\varepsilon^2 > 0$, $\kappa_1 < 1$ and (42), we must have $\frac{c_{RV}}{1 - \kappa_1} > \sigma^2$. Hence, the sign of the second term in (45) is negative. From (42) and (45), we have:

$$\sigma_\varepsilon^2 = \frac{1}{2m}\sqrt{\frac{2c_{RV}^2}{(1 - \kappa_1)^2} + 2m(2m - 1)\omega_\varepsilon^2 - \frac{m\left(\gamma_0 - 2D\omega_1^2 + 2\kappa_1\omega_\varepsilon^2\right)}{(1 + \kappa_1^2)}}. \tag{46}$$

From (45) and (46), we obtain (27b) and (27c).

Appendix B: Results for the Two-factor SR-SARV Model

In this Appendix, we provide the results for the case of the two-factor SR-SARV model ($\omega_2^2 \neq 0$).

ARMA representation for the two-factor SR-SARV model

First, we summarize the ARMA representation result in Meddahi (2003). Meddahi (2003, Proposition 3.3) shows that if the process of $p(t)$ belongs to the two-factor SR-SARV model, then IV_t follows an ARMA(2, 2) process:

$$IV_t = c_{IV} + (\kappa_1 + \kappa_2)IV_{t-1} - \kappa_1\kappa_2 IV_{t-2} + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}, \tag{47}$$

where κ_1 and κ_2 are defined as in the statement below (2), η_t is a white noise process with $\mathrm{var}(\eta_t) = \sigma_\eta^2$. Let $\phi_1 = \kappa_1 + \kappa_2$ and $\phi_2 = -\kappa_1\kappa_2$.[24] Other ARMA(2, 2) model parameters in (47), 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:

$$c_{IV} = (1 - \phi_1 - \phi_2)\sigma^2, \quad \theta_1 = \frac{1 - \sqrt{4s+1}}{2}\frac{\rho_1}{\rho_2}, \quad \theta_2 = \frac{\sqrt{4s+1} - 2s - 1}{2s},$$

$$\sigma_\eta^2 = \frac{\pi_1\mathrm{var}[IV_t] - 2\pi_2\mathrm{cov}[IV_t, IV_{t-1}] - 2\phi_2\mathrm{cov}[IV_t, IV_{t-2}]}{1 + \theta_1^2 + \theta_2^2}, \tag{48}$$

where

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

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

$$\rho_1 \equiv \frac{-\phi_1(1 - \phi_2) + (1 + \phi_1^2 - \phi_2)\mathrm{corr}[IV_t, IV_{t-1}] - \phi_1\mathrm{corr}[IV_t, IV_{t-2}]}{(1 + \phi_1^2 + \phi_2^2) - 2\phi_1(1 - \phi_2)\mathrm{corr}[IV_t, IV_{t-1}] - 2\phi_2\mathrm{corr}[IV_t, IV_{t-2}]}, \tag{49}$$

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

$$\mathrm{corr}[IV_t, IV_{t-2}] = \frac{\mathrm{cov}[IV_t, IV_{t-2}]}{\mathrm{var}[IV_t]}, \quad \mathrm{cov}[IV_t, IV_{t-2}] = \frac{\omega_1^2\kappa_1(1-\kappa_1)^2}{(\log\kappa_1)^2} + \frac{\omega_2^2\kappa_2(1-\kappa_2)^2}{(\log\kappa_2)^2},$$

and $\mathrm{sign}(\rho_2) = 1$ if $\rho_2 > 0$ and $\mathrm{sign}(\rho_2) = -1$ 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 (47), respectively. See Meddahi (2002b) and Meddahi (2003) for more details.

Assume that $\omega_2 \neq 0$. Let η_t and ξ_t be denoted by the state variables α_t and β_t, respectively. From (11), (17), (18), and (47),[25] we can express the NCRV in the following state space form:

Observation equation

$$RV_t^* = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_t \\ IV_{t-1} \\ u_t \\ \alpha_t \\ \alpha_{t-1} \\ \beta_t \end{bmatrix} + d_t, \tag{50a}$$

---

[24]Note that $\kappa_1 = \frac{1}{2}[\phi_1 + (\phi_1^2 + 4\phi_2)^{1/2}]$ and $\kappa_2 = \frac{1}{2}[\phi_1 - (\phi_1^2 + 4\phi_2)^{1/2}]$.
[25]The proof of Lemma 2 is valid for the two-factor model without any modification.

State equation

$$
\begin{bmatrix} IV_t \\ IV_{t-1} \\ u_t \\ \alpha_t \\ \alpha_{t-1} \\ \beta_t \end{bmatrix} = \begin{bmatrix} c_{IV} \\ 0 \\ c_u \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 & 0 & \theta_1 & \theta_2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \theta_u \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} IV_{t-1} \\ IV_{t-2} \\ u_{t-1} \\ \alpha_{t-1} \\ \alpha_{t-2} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}, \quad \text{(50b)}
$$

where the mean vector and variance matrix of $(d_t, \eta_t, \xi_t)'$ are as given in (19c).

Autocovariance functions

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

$$
(1 - \phi_1 L - \phi_2 L^2) RV_t^* = c_{RV} + (1 + \delta_1 L + \delta_2 L^2 + \delta_3 L^3)\tau_t, \ \tau_t \sim WN(0, \sigma_\tau^2). \tag{51}
$$

The same NCRV can alternatively be expressed as:

$$
\begin{aligned}
(1 - \phi_1 L - \phi_2 L^2) RV_t^* = \ & c_{IV} + \eta_t + \theta_1 \eta_{t-1} + \theta_2 \eta_{t-2} + d_t - \phi_1 d_{t-1} - \phi_2 d_{t-1} \\
& + (1 - \phi_1 - \phi_2) c_u + \xi_t + (\theta_u - \phi_1)\xi_{t-1} \\
& - (\phi_2 + \phi_1 \theta_u)\xi_{t-2} - \phi_2 \theta_u \xi_{t-3},
\end{aligned} \tag{52}
$$

The autocovariance functions of the MA process in (51) are given as:

$$
\begin{aligned}
& \gamma_0 = (1 + \delta_1^2 + \delta_2^2 + \delta_3^2)\sigma_\tau^2, \quad \gamma_1 = (\delta_1 + \delta_1 \delta_2 + \delta_2 \delta_3)\sigma_\tau^2, \\
& \gamma_2 = (\delta_2 + \delta_1 \delta_3)\sigma_\tau^2, \qquad\qquad \gamma_3 = \delta_3 \sigma_\tau^2,
\end{aligned} \tag{53}
$$

and $\gamma_j = 0$, for $j \geq 4$. Also, simple calculations lead us to the following autocovariance functions of the MA process in (52):

$$
\gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_\eta^2 + \pi_1 \sigma_d^2 + \left[\pi_1 \left(\frac{1}{\theta_u} + \theta_u \right) - 2\pi_2 \right] \theta_u \sigma_\xi^2, \tag{54a}
$$

$$
\gamma_1 = (\theta_1 + \theta_1 \theta_2)\sigma_\eta^2 - \pi_2 \sigma_d^2 + \left[\pi_1 - \phi_2 - \pi_2 \left(\frac{1}{\theta_u} + \theta_u \right) \right] \theta_u \sigma_\xi^2, \tag{54b}
$$

$$
\gamma_2 = \theta_2 \sigma_\eta^2 - \phi_2 \sigma_d^2 - \left[\phi_2 \left(\frac{1}{\theta_u} + \theta_u \right) + \pi_2 \right] \theta_u \sigma_\xi^2, \tag{54c}
$$

$$
\gamma_3 = -\phi_2 \theta_u \sigma_\xi^2, \tag{54d}
$$

and $\gamma_j = 0$, for $j \geq 4$. By equating the means of the MA processes on the right hand sides in (51) and (52), we obtain:

$$
c_{RV} = c_{IV} + (1 - \phi_1 - \phi_2)c_u. \tag{54e}
$$

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 given in (50a) and (50b) is not identified.

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 $j = 0 \sim 3$.

Proposition 2 *Under the condition $\sigma_\varepsilon^2 > 0$, Equations (54a)~(54e) can be uniquely solved for σ^2, ω_1^2, ω_2^2, σ_ε^2 and ω_ε^2 as:*

$$\omega_\varepsilon^2 = -\frac{\gamma_3}{\phi_2}, \quad \omega_1^2 = \frac{(\log \kappa_1)^2[(\alpha_1\beta_0 - \alpha_0\beta_1) - \kappa_2(\alpha_0\phi_1 + \alpha_2\beta_0)]}{(1 - \kappa_1)^2(\kappa_1 - \kappa_2)(\alpha_1\phi_1 - \alpha_2\beta_1)}, \tag{55a}$$

$$\omega_2^2 = \frac{(\log \kappa_2)^2[(\alpha_1\beta_0 - \alpha_0\beta_1) - \kappa_1(\alpha_0\phi_1 + \alpha_2\beta_0)]}{(1 - \kappa_2)^2(\kappa_2 - \kappa_1)(\alpha_1\phi_1 + \alpha_2\beta_1)}, \tag{55b}$$

$$\sigma_\varepsilon^2 = \frac{1}{2m}\sqrt{\frac{2c_{RV}^2}{(1 - \phi_1 - \phi_2)^2} + 2m(2m - 1)\omega_\varepsilon^2 + H}, \tag{55c}$$

and

$$\sigma^2 = \frac{c_{RV}}{1 - \phi_1 - \phi_2} - 2m\sigma_\varepsilon^2, \tag{55d}$$

where

$$\begin{aligned}
&\alpha_0 = \pi_2\gamma_0 + \pi_1\gamma_1 - [\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2, \quad \alpha_1 = \pi_1(1 + \phi_1^2 - \phi_2) - 2\pi_2^2, \\
&\alpha_2 = 2\phi_2\pi_2 + \phi_1\pi_1, \\
&\beta_0 = \pi_2\gamma_2 - \phi_2\gamma_1 + (\phi_1^2 + \phi_2)(1 + \phi_2^2 - \phi_2)\omega_\varepsilon^2, \quad \beta_1 = \phi_2^2 - \phi_1^2 - \phi_2,
\end{aligned} \tag{56}$$

$$H = \frac{m}{\phi_2}\left[\gamma_2 + \sum_{j=1}^{2}(2\phi_2 C_{1,j} + \phi_1 C_{2,j} - C_{3,j} + 2m\phi_2 C_{4,j})\omega_j^2 + \pi_2\omega_\varepsilon^2\right], \tag{57}$$

$$C_{1,j} \equiv \frac{\kappa_j - \log \kappa_j - 1}{(\log \kappa_j)^2}, \quad C_{2,j} \equiv \frac{(1 - \kappa_j)^2}{(\log \kappa_j)^2}, \quad C_{3,j} \equiv \frac{\kappa_j(1 - \kappa_j)^2}{(\log \kappa_j)^2}, \quad and \tag{58}$$

$$C_{4,j} = \frac{2(\kappa_j^{\frac{1}{m}} - \log \kappa_j^{\frac{1}{m}} - 1)}{(\log \kappa_j)^2} \quad for \quad j = 1, 2.$$

Proof. From $\theta_u\sigma_\xi^2 = \omega_\varepsilon^2$ in (24c) and $\gamma_3 = -\phi_2\theta_u\sigma_\xi^2$ in (54d), we have $\omega_\varepsilon^2 = -\frac{\gamma_3}{\phi_2}$, which is the first result in (55a). From (3), (4), (5), and (48), after simple calculations, it follows that:

$$\sigma_\eta^2 = \frac{2B_1\omega_1^2}{1 + \theta_1^2 + \theta_2^2} + \frac{2B_2\omega_2^2}{1 + \theta_1^2 + \theta_2^2} \quad and \quad \sigma_d^2 = \frac{2\sigma^4}{m} + 2mC_{4,1}\omega_1^2 + 2mC_{4,2}\omega_2^2, \tag{59}$$

where

$$B_j = \pi_1 C_{1,j} - \pi_2 C_{2,j} - \phi_2 C_{3,j} \quad for \quad j = 1, 2. \tag{60}$$

Substituting (59) into the autocovariance functions in (54a)~(54c), we have:

$$\gamma_0 = 2D_1\omega_1^2 + 2D_2\omega_2^2 + 2\frac{\pi_1}{m}\sigma^4 + \left[\pi_1\left(\frac{1}{\theta_u} + \theta_u\right) - 2\pi_2\right]\omega_\varepsilon^2, \tag{61a}$$

$$\gamma_1 = 2E_1\omega_1^2 + 2E_2\omega_2^2 - 2\frac{\pi_2}{m}\sigma^4 - \left[\pi_2\left(\frac{1}{\theta_u} + \theta_u\right) - (\pi_1 - \phi_2)\right]\omega_\varepsilon^2, \tag{61b}$$

$$\gamma_2 = 2F_1\omega_1^2 + 2F_2\omega_2^2 - 2\frac{\phi_2}{m}\sigma^4 - \left[\phi_2\left(\frac{1}{\theta_u} + \theta_u\right) + \pi_2\right]\omega_\varepsilon^2, \tag{61c}$$

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

$$\pi_2\gamma_0 + \pi_1\gamma_1 = (\pi_2 + \rho_1\pi_1)\left(2B_1\omega_1^2 + 2B_2\omega_2^2\right) + [\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2, \tag{62a}$$

and

$$\pi_2\gamma_2 - \phi_2\gamma_1 = (\rho_2\pi_2 - \rho_1\phi_2)\left(2B_1\omega_1^2 + 2B_2\omega_2^2\right) - [\phi_2(\pi_1 - \phi_2) + \pi_2^2]\omega_\varepsilon^2. \tag{62b}$$

Noting that ρ_1 and ρ_2 can be expressed as in (49) (see the explanations below (49)), we have:

$$
\begin{aligned}
\rho_1 &= \frac{-\pi_2\mathrm{var}[IV_t] + (1 + \phi_1^2 - \phi_2)\mathrm{cov}[IV_t, IV_{t-1}] - \phi_1\mathrm{cov}[IV_t, IV_{t-2}]}{(1 + \theta_1^2 + \theta_2^2)\sigma_\eta^2} \\
&= \frac{\sum_{j=1}^2[-2\pi_2 C_{1,j} + (1 + \phi_1^2 - \phi_2)C_{2,j} - \phi_1 C_{3,j}]\omega_j^2}{2B_1\omega_1^2 + 2B_2\omega_2^2},
\end{aligned} \tag{63a}
$$

and

$$
\begin{aligned}
\rho_2 &= \frac{-\phi_2\mathrm{var}[IV_t] - \phi_1\mathrm{cov}[IV_t, IV_{t-1}] + \mathrm{cov}[IV_t, IV_{t-2}]}{(1 + \theta_1^2 + \theta_2^2)\sigma_\eta^2} \\
&= \frac{\sum_{j=1}^2(-2\phi_2 C_{1,j} - \phi_1 C_{2,j} + C_{3,j})\omega_j^2}{2B_1\omega_1^2 + 2B_2\omega_2^2}.
\end{aligned} \tag{63b}
$$

Substituting B_j in (60), ρ_1 in (63a) and ρ_2 in (63b) into (62a) and (62b), we have:

$$
\begin{aligned}
\pi_2\gamma_0 + \pi_1\gamma_1 =\ & 2\pi_2\sum_{j=1}^2(\pi_1 C_{1,j} - \pi_2 C_{2,j} - \phi_2 C_{3,1})\omega_j^2 \\
& + \pi_1\sum_{j=1}^2[-2\pi_2 C_{1,j} + (1 + \phi_1^2 - \phi_2)C_{2,j} - \phi_1 C_{3,j}]\omega_j^2 \\
& + [\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2 \\
=\ & \sum_{j=1}^2\{[-2\pi_2^2 + \pi_1(1 + \phi_1^2 - \phi_2)]C_{2,j} - (2\phi_2\pi_2 + \phi_1\pi_1)C_{3,j}\}\omega_j^2 \\
& + [\pi_1(\pi_1 - \phi_2) - 2\pi_2^2]\omega_\varepsilon^2,
\end{aligned} \tag{64}
$$

and

$$
\begin{aligned}
\pi_2\gamma_2 - \phi_2\gamma_1 =\ & \pi_2\sum_{j=1}^2(-2\phi_2 C_{1,j} - \phi_1 C_{2,j} + C_{3,j})\omega_j^2 \\
& - \phi_2\sum_{j=1}^2[-2\pi_2 C_{1,j} + (1 + \phi_1^2 - \phi_2)C_{2,j} - \phi_1 C_{3,j}]\omega_j^2 \\
& - [\phi_2(\pi_1 - \phi_2) + \pi_2^2]\omega_\varepsilon^2 \\
=\ & \sum_{j=1}^2\{[-\phi_1\pi_2 - \phi_2(1 + \phi_1^2 - \phi_2)]C_{2,j} + (\pi_2 + \phi_1\phi_2)C_{3,j}\}\omega_j^2 \\
& - [\phi_2(\pi_1 - \phi_2) + \pi_2^2]\omega_\varepsilon^2 \\
=\ & \sum_{j=1}^2\{(\phi_2^2 - \phi_1^2 - \phi_2)C_{2,j} + \phi_1 C_{3,j}\}\omega_j^2 \\
& - (\phi_1^2 + \phi_2)(\phi_2^2 - \phi_2 + 1)\omega_\varepsilon^2.
\end{aligned} \tag{65}
$$

We can regard (64) and (65) as the following system of two equations for ω_1^2 and ω_2^2:

$$
\begin{aligned}
\alpha_0 &= (\alpha_1 C_{2,1} - \alpha_2 C_{3,1})\omega_1^2 + (\alpha_1 C_{2,2} - \alpha_2 C_{3,2})\omega_2^2 \\
\beta_0 &= (\beta_1 C_{2,1} + \phi_1 C_{3,1})\omega_1^2 + (\beta_1 C_{2,2} + \phi_1 C_{3,2})\omega_2^2,
\end{aligned} \tag{66}
$$

where α_0, α_1, α_2, β_0 and β_1 are as given in (56). Solving (66), we have

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

and

$$\omega_2^2 = \frac{(\alpha_0\beta_1 - \alpha_1\beta_0)C_{2,1} + (\alpha_0\phi_1 + \alpha_2\beta_0)C_{3,1}}{(\alpha_1\phi_1 + \alpha_2\beta_1)C_{2,2}C_{3,1} - (\alpha_1\phi_1 + \alpha_2\beta_1)C_{3,2}C_{2,1}}$$

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

From (54e) and (24a), we have:

$$\sigma_\varepsilon^2 = \frac{c_{RV} - (1 - \phi_1 - \phi_2)\sigma^2}{2(1 - \phi_1 - \phi_2)m}. \tag{67}$$

Substituting σ_ε^2 in (67) into A in (25), we have:

$$2A = \frac{2c_{RV}\sigma^2}{(1 - \phi_1 - \phi_2)m\omega_\varepsilon^2} - \frac{3\sigma^4}{m\omega_\varepsilon^2} + 2(2m - 1) + \frac{c_{RV}^2}{(1 - \phi_1 - \phi_2)^2 m\omega_\varepsilon^2}. \tag{68}$$

From (41), (61c) and (68), we have:

$$\gamma_2 = 2\omega_1^2(\rho_2 B_1 - m\phi_2 C_{4,1}) + 2\omega_2^2(\rho_2 B_2 - m\phi_2 C_{4,2}) - 2\sigma^4\frac{\phi_2}{m}$$
$$- \left\{\phi_2\left[\frac{2c_{RV}\sigma^2}{(1-\phi_1-\phi_2)m\omega_\varepsilon^2} - \frac{3\sigma^4}{m\omega_\varepsilon^2} + 2(2m-1) + \frac{c_{RV}^2}{(1-\phi_1-\phi_2)^2 m\omega_\varepsilon^2}\right] + \pi_2\right\}\omega_\varepsilon^2$$

$$= \rho_2(2B_1\omega_1^2 + 2B_2\omega_2^2) - \phi_2(2m\omega_1^2 C_{4,1} + 2m\omega_2^2 C_{4,2})$$
$$+ \sigma^4\frac{\phi_2}{m} - \sigma^2\frac{2\phi_2 c_{RV}}{(1-\phi_1-\phi_2)m} - 2\phi_2\omega_\varepsilon^2(2m-1) - \frac{\phi_2 c_{RV}^2}{(1-\phi_1-\phi_2)^2 m} - \pi_2\omega_\varepsilon^2 \tag{69}$$

$$= \sum_{j=1}^{2}(-2\phi_2 C_{1,j} - \phi_1 C_{2,j} + C_{3,j} - 2m\phi_2 C_{4,j})\omega_j^2$$
$$+ \sigma^4\frac{\phi_2}{m} - \sigma^2\frac{2\phi_2 c_{RV}}{(1-\phi_1-\phi_2)m} - 2\phi_2\omega_\varepsilon^2(2m-1) - \frac{\phi_2 c_{RV}^2}{(1-\phi_1-\phi_2)^2 m} - \pi_2\omega_\varepsilon^2.$$

Multiplying both sides in (69) 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 (57). Solving the quadratic equation for σ^2, and by the same argument as used in (46), we have:

$$\sigma^2 = \frac{c_{RV}}{1 - \phi_1 - \phi_2} - \sqrt{\frac{2c_{RV}^2}{(1 - \phi_1 - \phi_2)^2} + 2m(2m - 1)\omega_\varepsilon^2 + H}, \tag{70}$$

and

$$\sigma_\varepsilon^2 = \frac{1}{2m}\sqrt{\frac{2c_{RV}^2}{(1 - \phi_1 - \phi_2)^2} + 2m(2m - 1)\omega_\varepsilon^2 + H}. \tag{71}$$

From (70) and (71), we have (55c) and (55d).

Lastly, we summarize direct and indirect approaches for estimating the parameters for the two factor model.

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 (51) by QML estimation assuming Gaussian innovations.

Step 3 Given the estimates of $c_{RV}^{(m)}$, κ_1, κ_2,[26] $\delta_1^{(m)}$, $\delta_2^{(m)}$, $\delta_3^{(m)}$ 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 (53).

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, calculate ω_ε^2, σ_ε^2, ω_1^2, ω_2^2 and σ^2, applying the results in (55a) $-$ (55c).

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), (4), (48), and (25).

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 (50a) $-$ (50b), 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.

---

[26]These can be obtained from the estimates of ϕ_1 and ϕ_2.

References

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005), How Often to Sample a Continuous-time Process in the Presence of Market Microstructure Noise, *Review of Financial Studies*, Vol. 18, No. 2, pp. 351–416.

Anderson, B. D. O., and Moore, J. B. (1979), *Optimal Filtering*, Prentice-Hall, NJ.

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

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001), The Distribution of Stock Return Volatility, *Journal of Financial Economics*, Vol. 61, pp. 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), The Distribution of Exchange Rate Volatility, *Journal of the American Statistical Association*, Vol. 92, pp. 42–55.

Andersen, T. G., Bollerslev, T., Diebold. F. X., and Labys, P. (2003) Modeling and Forecasting Realized Volatility, *Econometrica*, Vol. 71, No. 2, pp.579-625.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005), Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities, *Econometrica*, Vol. 73, No. 1, pp. 279-296.

Andersen, T. G., Dobrev D., and Schaumburg, E. (2009), Duration–Based Volatility Estimation, *Global COE Hi-Stat Discussion Paper Series* 034. Available at http://gcoe.ier.hit-u.ac.jp/research/discussion/2008/pdf/gd08-034.pdf

Bandi, F. M., and Russell, J. R. (2006), Separating Microstructure Noise from Volatility, *Journal of Financial Economics*, Vol. 79, pp. 655–692.

Bandi, F. M., and Russell, J. R. (2008), Microstructure Noise, Realized Variance, and Optimal Sampling, *Review of Economic Studies*, Vol. 75, pp. 339–369.

Barndorff-Nielsen, O. E., and 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)*, Vol. 63, pp. 167–241.

Barndorff-Nielsen, O. E., and Shephard, N. (2002), Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Vol. 64, pp. 253–280.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008), Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise, *Econometrica*, Vol. 76, pp. 1481–1536.

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

Beine, M., Lahaye, J., Laurent, S. C., Neely, J., and Palm, F. C. (2007), Central Bank Intervention and Exchange Rate Volatility, Its Continuous and Jump Components, *International Journal of Finance and Economics*, Vol. 12, pp. 201–223.

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

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994), ARCH Models, pp. 2959–3038, in *Handbook of Econometrics* Vol. 4, edited by R. F. Engle, D. L. McFadden. Elsevier, NY.

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

Brockwell, P.J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd ed., Springer–Verlag, New York.

Durbin, J., and S. J. Koopman, (2001), *Time Series Analysis by State Space Methods*. Oxford University Press, NY.

Ghysels, E., Harvey, A. C., and Renault, E. (1996), Stochastic Volatility, pp. 119–192, in *Handbook of Statistics* 14, edited by G. S. Maddala and C. R. Rao. Elsevier, NY.

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

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

Hansen, P. R., and Horel, G. (2009), Quadratic Variation by Markov Chains, University of Aarhus Department of Economics Research Paper No. 2009-13. Available at SSRN: http://ssrn.com/abstract=1367519

Hansen, P. R., and Lunde, A. (2006), Realized Variance and Market Microstructure Noise, *Journal of Business and Economic Statistics*, Vol. 24, pp. 127–161.

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

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009), Microstructure Noise in the Continuous Case: The Pre-averaging Approach, *Stochastic Processes and their Applications*, Vol. 119, pp. 2249-2276.

Jungbacker, B., and Koopman, S. J. (2006), Model-based Measurement of Actual Volatility in High-frequency Data, *Advances in Econometrics*, Vol. 20, pp.183-210.

McAleer, M., and Medeiros, M. C. (2008), Realized Volatility: A Review, *Econometrics Review*, Vol. 27, pp.10-45.

Meddahi, N., (2001), An Eigenfunction Approach for Volatility Modeling, CIRANO Working Paper, 2001s–70.

Meddahi, N. (2002a), A Theoretical Comparison Between Integrated and Realized Volatility, *Journal of Applied Econometrics*, Vol. 17, pp. 479-508.

Meddahi, N. (2002b), ARMA Representation of Two-factor Models, CIRANO Working Paper, 2002s–92.

Meddahi, N. (2003), ARMA Representation of Integrated and Realized Variances, *Econometrics Journal*, Vol. 6, pp. 335–356.

Meddahi, N., and Renault, E. (2004), Temporal Aggregation of Volatility Models, *Journal of Econometrics*, Vol. 119, pp. 355-379.

Morley, J., Nelson, C. R., and Zivot, E. (2003), Why are the Beveridge-Nelson and Unobserved Components Decompositions of GDP So Different? *The Review of Economics and Statistics*, Vol.85, pp. 243-253.

Nagakura, D., and Watanabe, T. (2009), A State Space Approach to Estimating the Integrated Variance and Microstructure Noise Component, *IMES Discussion Paper Series*, 09-E-11

Nelson, D. B. (1990), ARCH Models as Diffusion Approximations, *Journal of Econometrics* Vol. 45, pp. 7–39.

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

Palm, F. C. (1996), GARCH Models of Volatility, pp. 209-240, in *Handbook of Statistics* Vol. 14, edited by G. S. Maddala and C. R. Rao. Elsevier, NY.

Romano, P. J., and Thombs L. A. (1996), *Journal of the American Statistical Association*, Vol. 91, No. 434, pp.590-600.

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

Ubukata, M., and Oya, K. (2009), Estimation and Testing for Dependence in Market Microstructure Noise, *Journal of Financial Econometrics*, Vol. 7, No. 2, pp. 106-151.

Zhang, L. (2006), Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi Scale Approach, *Bernoulli*

Zhang, L., Mykland, P., and Aït-Sahalia, Y. (2005), A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-frequency Data, *Journal of the American Statistical Association*, Vol. 100, pp. 1394–1411.

Zhou, B. (1996), High-frequency Data and Volatility in Foreign-exchange Rates, *Journal of Business and Economic Statistics*, Vol. 14, pp. 45–52.

Zivot, E. (2008), Practical Issues in the Analysis of GARCH Models, forthcoming in *Handbook of Financial Time Series*, edited by T.G. Andersen, R.A. Davis, J-P Kreiss and T. Mikosch. Springer.

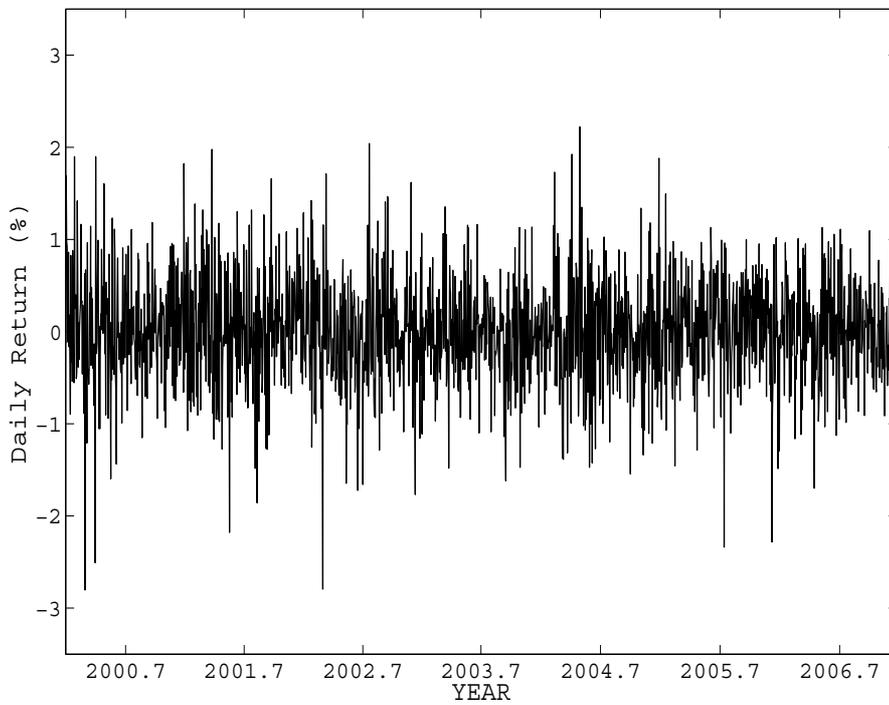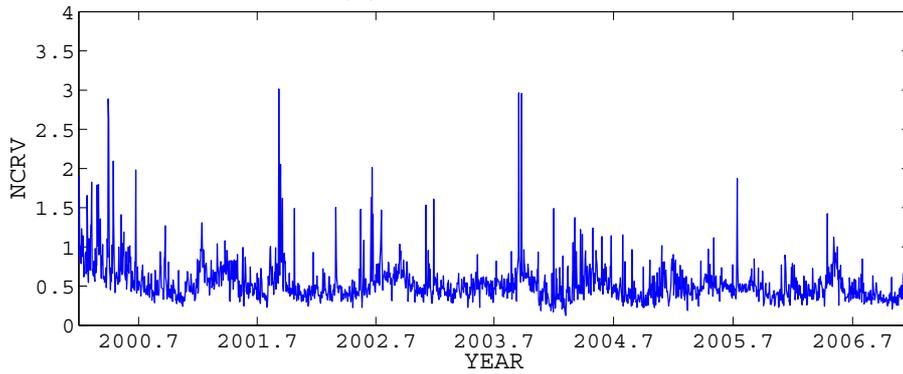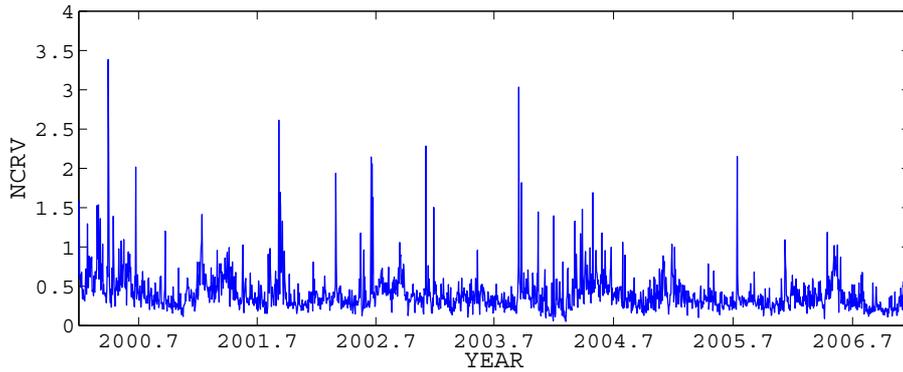Figure 1: Daily Returns of the Yen/Dollar Exchange Rate
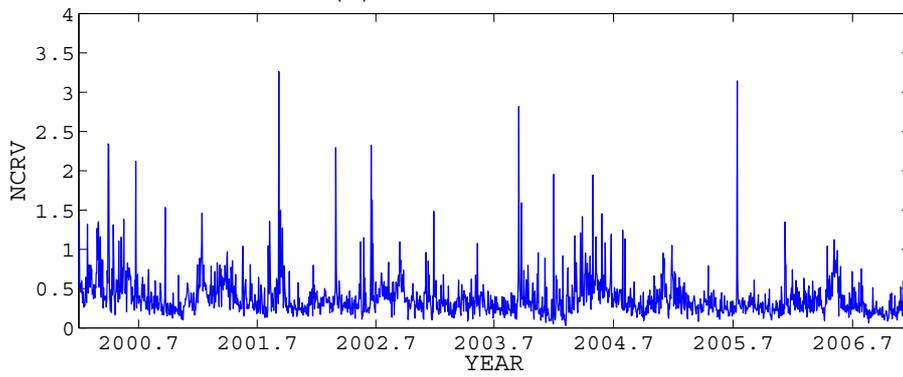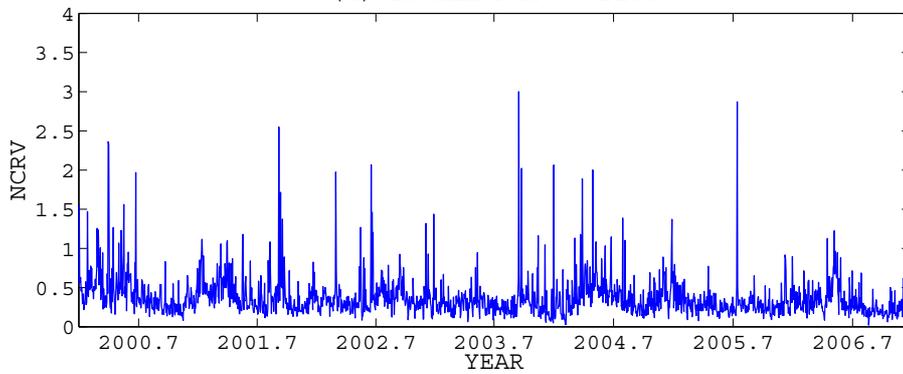
Figure 2: NCRV Series with Various m

Figure 2: (continued) NCRV Series with Various m

(d) 15-min NCRV series

(e) 30-min NCRV series
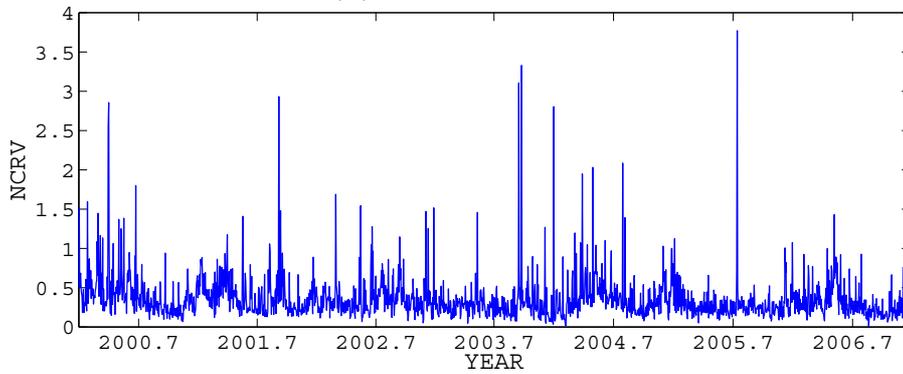
Table 1: Descriptive Statistics of Returns

	$r(1)$	$r(5)$	$r(10)$	$r(15)$	$r(30)$
m	1440	244	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 sample autocorrelations (SAC) of k-minute returns, which are denoted by $r(k)$, m is the number of intra-day returns for each return series, and SAC(k) is the value of the SAC of order k. 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 sample autocorrelation (SAC) of k-minute NCRV series, which are denoted by NCRV(k), m is the number of intra-day returns used for calculating each NCRV series, and SAC(k) is the value of the SAC of order k. 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 $\log(\widehat{\mu})$, 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. Unreported simulation experiment confirms that these standard errors are reasonably close to the actual standard errors in finite samples.

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
$\widehat{\phi}_1$	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
$\widehat{c}_u^{(m)}$	0.2450	0.0586	0.0308	0.0201
$\widehat{\theta}_u^{(m)}$	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
$\widehat{c}_u^{(m)}$	0.2121	0.0604	0.0293	0.0196
$\widehat{\theta}_u^{(m)}$	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

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: Mean Absolute Differences (and Correlations) among Smoothed IV Series with Various m

(a) One-factor model

	$n = 288$	$n = 144$	$n = 96$
$m = 1440$	0.0822 (0.9060)	0.0896 (0.8799)	0.0913 (0.8700)
$m = 288$	–	0.0179 (0.9841)	0.0244 (0.9718)
$m = 144$	–	–	0.0165 (0.9843)

(b) Two-factor model

	$n = 288$	$n = 144$	$n = 96$
$m = 1440$	0.0601 (0.9087)	0.0699 (0.8821)	0.0730 (0.8667)
$m = 288$	–	0.0205 (0.9779)	0.0257 (0.9702)
$m = 144$	–	–	0.0189 (0.9819)

Note: the table reports the mean absolute differences, which is defined as $T^{-1} \sum_{t=1}^{T} |\widehat{IV}_{t|T}^{(m)} - \widehat{IV}_{t|T}^{(n)}|$ for $m = 1440, 288, 144 \ n = 288, 144, 96$, where $\widehat{IV}_{t|T}^{(m)}$ is the smoothed value of IV_t, Correlation coefficients between two series are in parentheses. The numbers, 1440, 288, 144, 96 of m, n correspond to 1−, 5−, 10−, 15− minute, respectively.

Table 7: Results of the Jarque - Bera Normality Test for Normalized Returns

(a) Daily return without normalization

m	Mean	S.D.	Skewness	Kurtosis	JB test
1	0.0087	0.5802	−0.1437	4.3475	143.0909**

(b) Daily return normalized by the square root of the NCRV

m	Mean	S.D.	Skewness	Kurtosis	JB test
1440	0.0254	0.7586	0.0576	3.0101	1.0090
288	0.0335	0.8625	0.0695	2.7444	6.3786*
144	0.0348	0.8890	0.0603	2.6722	9.1963*
96	0.0349	0.9065	0.0368	2.6617	9.0329*

(c) Daily return normalized by the square root of the smoothed IV

One-factor model					
m	Mean	S.D.	Skewness	Kurtosis	JB test
1440	0.0282	1.1065	−0.0460	3.5032	19.7229**
288	0.0255	0.9537	−0.0290	3.2349	4.4120
144	0.0235	0.9445	−0.0398	3.2347	4.6294
96	0.0221	0.9475	−0.0575	3.3104	8.2567*
Two-factor model					
m	Mean	S.D.	Skewness	Kurtosis	JB test
1440	0.0326	1.0086	0.0289	3.1168	1.2808
288	0.0275	0.9595	−0.0079	3.1002	0.7762
144	0.0259	0.9309	−0.0107	3.0218	0.0701
96	0.0250	0.9330	−0.0235	3.0472	0.3347

Note: that table shows the sample mean (Mean), sample standard deviation (S.D.), sample skewness (Skewness), sample kurtosis (Kurtosis), and the results of the Jarque-Bera normality test for normalized daily returns, where m is the number of intra-daily returns used in calculating the NCRV and the smoothed IV. The superscripts, "∗" and "∗∗", indicate that the null of normality is rejected at the nominal level of 5% and 1%, respectively.

Table 8: Forecast Evaluation (In-sample Case)

(a) Results of Mincer - Zarnowitz style regression

m	One-factor model a_0	a_1		R^2	Two-factor model 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	0.02639	0.1171
	(0.0284)	(0.1687)	(0.2319)	−	(0.0308)	(0.6835)	(0.6866)	−

(b) Mean absolute error (MAE) and difference (MAD)

m	One-factor model $\mathrm{MAE}^{(NR)}$	$\mathrm{MAE}^{(BSM)}$	$\mathrm{MAD}^{(BSM,NR)}$	Two-factor model $\mathrm{MAE}^{(NR)}$	$\mathrm{MAE}^{(BSM)}$	$\mathrm{MAD}^{(BSM,NR)}$
1440	0.1593	0.2623	0.2594 (0.6929)	0.1586	0.2604	0.2502 (0.9978)
288	0.1547	0.1831	0.0841 (0.7327)	0.1536	0.1754	0.0606 (0.9884)
144	0.1553	0.1686	0.0461 (0.9299)	0.1552	0.1651	0.0304 (0.9918)
96	0.1545	0.1641	0.0395 (0.9383)	0.1541	0.1601	0.0210 (0.9929)

Note: the table (a) shows the results of the Mincer–Zarnowitz style regressions, and the table (b) reports the mean absolute error (MAE) of one-day-ahead in-sample forecasts, and the mean absolute difference (MAD) between the forecasts of our and BSM methods, which are defined as

$$T^{-1} \sum_{t=1}^{T-1} |RV_{t+1}^{*(48)} - IV_{t+1|t}^{X(m)}|, \quad \text{and} \quad T^{-1} \sum_{t=1}^{T-1} |IV_{t+1|t}^{NR(m)} - IV_{t+1|t}^{BSM(m)}|,$$

respectively, where $T = 1809$. The correlation coefficient between the two forecasts for a given m is in the parenthesis next to the MAD.

Table 9: 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 absolute error (MAE) and difference (MAD)

	One-factor model			Two-factor model		
m	$\text{MAE}^{(NR)}$	$\text{MAE}^{(BSM)}$	$\text{MAD}^{(BSM,NR)}$	$\text{MAE}^{(NR)}$	$\text{MAE}^{(BSM)}$	$\text{MAD}^{(BSM,NR)}$
1440	0.1250	0.2578	0.2717 (0.6349)	0.1264	0.2317	0.2485 (0.9730)
288	0.1253	0.1408	0.0468 (0.9643)	0.1257	0.1501	0.0623 (0.9864)
144	0.1299	0.1327	0.0288 (0.9301)	0.1319	0.1406	0.0278 (0.9583)
96	0.1294	0.1295	0.0263 (0.9428)	0.1287	0.1361	0.0244 (0.9734)

Note: the table (a) shows the results of the Mincer–Zarnowitz style regressions, and the table (b) reports the mean absolute error (MAE) of one-day-ahead out-of-sample forecasts, and the mean absolute difference (MAD) between the forecasts of our and BSM methods, which are defined as

$$T^{-1} \sum_{t=1}^{T-1} |RV_{t+1}^{*(48)} - IV_{t+1|t}^{X(m)}|, \quad \text{and} \quad T^{-1} \sum_{t=1}^{T-1} |IV_{t+1|t}^{NR(m)} - IV_{t+1|t}^{BSM(m)}|,$$

respectively, where $T = 1809$. The correlation coefficient between the two forecasts for a given m is in the parenthesis next to the MAD.

Figure 3: NCRV, Smoothed IV, and MN component Series (One-factor Model)
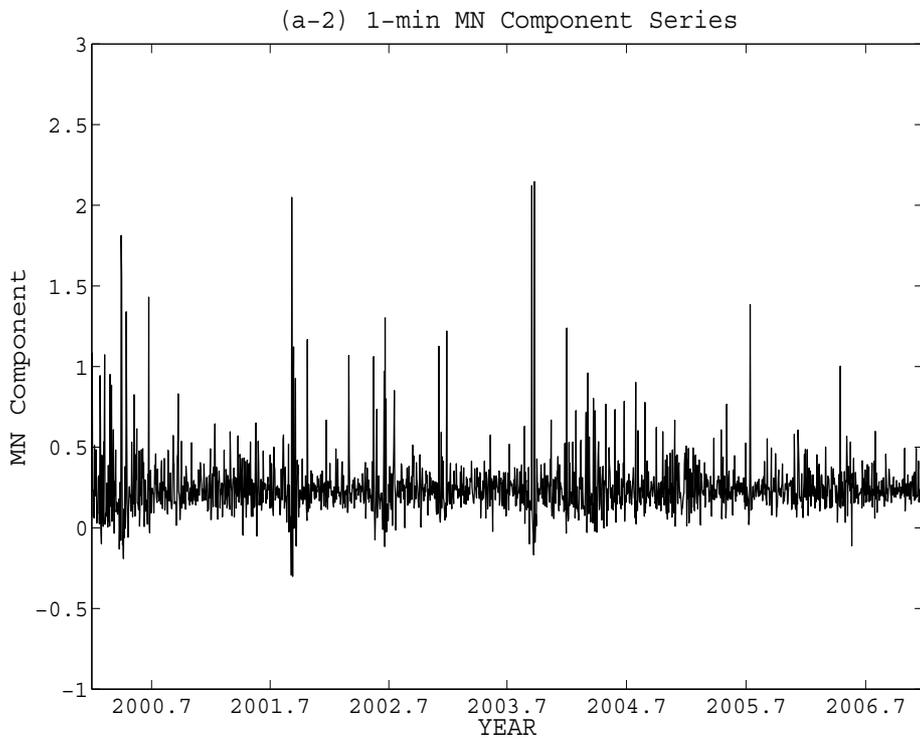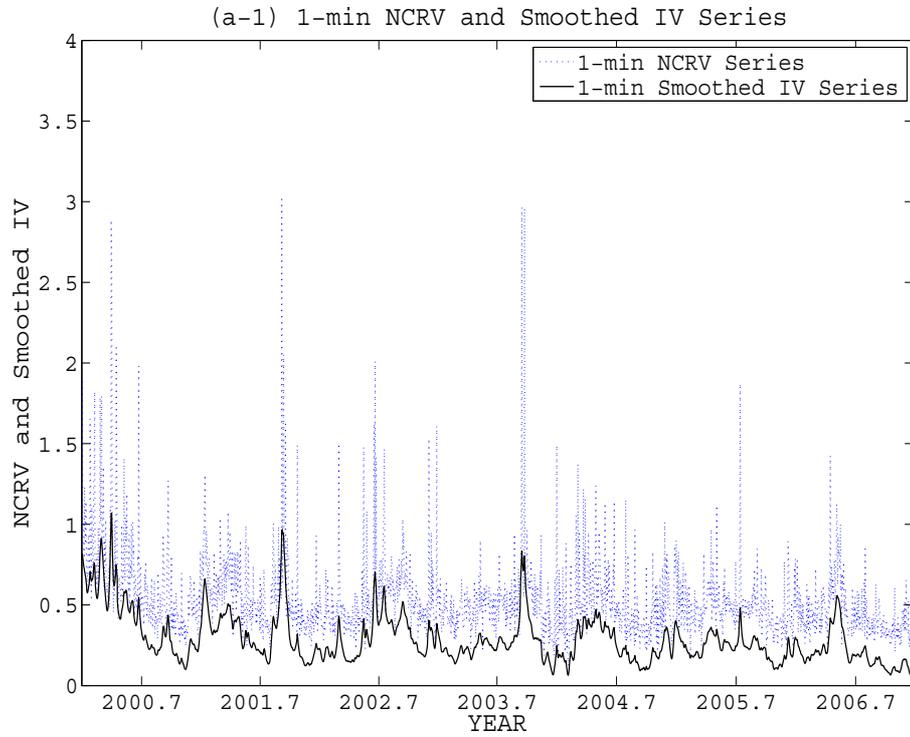
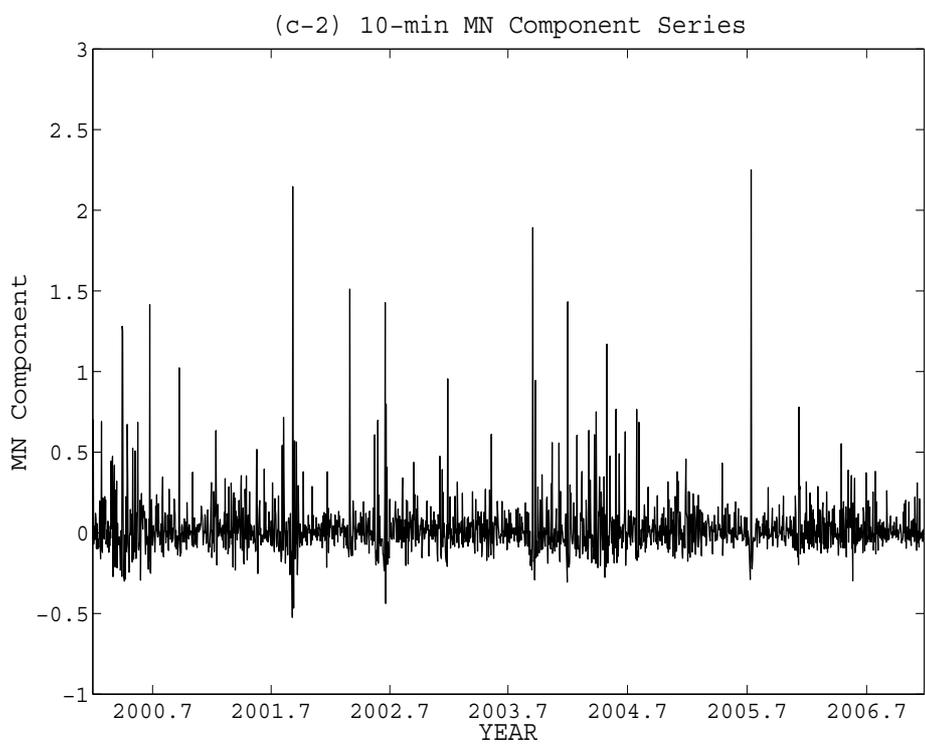(a-1) 1-min NCRV and Smoothed IV Series

(a-2) 1-min MN Component Series

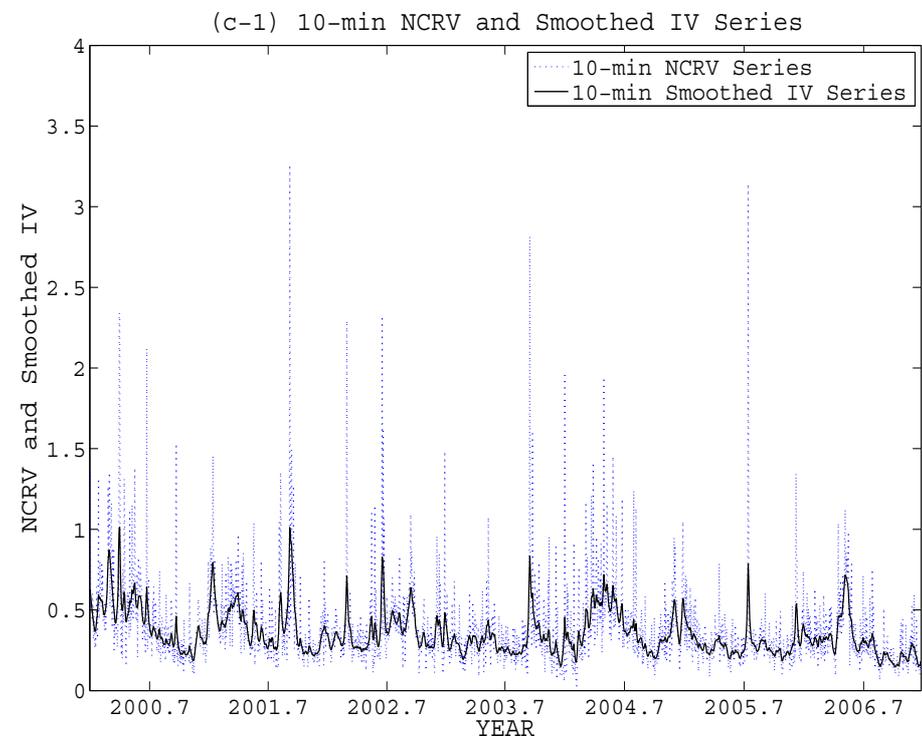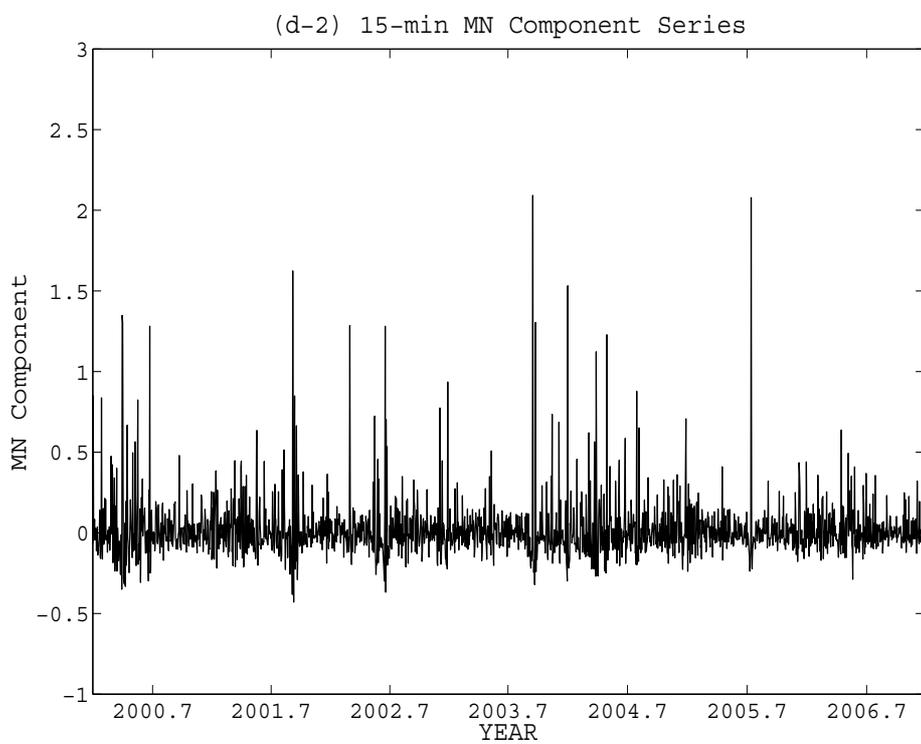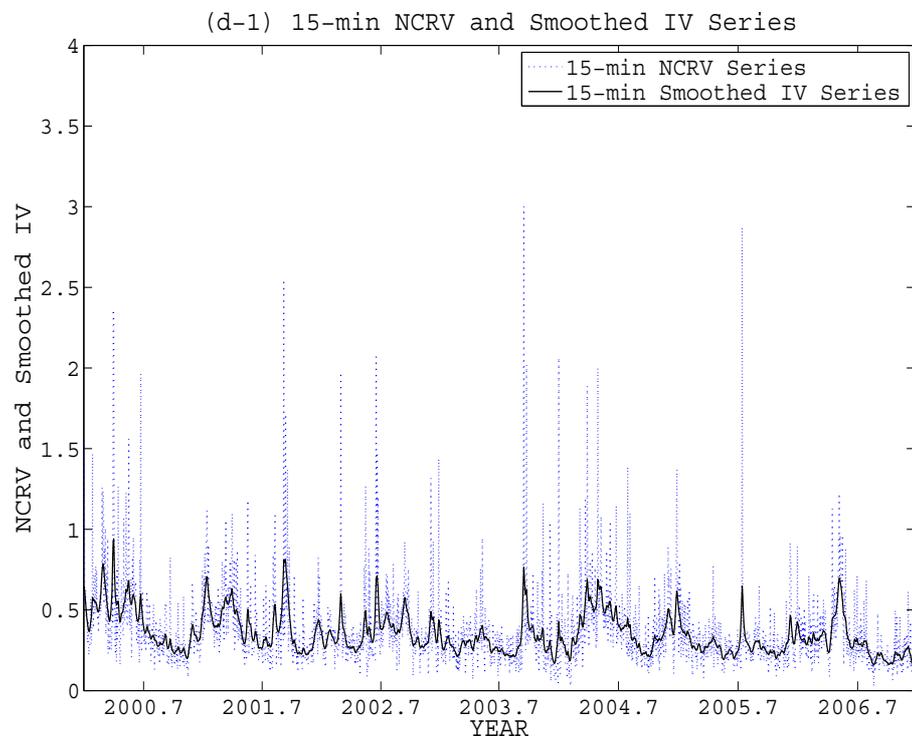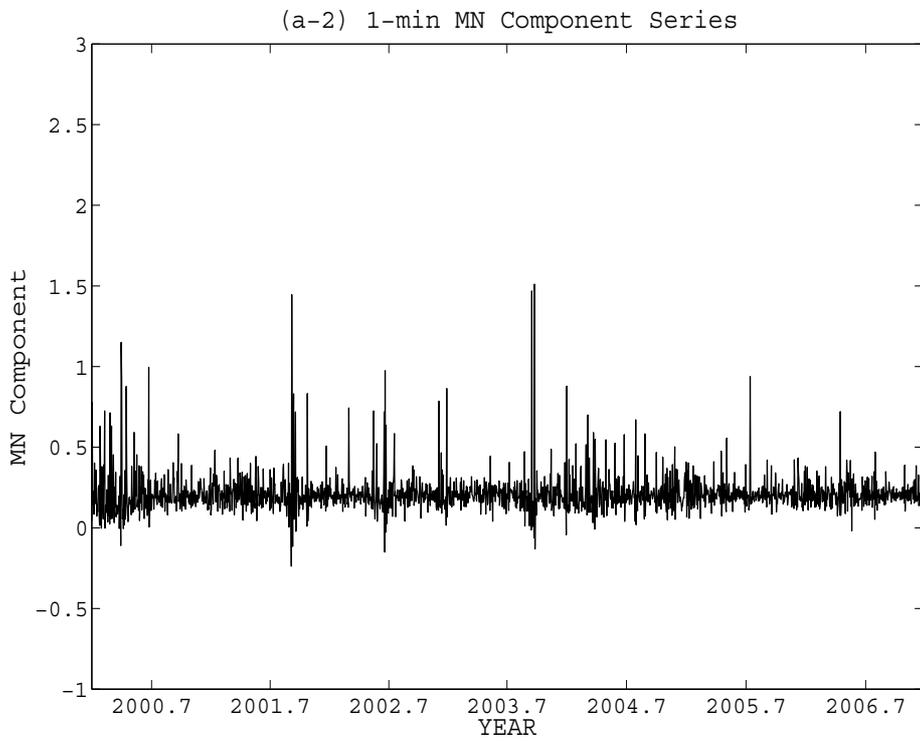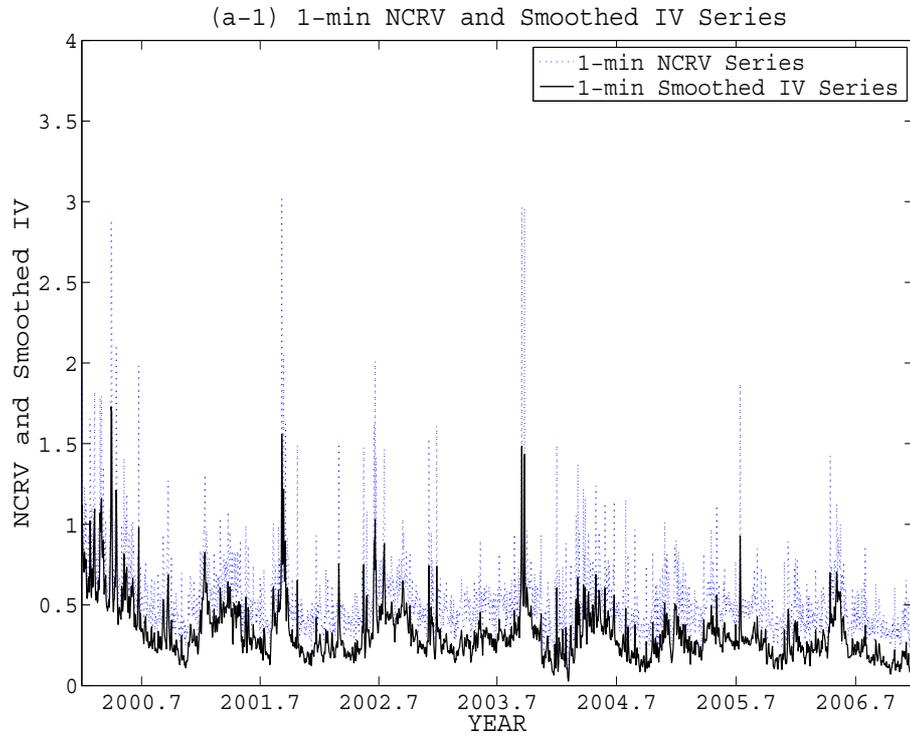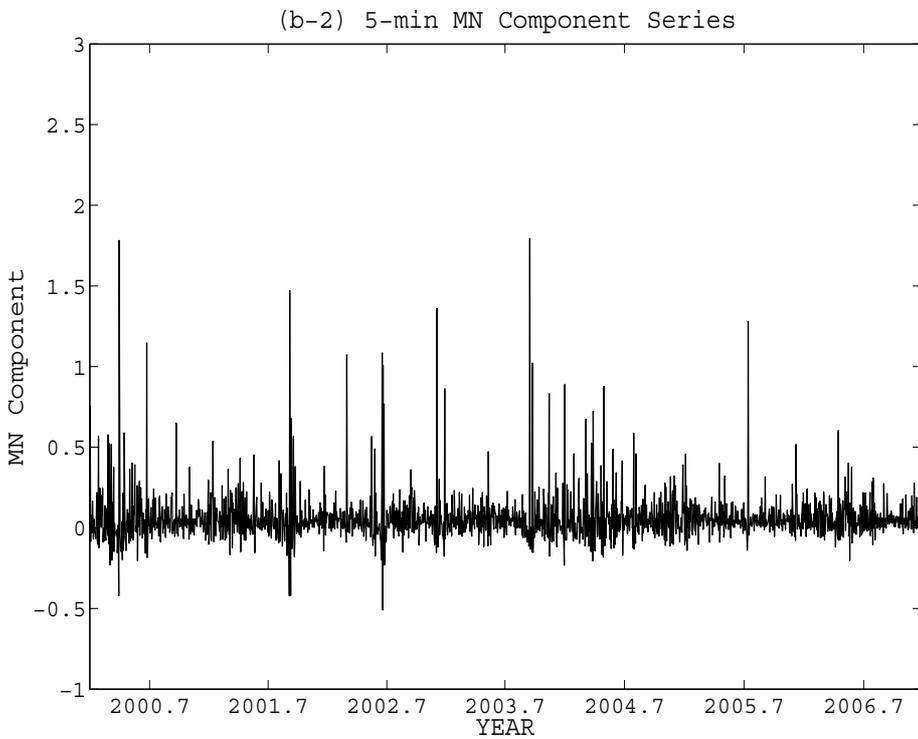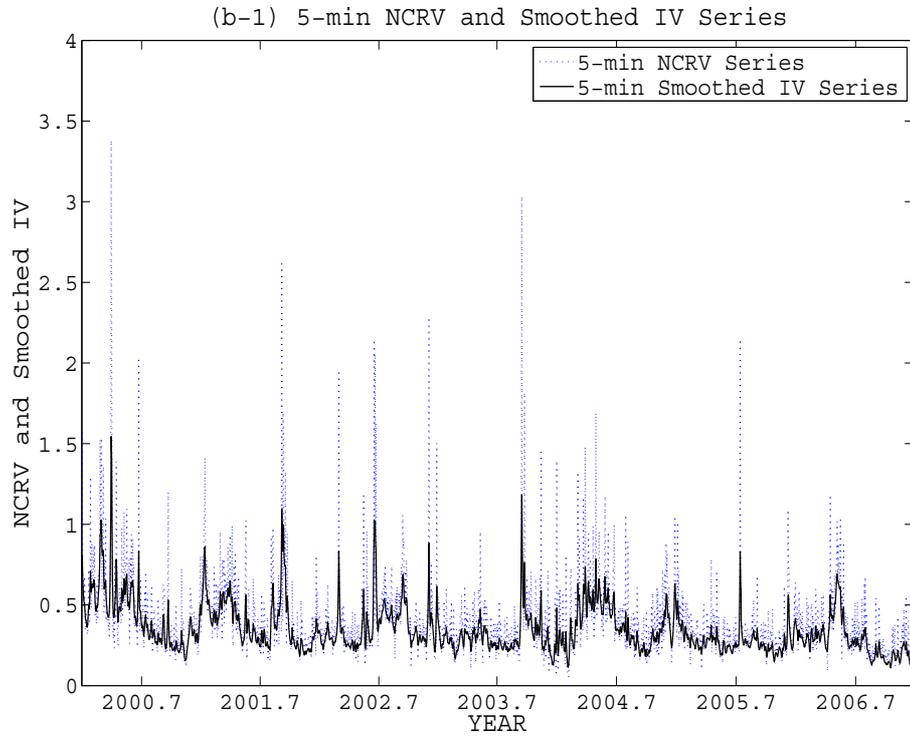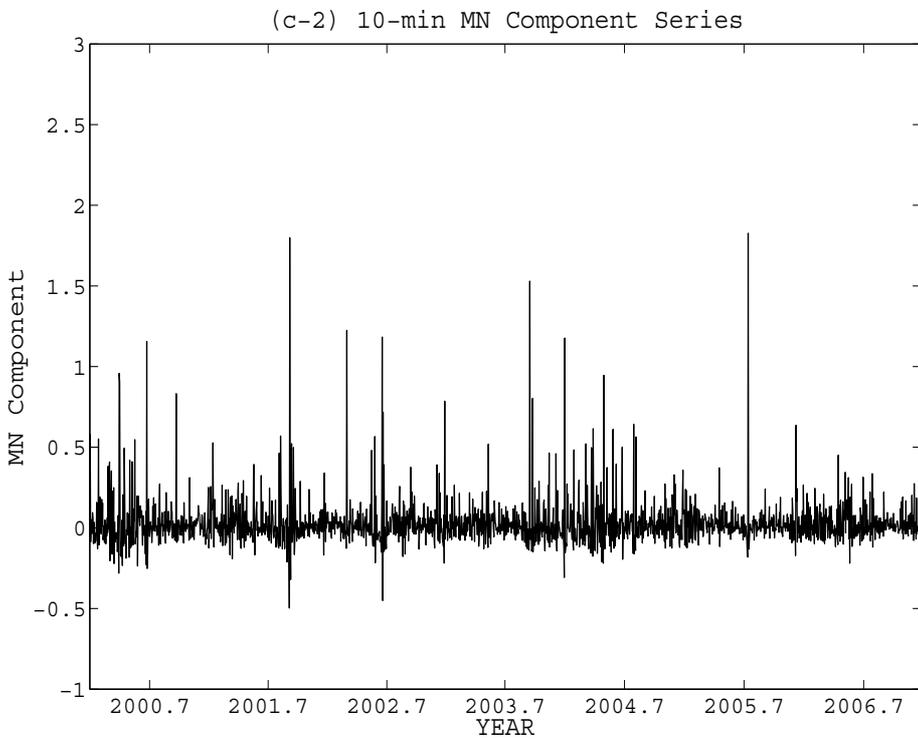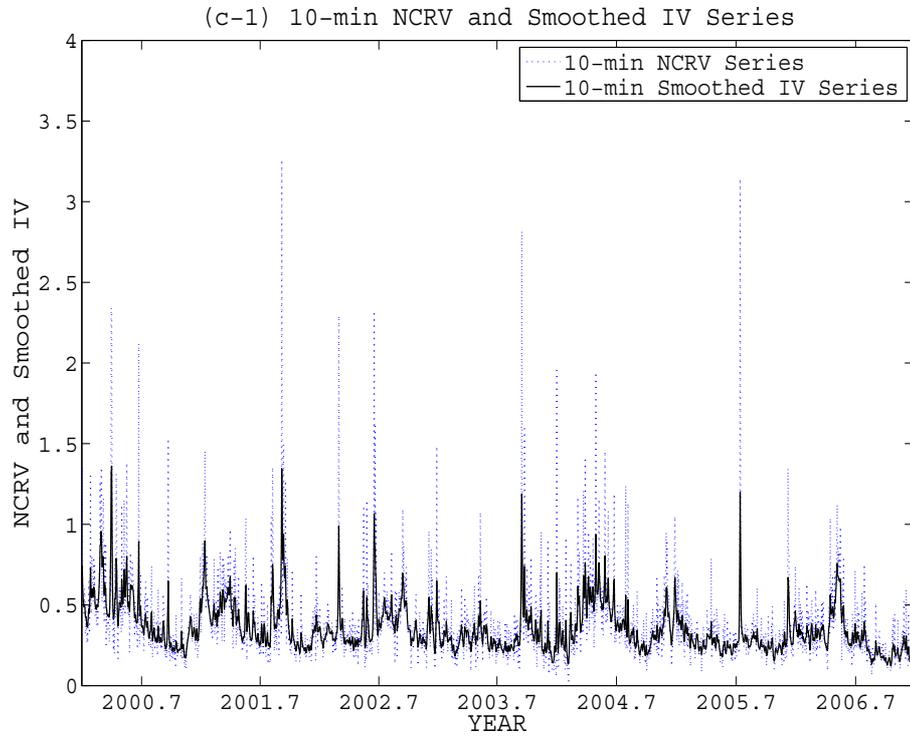Figure 3: (continued) NCRV, Smoothed IV, and MN component Series (One-factor Model)

(b-1) 5-min NCRV and Smoothed IV Series

(b-2) 5-min MN Component Series

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

(c-1) 10-min NCRV and Smoothed IV Series

(c-2) 10-min MN Component Series

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

(d-1) 15-min NCRV and Smoothed IV Series
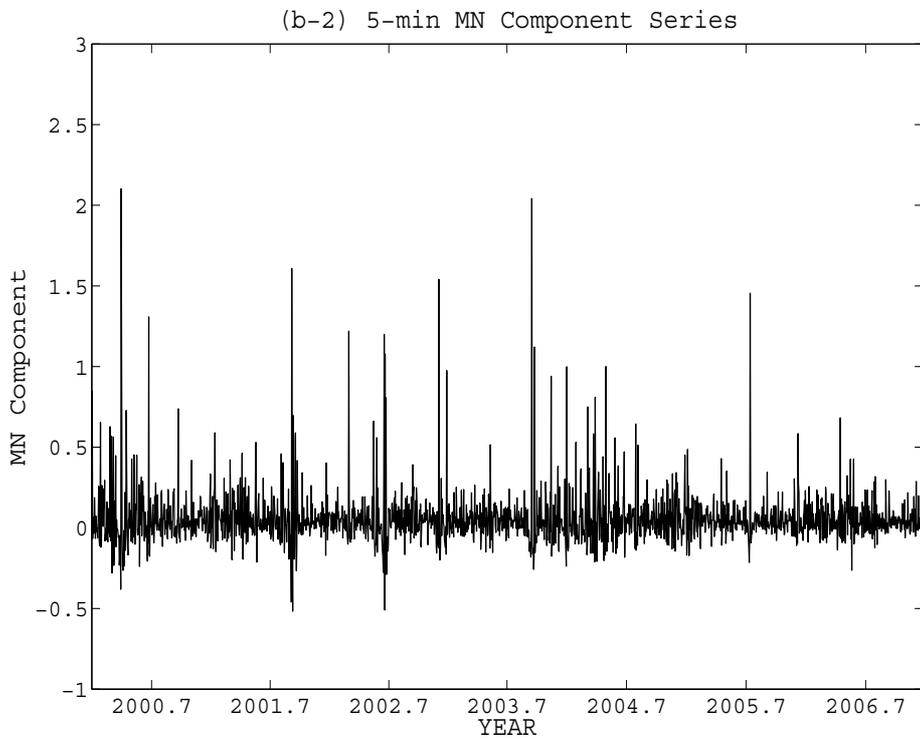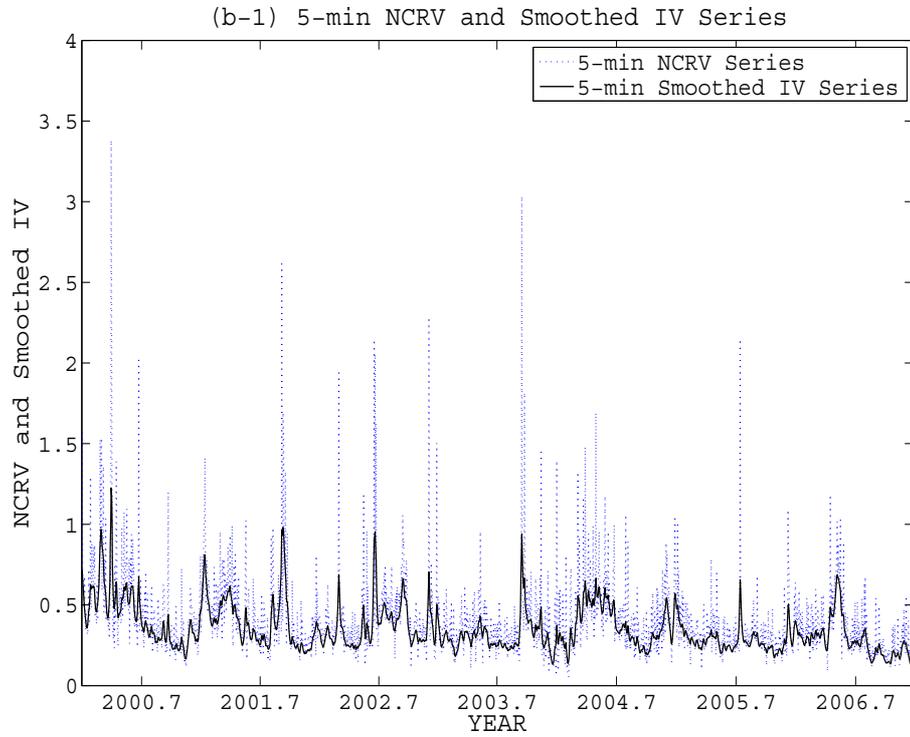
(d-2) 15-min MN Component Series

Figure 4: NCRV, Smoothed IV, and MN component Series (Two-factor Model)

(a-1) 1-min NCRV and Smoothed IV Series

(a-2) 1-min MN Component Series

Figure 4: (continued) NCRV, Smoothed IV, and MN component Series (Two-factor Model)

(b-1) 5-min NCRV and Smoothed IV Series

(b-2) 5-min MN Component Series

Figure 4: (continued) NCRV, Smoothed IV, and MN component Series (Two-factor Model)

(c-1) 10-min NCRV and Smoothed IV Series

(c-2) 10-min MN Component Series

Figure 4: (continued) NCRV, Smoothed IV, and MN component Series (Two-factor Model)

(d-1) 15-min NCRV and Smoothed IV Series

(d-2) 15-min MN Component Series
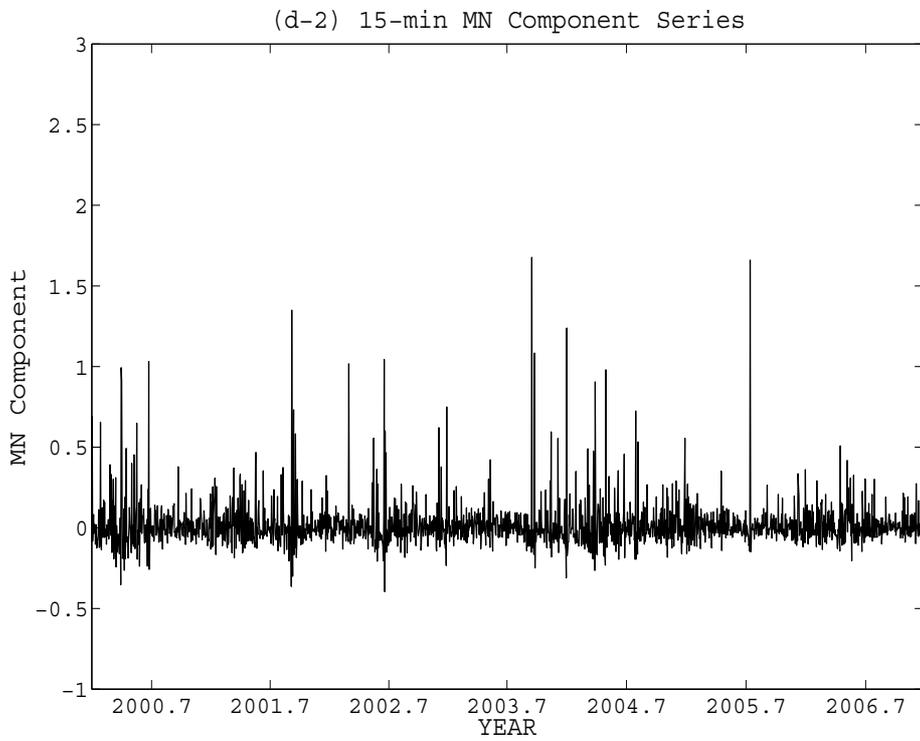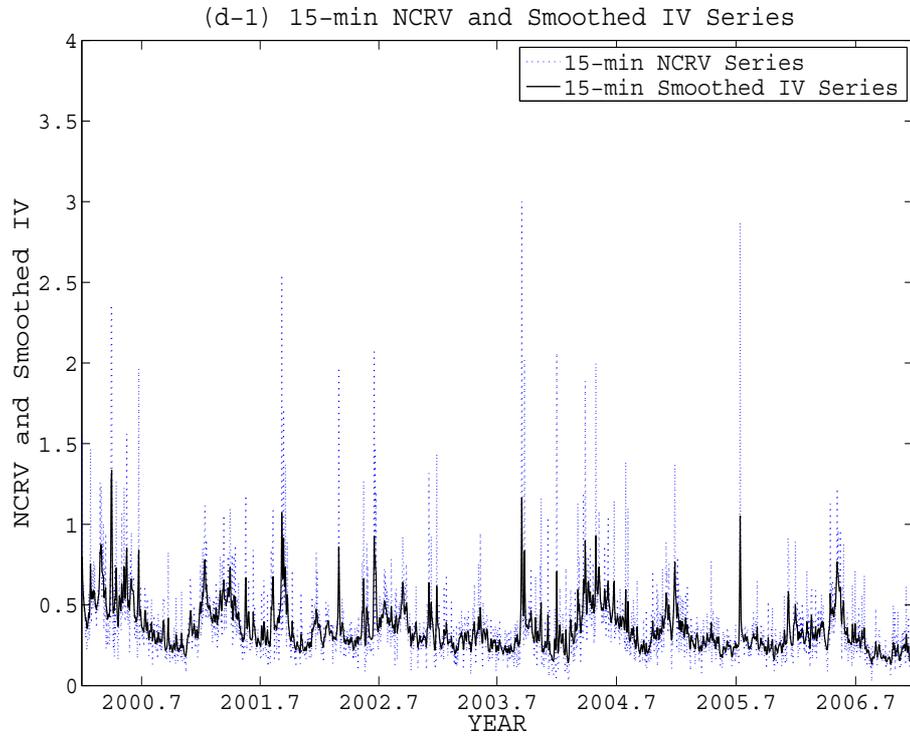
Figure 5: In Sample Forecasts (One-factor Model)

(a) 30-min NCRV Series and 1-min Predicted IV Series
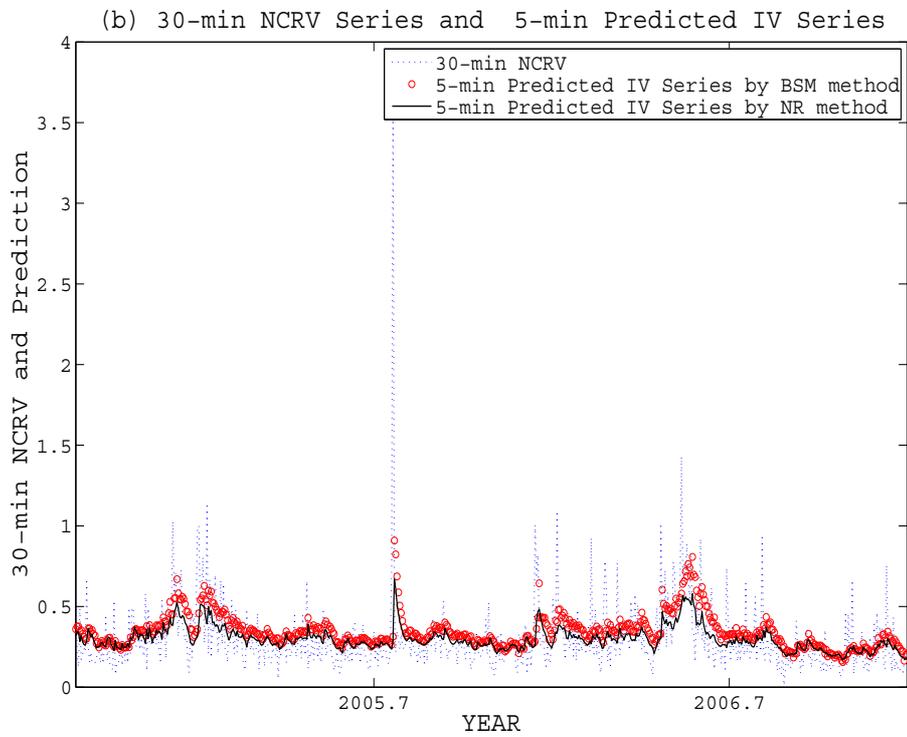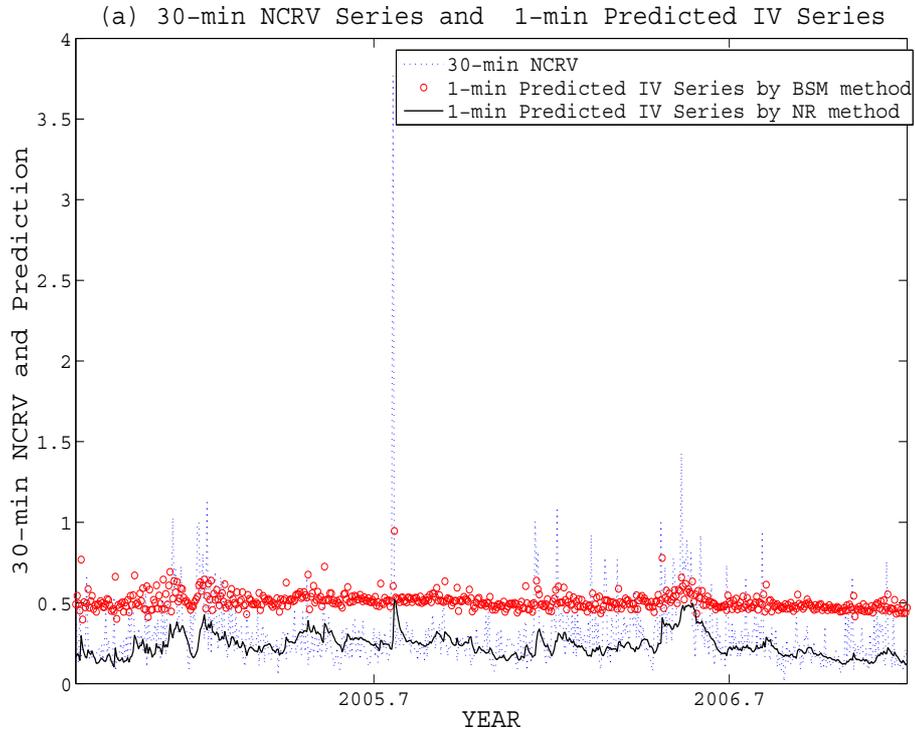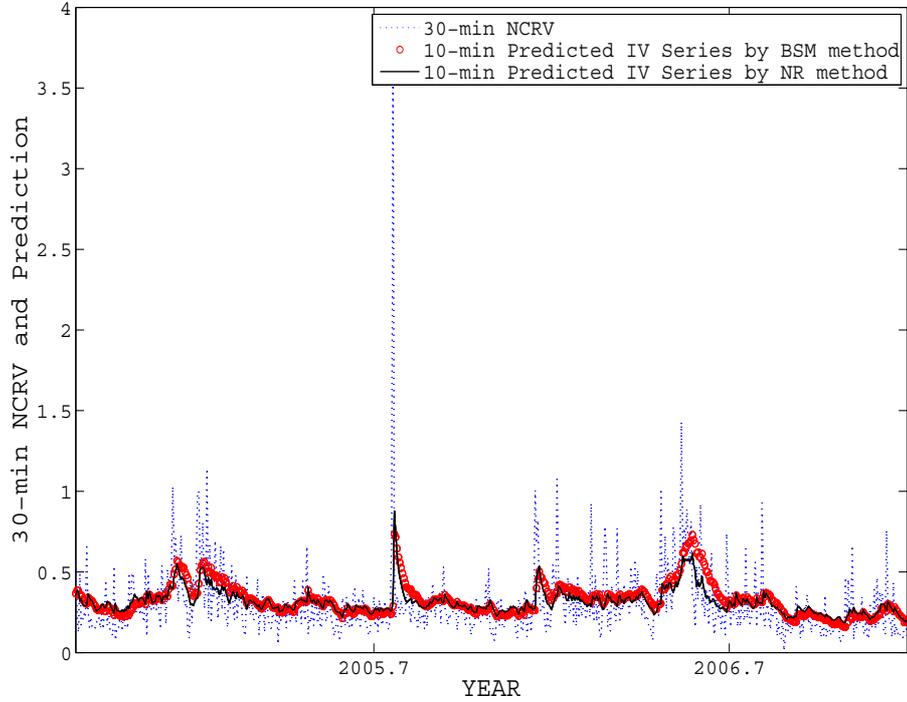

(b) 30-min NCRV Series and 5-min Predicted IV Series
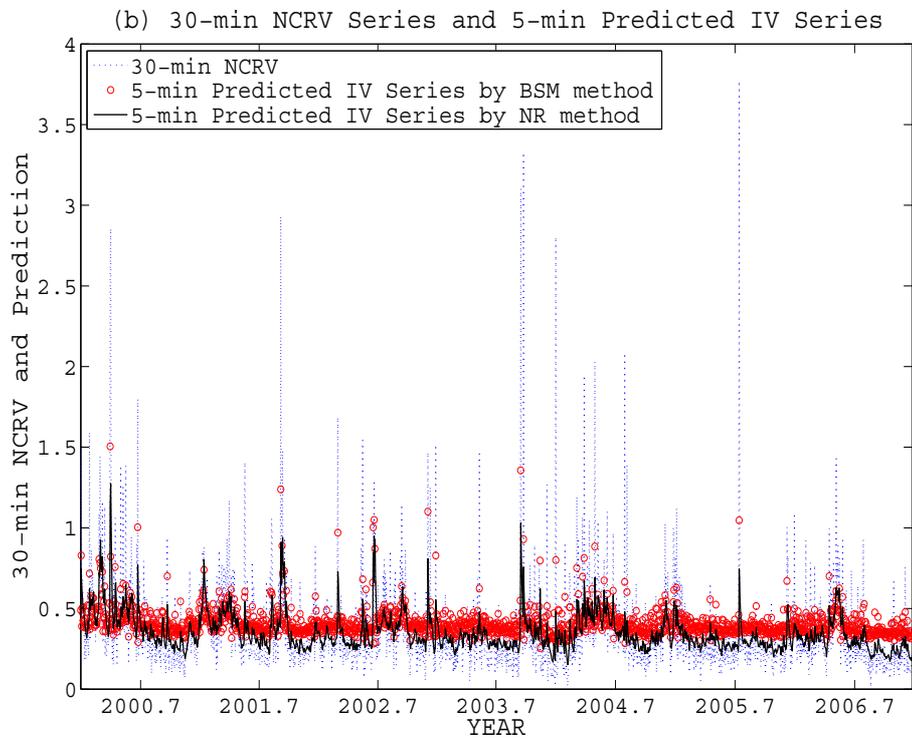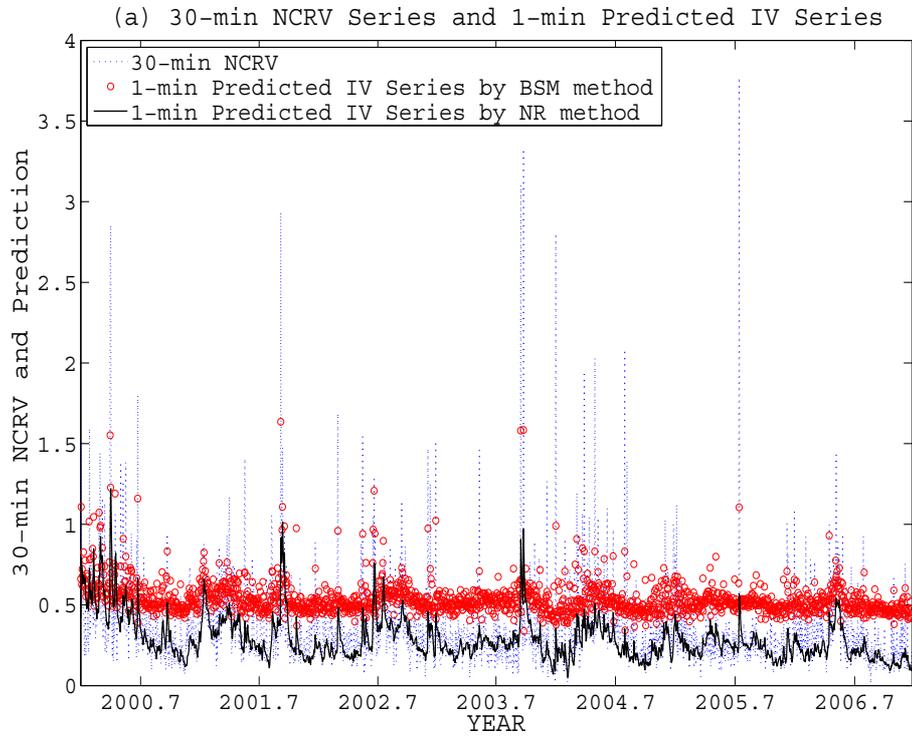
Figure 5: (continued) In Sample Forecasts (One-factor Model)

(c) 30-min NCRV Series and 10-min Predicted IV Series

(d) 30-min NCRV Series and 15-min Predicted IV Series
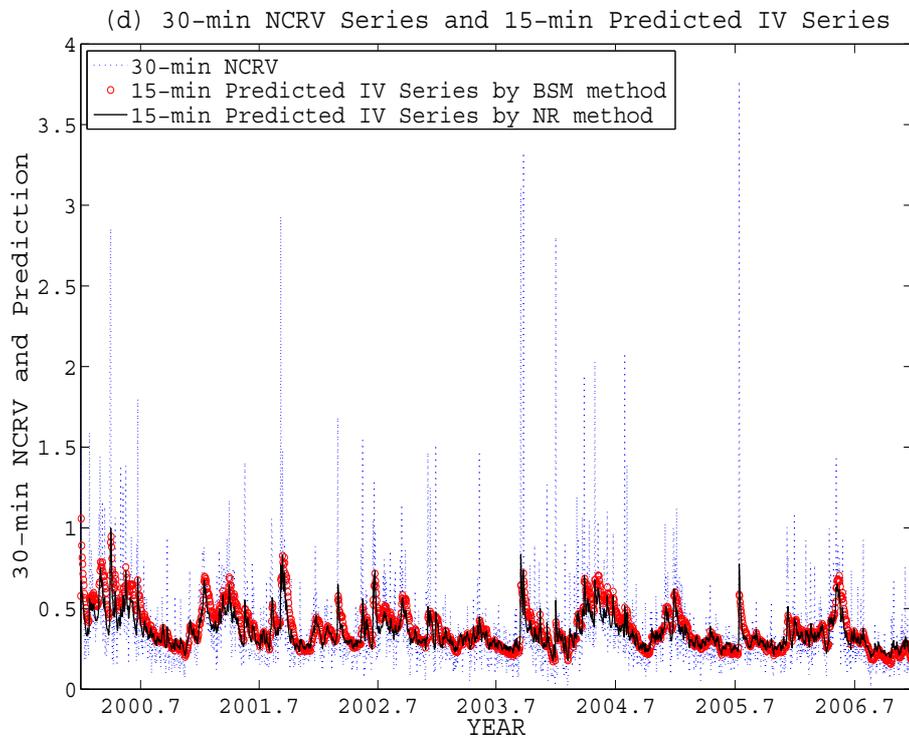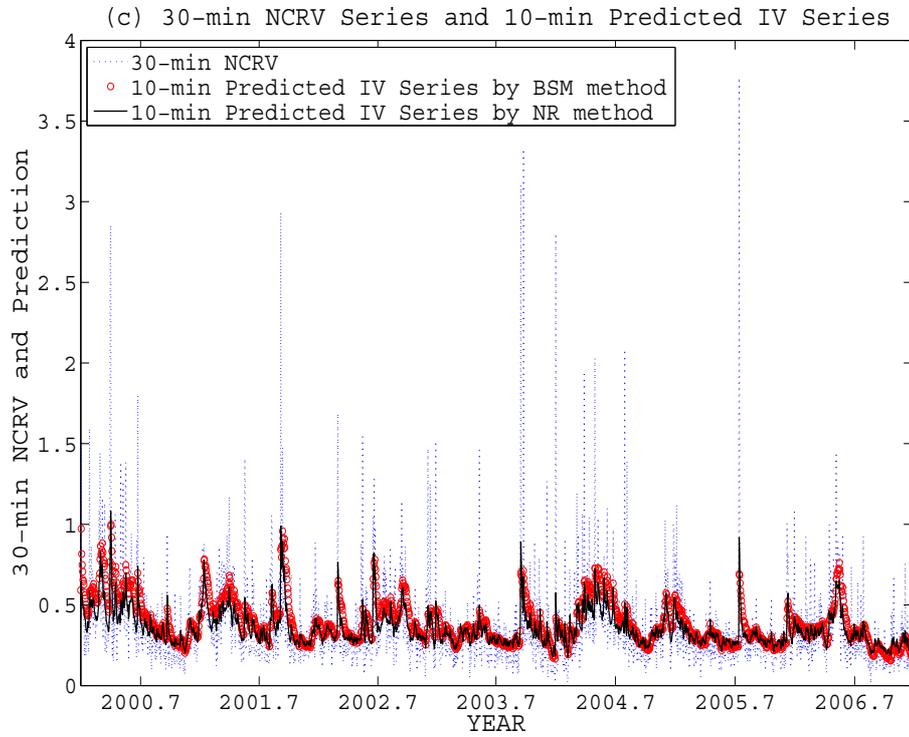
Figure 6: In Sample Forecasts (Two-factor Model)

(a) 30-min NCRV Series and 1-min Predicted IV Series

(b) 30-min NCRV Series and 5-min Predicted IV Series
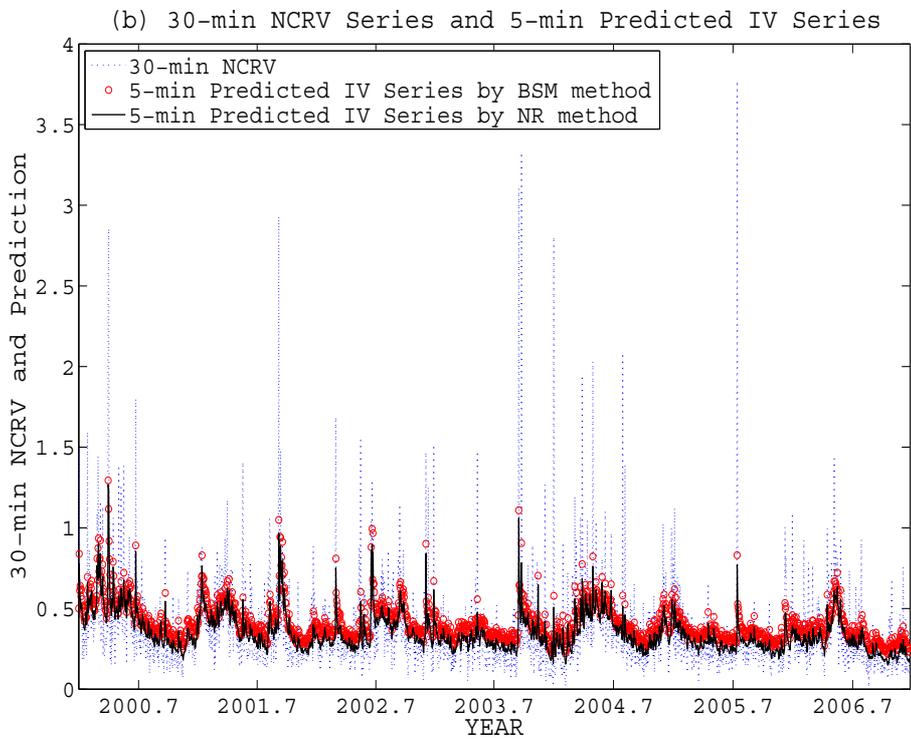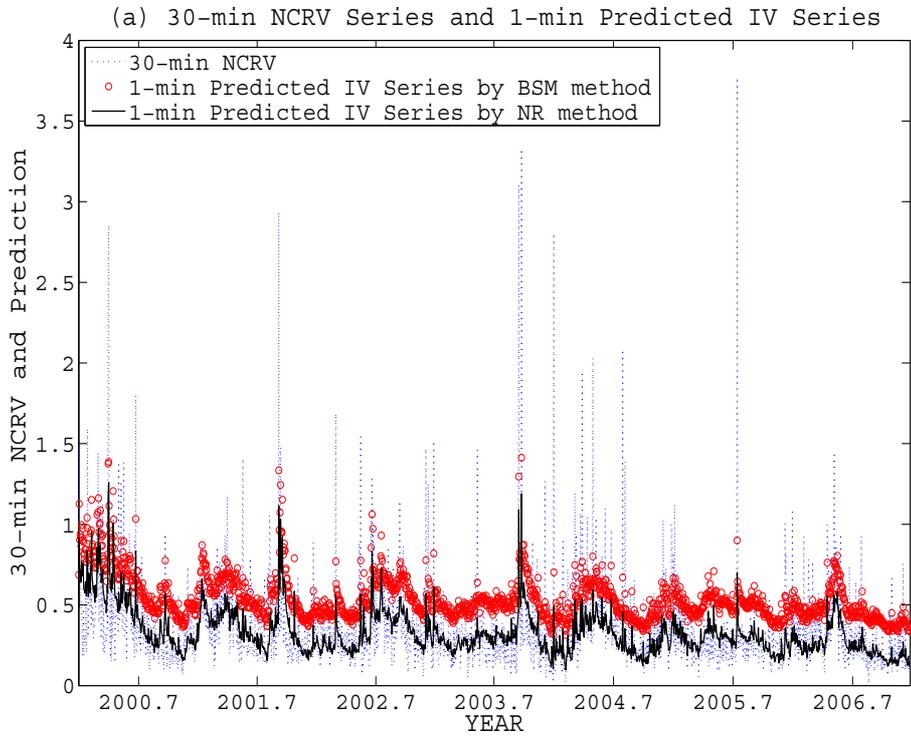
Figure 6: (continued) In Sample Forecasts (Two-factor Model)

(c) 30-min NCRV Series and 10-min Predicted IV Series
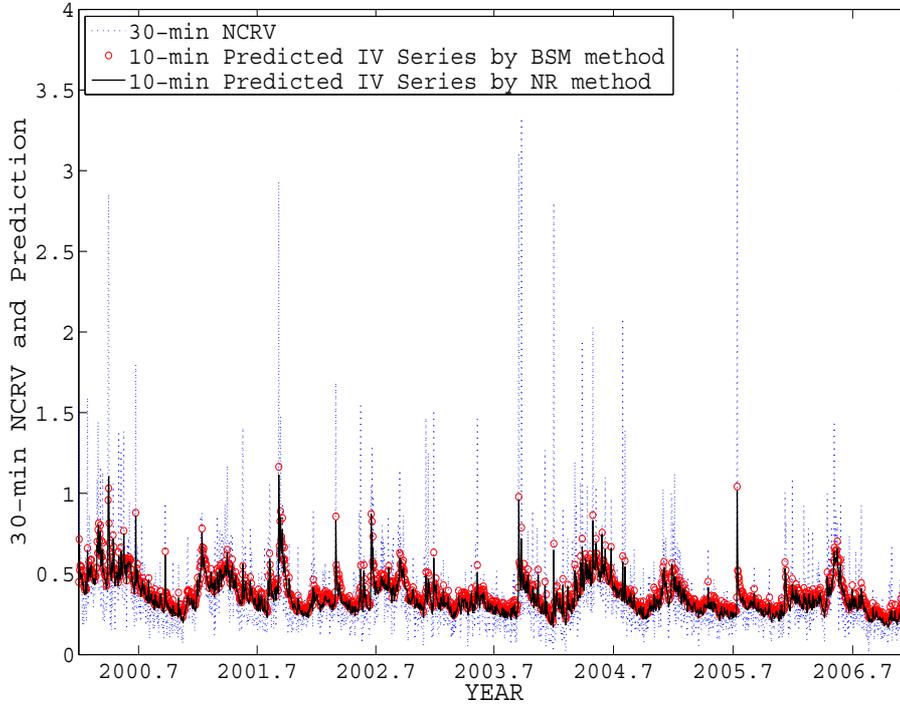
(d) 30-min NCRV Series and 15-min Predicted IV Series
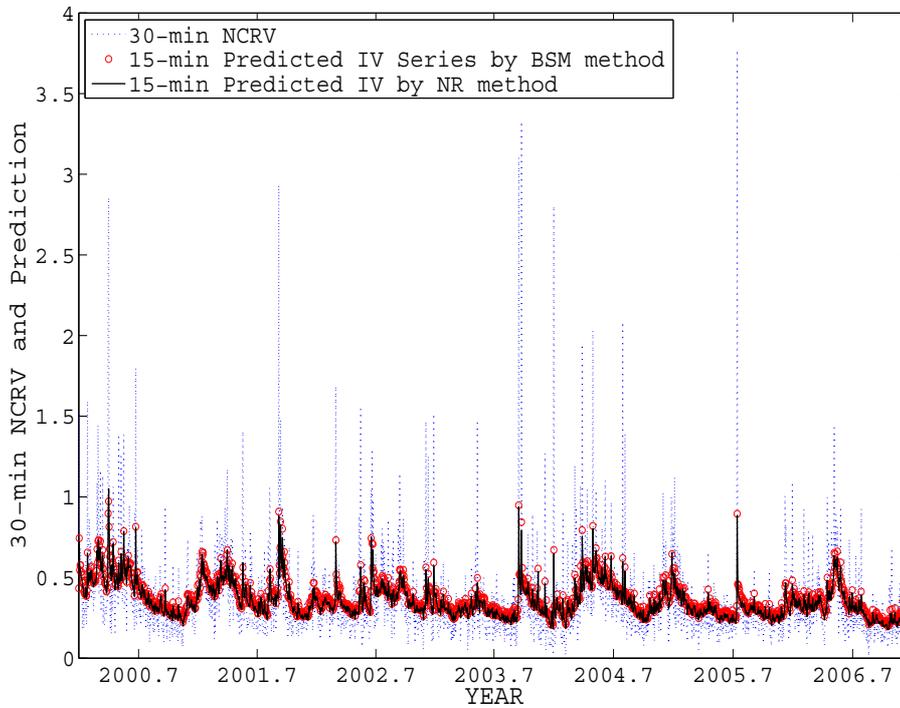
Figure 7: Out of Sample Forecasts (One-factor Model)

(a) 30-min NCRV Series and 1-min Predicted IV Series
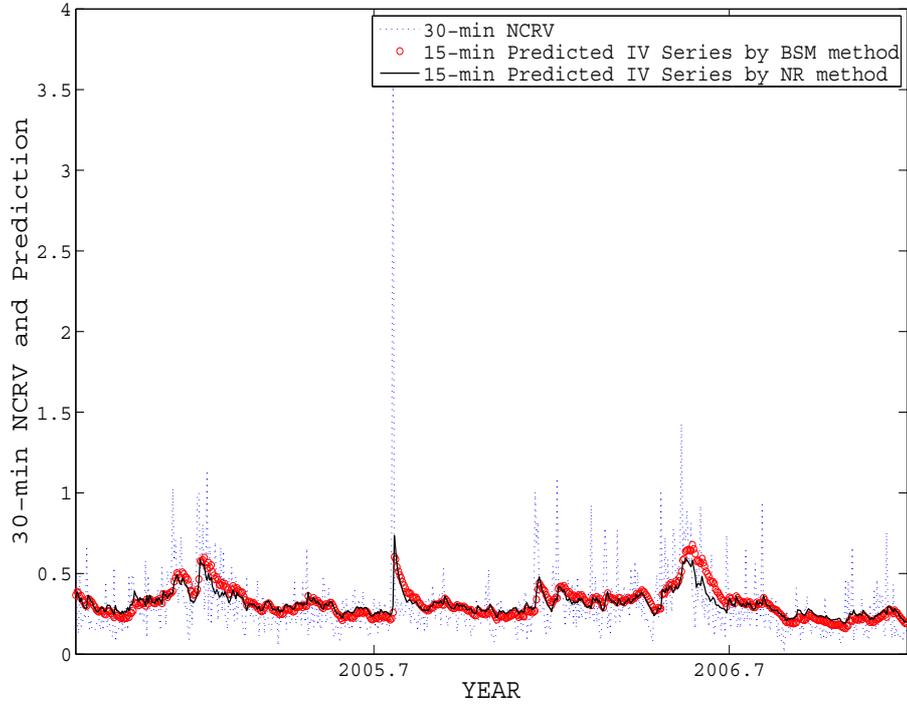

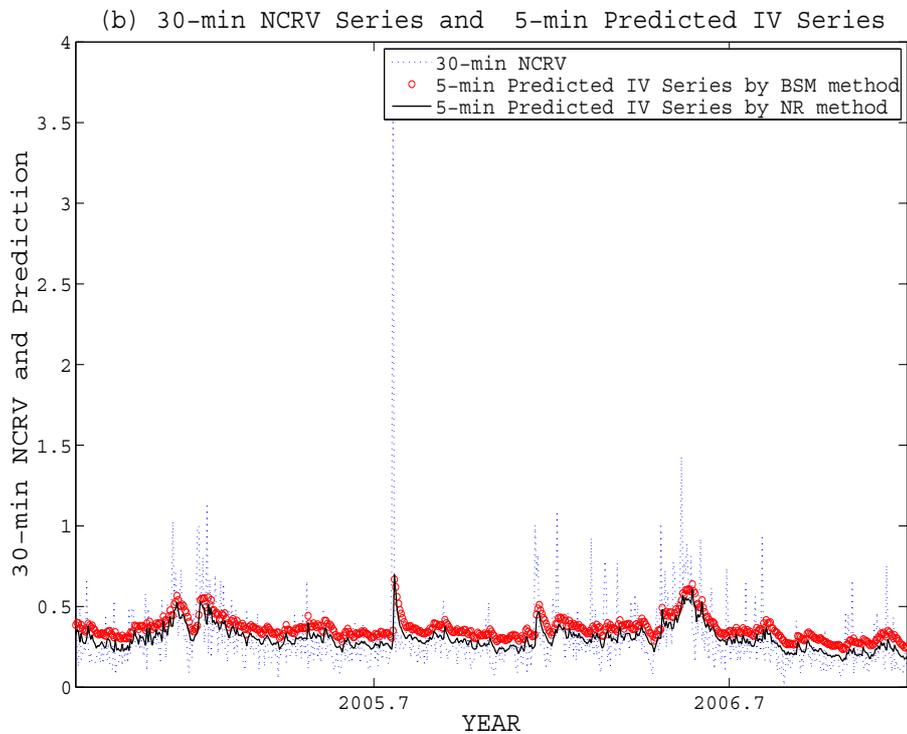(b) 30-min NCRV Series and 5-min Predicted IV Series

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

(c) 30-min NCRV Series and 10-min Predicted IV Series

(d) 30-min NCRV Series and 15-min Predicted IV Series

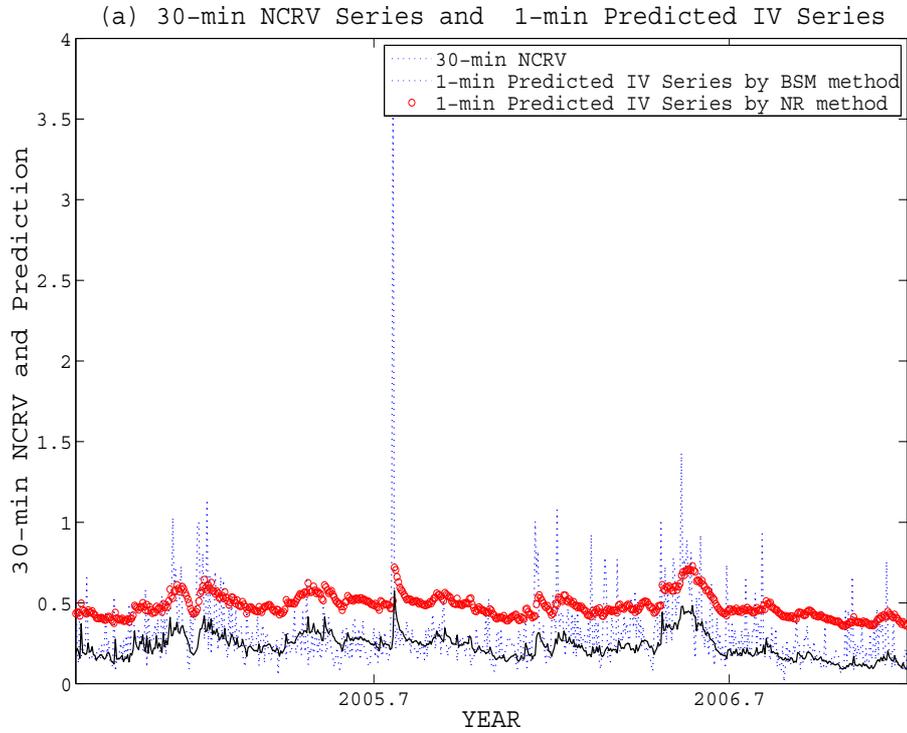Figure 8: Out of Sample Forecasts (Two-factor Model)

(a) 30-min NCRV Series and 1-min Predicted IV Series

(b) 30-min NCRV Series and 5-min Predicted IV Series
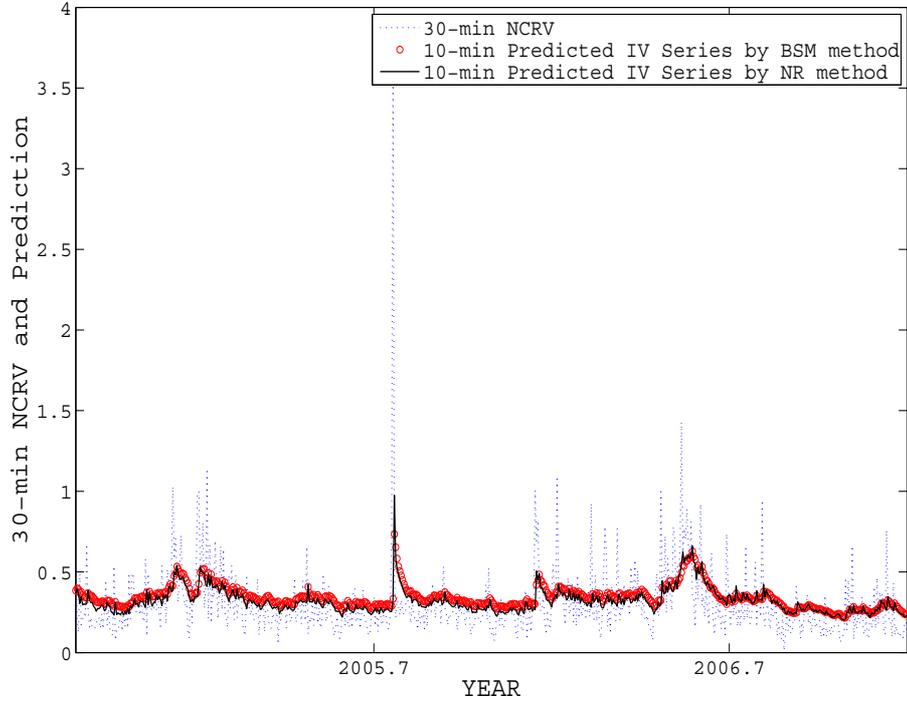

57

Figure 8: (continued) Out of Sample Forecasts (Two-factor Model)

(c) 30-min NCRV Series and 10-min Predicted IV Series

(d) 30-min NCRV Series and 15-min Predicted IV Series