Simple estimators for higher-order stochastic volatility models and forecasting *

Md. Nazmul Ahsan† Jean-Marie Dufour ‡
McGill University McGill University

November 5, 2018

*The authors thank Manabu Asai, Aman Ullah, Russell Davidson, M. Hashem Pesaran, René Garcia, Eric Renault, Lynda Khalaf, John Galbraith, Victoria Zinde-Walsh, Lars Stentoft, Denis Pelletier, Pascale Valéry, Firmin Doko Tchatoka, Masaya Takano, Purevdorj Tuvaandorj, Byunguk Kang, Jinjing Liu for useful comments and constructive discussions. Earlier versions of this paper were presented at Canadian Econometrics Study Group Meeting (Ottawa 2018), International conference on Time Series and Forecasting (Granada, 2018), Annual Conference of IAAE (Montreal, 2018), 52nd Annual Conference of the CEA (Montreal, 2018), CIREQ Econometrics Conference (Recent Advances in the Method of Moments, Montreal, 2018), University of Bergamo (Bergamo, 2018), University of Southern California (Los Angeles, 2018), Camp Econometrics (New York, 2018), York University (Toronto, 2018), CIREQ-McGill Lunch Seminar (Montreal, 2016, 2017), 11th World Congress of the Econometric Society (Montreal, 2015), and 11th CIREQ PhD Students' Conference (Montreal, 2015). This paper was previously circulating under the title "Simple estimators and inference for higher-order stochastic volatility models" and "Closed-Form Estimator for Higher-order Gaussian Stochastic Volatility Models".

This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† Ph.D. Candidate, Department of Economics, McGill University and Centre interuniversitaire de recherche en analyse des organisations (CIRANO). Mailing address: Department of Economics, McGill University, Leacock Building, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada; e-mail: md.ahsan@mail.mcgill.ca.

‡ William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: `jean-marie.dufour@mcgill.ca`. Web page: `http://www.jeanmariedufour.com`

ABSTRACT

We focus on higher-order stochastic volatility models [SV(p)] and propose several estimators: two simple estimators [moment-based and ARMA-based] and GMM estimators. The estimation and inference are challenging in SV models due to the inherent problem of evaluating the likelihood function – a general feature of most non-linear latent variable models. Most of the existing estimation methods are confined to estimate an SV(1) model and are computationally expensive, inflexible [not easy to generalize for SV(p) models], difficult to implement and additionally inefficient. The estimation of SV(p) models is even more challenging and rarely considered in the literature. Compared to the existing methods for SV(p) models, our simple estimators are computationally simple and very easy to implement. These estimators do not require choosing a sampling algorithm, initial parameter values, and an auxiliary model. Also, we suggest winsorized versions of the simple ARMA-based estimator to ensure the stationarity condition, especially in a small sample or in the presence of outliers. We develop recursive estimation procedures which allow for recursive-in-order calculation of the parameters of higher-order SV processes. We derive asymptotic theories for simple estimators and show the usefulness of these estimators in the context of simulation-based inference technique, i.e., Monte Carlo (MC) tests. By simulation, we compare our proposed estimators to the Bayesian MCMC estimator. The results show that the simple winsorized ARMA-based estimator is uniformly superior to other estimators in terms of bias and root mean square error. We present empirical applications related to SV(p) models and the ARMA-based estimator. First, using the daily return on the S&P 500 index from 1928 to 2016, we find that higher-order SV models may be preferable for in-sample model fitting and this result confirms by both asymptotic and finite-sample tests. Second, using different volatility proxies [the squared return of S&P 500 index and the realized volatility of S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices], we conduct two out-of-sample forecast experiments: (1) we forecast a moderately volatile period after the late-2000s Financial Crisis; (2) we forecast a highly volatile period, i.e., the core Financial Crisis. We compare the accuracy of volatility forecasts among SV(p) models, GARCH models, and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models. The results suggest that SV(p) models perform better than other models in most cases. This finding holds even if a high volatility period (such as Financial Crisis) is included in the estimation sample or the forecasted sample. Formal prediction tests, i.e., model confidence set procedure, also support these inferences. Our findings highlight the usefulness of higher-order SV models for volatility forecasting.

Key words: generalized method of moments, Markov Chain Monte Carlo, Monte Carlo tests, stochastic volatility, asymptotic distribution, stock returns, realized variance, volatility forecasting, high frequency data.

Journal of Economic Literature classification: C15, C22, C53, C58.

Contents

List of Tables

1. Introduction

Time-varying volatility is a widespread feature of financial markets. This property has been known for a long time; early comments include Mandelbrot (1963) and Fama (1965). Two main classes of parametric models have been proposed in the literature to model time-varying volatility: (1) GARCH-type models, where the conditional variance is a deterministic process [Engle (1982), Bollerslev (1986)]; (2) stochastic volatility (SV) models, where volatility is a latent stochastic process [Taylor (1982, 1986)]. The main distinction between GARCH and SV models is that the variance process of the latter has an additional error term which captures the effect of any new information coming to the market. Several reviews of GARCH and SV literature are available; see for GARCH [Bollerslev (2009) and for SV [Ghysels, Harvey and Renault (1996), Broto and Ruiz (2004), and Shephard (2005)]. SV models are also crucial in macroeconomics; see Cogley and Sargent (2005), Primiceri (2005), Benati (2008), Koop, Leon-Gonzalez and Strachan (2009), Koop and Korobilis (2013), and Liu and Morley (2014).

SV models may be preferable to GARCH-type models for several reasons. *First*, SV models are directly connected to diffusion processes used in theoretical finance and allow for a volatility process that does not depend on observable variables; see Shephard and Andersen (2009). In particular, SV models have simple continuous-time analogs which are used for option pricing; see Taylor (1994). *Second*, SV models are more robust to model misspecification than GARCH models. GARCH models often require adding a random jump component or allowing non-Gaussian innovations. These modifications improve the performance of the standard GARCH, but these are apparently unnecessary for SV models; see Carnero, Peña and Ruiz (2004), Chan and Grant (2016). *Third*, SV models often provide more accurate forecasts of volatility than those provided by GARCH models, indicating that the time-varying volatility is better modelled as a latent stochastic process; see Kim, Shephard and Chib (1998), Yu (2002), Poon and Granger (2003), Koopman, Jungbacker and Hol (2005). *Finally*, in comparison with GARCH models, SV models have a much simpler probabilistic structure: one can easily derive the stationarity, ergodicity and mixing properties of SV model than GARCH models; see Davis and Mikosch (2009), Lindner (2009).

Despite these attractive features, SV models are used less than the GARCH-type models. There seem to be two reasons for that. *First*, SV models have no closed-form likelihood function, so estimating the parameters of an SV model is much more complicated than it is for GARCH-type models. A single estimation method (for example, QML) is applicable for many variants of the GARCH model, whereas the variety (and potential incompatibility) of estimation methods for SV models [no estimation technique has turned out as clearly the simplest or most efficient]. *Second*, many statistical packages (such as EVIEWS, GAUSS, MATLAB, R, S+, SAS, TSP, STATA, PYTHON, OX, etc.) have many options for incorporating GARCH effects, whereas SV models lack statistical packages [some routines in R and MATLAB for SV models are available].

Several methods have been proposed to estimate the first-order SV model. These include: quasi-maximum likelihood (QML) [Nelson (1988), Harvey, Ruiz and Shephard (1994), Ruiz (1994)], the generalized method of moments (GMM) [Melino and Turnbull (1990), Andersen and Sørensen (1996)], the simulated method of moments (SMM) [Gallant and Tauchen (1996), Monfardini (1998), Andersen, Chung and Sørensen (1999)], Monte Carlo likelihood (MCL) [Sandmann and Koopman (1998)], simulated maximum likelihood (SML) [Danielsson and Richard (1993), Danielsson (1994), Durham (2006, 2007), Richard and Zhang (2007)], the method based on linear representation [Francq and Zakoïan (2006)], a closed-form moment-based estimator (DV) [Dufour and Valéry (2006, 2009)], and Bayesian techniques based on Markov Chain Monte Carlo (MCMC) methods [Jacquier, Polson and Rossi (1994), Kim et al. (1998), Chib, Nardari and Shephard (2002), Flury and Shephard (2011)].

Except for the closed-form estimator of Dufour and Valéry (2006, 2009), these estimation methods are based on simulation and numerical optimization procedures. Simulation-based methods such as SML, MCL, SMM, and Bayesian MCMC method (based on Metropolis-Hastings algorithm or Gibbs sampler) are costly from a computational viewpoint, inflexible across models, not easy to implement, and converge very slowly [see Broto and Ruiz (2004)]. Furthermore, some of these methods require choosing a sampling algorithm, initial parameters, and an arbitrary choice of auxiliary model. Other methods based on numerical optimization [such as QML or GMM] require choosing initial parameter values. Broto and Ruiz (2004) pointed out that the GMM criterion surface of the SV model is highly irregular, so optimization often fails to converge, especially with small samples. A large number of non-converging GMM estimations has been reported by Andersen and Sørensen (1996), which is consistent with our simulation findings. Further, GMM requires choosing and estimating a weighting matrix, which can easily be ill-conditioned. QML may produce imprecise estimates due to an inefficient implementation of the procedure (poor starting values, different convergence criteria, etc.).

By contrast, the closed-form estimator of Dufour and Valéry (2006, 2009) is computationally simple, and very easy to implement. However, it tends to be less precise than some other estimators. In Ahsan and Dufour (2015), we propose a simple closed-form estimator for the SV(1) model by exploiting its non-Gaussian ARMA representation.[1] The ARMA-based estimator not only has the advantages of Dufour and Valéry (2006) but it also solves the efficiency problem. The estimation of higher-order stochastic volatility [SV(p)] models is even more challenging than an SV(1) model. In an SV(p) model the latent volatility process follows an autoregressive process of order p. In this paper, we extend the methods of Dufour and Valéry (2006, 2009) and Ahsan and Dufour (2015) to propose simple estimators for SV(p) models. Further, we also suggest GMM-type estimators for SV(p) models. These GMM estimators are an extension of Andersen and Sørensen (1996).

In financial econometrics literature, SV(p) models are rarely estimated, exceptions are Gallant, Hsieh and Tauchen (1997), Asai (2008) and Chan and Grant (2016). In line with these studies, the motivation for SV(p) models can be described as follows. *First*, it is a natural extension of the basic SV model. *Second*, as pointed out by Asai (2008) and Meddahi (2003) the multi-factor stochastic volatility (MFSV) model can be interpreted as a linear combination of latent and independent AR(1) processes that aggregate to an ARMA(p,q) process. Since the latent volatility process of an MFSV model can be written as an ARMA process, the higher-order autoregressive terms in SV models are reasonable. *Third*, the empirical results of these studies suggest that higher-order models provide more flexibility to represent volatility persistence, heavy tails and may capture the effects of jumps as well. *Fourth*, empirical evidence in this paper suggests that higher-order SV models may be preferable for both in-sample model fitting and out-of-sample volatility forecasting.

The proposed simple estimators are analytically tractable, computationally simple, and very easy to implement. These estimators can readily be implemented without using numerical optimization methods, and these estimators do not require choosing arbitrary initial values for the parameters or auxiliary model. One limitation of the ARMA-based estimator is that it may give you the solution outside the admissible area (violates the stationarity), particularly in a small sample or in the presence of outliers. To avoid these problems, we suggest winsorized versions of the ARMA-based estimator to ensure the stationarity condition. In Section 9, we thoroughly check the robustness of these winsorized estimators.

We use these simple estimators to develop recursive estimation procedures for SV(p) models by exploiting Durbin-Levinson (DL) type algorithms. We discuss two algorithms which allow for recursive-in-order calculation of the parameters of higher-order SV processes. The proposed procedures generalize the recursion of Durbin (1960) [which applies to pure autoregressive models] and the recursion of Tsay and Tiao (1984) [which applies to autoregressive-moving average models]. These types of recursive estimations are

[1] In the context of continuous time stochastic volatility models, Meddahi (2003) derives the ARMA representation of integrated and realized variances when the spot variance depends linearly on two autoregressive factors.

improbable with other estimators.

The proposed simple estimators are useful in the context of simulation-based inference procedures. The fact that SV models are parametric models involving only a finite number of unknown parameters, so these computationally inexpensive estimators can be easily exploited within simulation-based inference procedures as opposed to procedures based on establishing asymptotic distributions. For that purpose, one can obtain finite-sample inference based on Monte Carlo (MC) tests; see Dufour (2006). When the distribution of a test statistic does not depend on (unknown) nuisance parameters, the technique of MC tests yields an exact test provided one can generate a few i.i.d. (or exchangeable) replications of the test statistic under the null hypothesis; for example, 19 replications are sufficient to get a test with level 0.05; see Dufour and Khalaf (2001). The MC technique can be extended to test statistics which depend on nuisance parameters by considering maximized Monte Carlo (MMC) tests; see Dufour (2006).

We derive the asymptotic properties of the proposed simple estimators under standard regularity assumptions, showing consistency and asymptotic normality when the fourth moment of latent volatility process exists. The simple estimators can be effortlessly applied to very large samples since these are $\sqrt{T}$-consistent. The estimators based on simulation or numerical optimization methods often require substantial computational effort to achieve convergence in those situations, which are not rare in empirical studies. So instead of using computationally costly estimators, one may prefer to use estimators that are available in analytical form.

We then study the statistical properties of our estimators by simulation and compare them with the Bayesian MCMC method. The simulation results confirm that the simple winsorized ARMA-based estimator has excellent statistical properties in terms of bias and root mean square error. It uniformly outperform all other estimators, including the Bayesian estimator regarding bias and RMSE. This result holds across different simulation designs and for all individual parameters. Furthermore, the simple estimators are highly time efficient compared to other estimators.

We present empirical applications related to SV(p) models and the ARMA-based estimator. *First*, using the daily return on the S&P 500 index from 1928 to 2016, we find that an SV(3) model may be preferable for in-sample model fitting and this result confirms by both asymptotic and finite-sample tests. *Second*, using different volatility proxies [the squared return of S&P 500 index and the realized volatility of S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices], we conduct two out-of-sample forecast experiments: (1) we forecast a moderately volatile period after the late-2000s Financial Crisis; (2) we forecast a highly volatile period, i.e., the core Financial Crisis. We compare the accuracy of volatility forecasts among SV(p) models, GARCH models, and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models. The results suggest that SV(p) models perform better than other models in most cases. This finding holds even if a high volatility period (such as Financial Crisis) is included in the estimation sample or the forecasted sample. These inferences are not only based on a standard forecasting precision assessment [such as using MSE and MAE statistics] but also based on formal prediction tests, using the MCS procedure of Hansen, Lunde and Nason (2011). Our findings highlight the usefulness of higher-order SV models for volatility forecasting.

The simple moment-based estimators proposed in this paper can be interpreted *parsimonious moment-based* estimators where only a small number of (well chosen) moments are used. In moment-based (or GMM) inference, using too many moments can be very costly from an estimation efficiency viewpoint as well as forecasting. Indeed, we show in our simulations and empirical applications that the proposed ARMA-based estimators exhibit the best performance in both estimation and forecasting, as well as numerical efficiency.

The rest of the paper is organized as follows. Section 2 specifies models and assumptions and Section 3 discusses the motivation for SV(p) models. Section 4 discusses the stationarity, ergodicity and mixing properties of SV(p) models. Section 5 proposes simple estimators and their recursive prediction algorithms. Section 6 proposes GMM type estimators for SV(p) models, and Section 8 discusses the MC test technique. Section 7 develops asymptotic theories for simple estimators. Section 9 presents the simulation study, and Section 10 presents the empirical applications. Section 11 offers conclusions. Proofs, tables, and figures are reported in Appendix.

2. Framework

We consider a standard discrete-time SV process of order p, which is described below following Taylor (1986), Ghysels et al. (1996). Specifically, we say that a variable y_t follows a discrete-time SV(p) process if it satisfies the following assumption, where $t \in \mathbb{N}_0$ and $\mathbb{N}_0$ represents the non-negative integers.

Assumption 2.1 STOCHASTIC VOLATILITY OF ORDER p. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *satisfies the equations*

$$y_t = \sigma_y \exp(\frac{w_t}{2}) z_t \,, \tag{2.1}$$

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t \,, \tag{2.2}$$

where the vectors $(z_t, v_t)'$ *are i.i.d. according to a* $\mathrm{N}[0, I_2]$ *distribution, while* $\phi_1, \ldots, \phi_p$, σ_y, σ_v *are fixed parameters.*

We also make a stationarity assumption as follows.

Assumption 2.2 STATIONARITY. *The process* $l_t = (y_t, w_t)'$ *is strictly stationary.*

The above stationarity condition entails all the roots of the characteristic equation of the volatility process $[\phi(B) = 0]$ are lie outside the unit circle, i.e., $(1 - \phi_1 B - \cdots - \phi_p B^p) = 0 \Leftrightarrow |B_i| > 1$ for $i = 1, \ldots, p$ and $w_0 \sim N[0, \sigma_v^2/(1 - \sum_{j=1}^{p} \phi_j^2)]$.

The SV(p) model consists of two stochastic processes which describe the dynamics of y_t and the latent log volatilities w_t. When y_t is an asset return, the latent process w_t in (2.2) can be interpreted as a random flow of uncertainty shocks or new information in financial markets, while ϕ_j's capture volatility persistence. This type of volatility models naturally fits into the theoretical framework of modern financial theory.

Let us now transform y_t by taking the logarithm of its squared value. We get in this way the following *measurement equation*:

$$\begin{aligned} \log(y_t^2) &= \log(\sigma_y^2) + w_t + \log(z_t^2) = \{\log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)]\} + w_t + \{\log(z_t^2) - \mathbb{E}[\log(z_t^2)]\} \\ &= \mu + w_t + \varepsilon_t \end{aligned} \tag{2.3}$$

where

$$\mu := \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)] \,, \quad \varepsilon_t := \log(z_t^2) - \mathbb{E}[\log(z_t^2)] \,. \tag{2.4}$$

Note that this logarithmic transformation entails no information loss since the distribution of z_t is symmetric (see Remark 1 of Francq and Zakoïan (2006)). Furthermore, even if v_t and z_t are not mutually independent, they are uncorrelated if the joint distribution of v_t and z_t is symmetric, that is $f(v_t, z_t) = f(-v_t, -z_t)$; see Harvey et al. (1994).

Under the normality assumption for z_t, the errors ε_t are i.i.d. according to the distribution of a centered $\log(\chi_1^2)$ random variable [*i.e.*, ε_t has mean zero and variance $\mathbb{E}(\varepsilon_t^2)$]. The moment generating function of

$\log(\chi_1^2)$ distribution is

$$\begin{aligned} M(s) &= \log \mathbb{E}\left[\exp\left(s\log(\chi_1^2)\right)\right] = \log\left[\mathbb{E}\left(\chi_1^2\right)^s\right] = \log\left[\frac{2^s\Gamma\left((1/2)+s\right)}{\Gamma(1/2)}\right] \\ &= s\log(2)+\log[\Gamma\left((1/2)+s\right)]-\log[\Gamma(1/2)], \quad s\in\mathbb{R}, \end{aligned} \tag{2.5}$$

where $\Gamma(z) := \int_0^\infty x^{z-1}e^{-x}\,dx$ is the *gamma function*; see Wishart (1947). The m^{th} moment of the $\log(\chi_1^2)$ random variable is the m^{th} derivative of $M(s)$ evaluated at $s=0$, and the corresponding central moments are:

$$\mu_m = \begin{cases} \log(2)+\psi(\frac{1}{2}), & \text{if } m=1, \\ \psi^{(m-1)}(\frac{1}{2}), & \text{if } m>1, \end{cases} \tag{2.6}$$

where

$$\psi(z) := \frac{d}{dz}\ln\left[\Gamma(z)\right] = \frac{\Gamma'(z)}{\Gamma(z)} \tag{2.7}$$

is the *digamma function* and

$$\psi^{(m)}(z) := \frac{d^m}{dz^m}\psi(z) = \frac{d^{m+1}}{dz^{m+1}}\ln[\Gamma(z)] \tag{2.8}$$

is the polygamma function of order m [*i.e.*, the $(m+1)$-th order derivative of the logarithm of the gamma function].

Form (2.6), we get:

$$\mathbb{E}[\log(z_t^2)] = \log(2)+\psi(\frac{1}{2}) \simeq -1.27\,, \tag{2.9}$$

$$\sigma_\varepsilon^2 := \mathbb{E}(\varepsilon_t^2) = \text{Var}[\log(z_t^2)] = \psi^{(1)}(\frac{1}{2}) = \pi^2/2\,, \tag{2.10}$$

$$\mathbb{E}(\varepsilon_t^3) = \psi^{(2)}(\frac{1}{2})\,, \quad \mathbb{E}(\varepsilon_t^4) = \psi^{(3)}(\frac{1}{2}) = \pi^4\,; \tag{2.11}$$

see Abramowitz and Stegun (1970, Chapter 6). The $\log(\chi_1^2)$ distribution is often approximated by a normal distribution with mean of -1.27 and variance of $\pi^2/2$; see Broto and Ruiz (2004).

On setting

$$y_t^* := \log(y_t^2) - \mu\,, \tag{2.12}$$

the SV model (2.3) can be written as

$$y_t^* = w_t + \varepsilon_t\,. \tag{2.13}$$

By combining (2.2) and (2.13), we see that the SV model can be written in state-space form:

$$w_t = \sum_{j=1}^{p}\phi_j w_{t-j} + v_t\,, \qquad \text{(State Transition Equation)} \tag{2.14}$$

$$y_t^* = w_t + \varepsilon_t\,, \qquad \text{(Measurement Equation)} \tag{2.15}$$

where w_t is a logarithm of latent daily volatility, y_t^* is a logarithm of the daily squared return corrected by its mean, where the variables v_t are i.i.d. $\text{N}(0,\sigma_v^2)$ and the ε_t's are i.i.d. $\log(\chi_1^2)$; for further discussion of this representation, see Nelson (1988), Harvey et al. (1994), Ruiz (1994), Shephard (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim et al. (1998), Sandmann and Koopman (1998), Steel (1998), Chib et al. (2002), Knight, Satchell and Yu (2002), Francq and Zakoïan (2006), Omori, Chib, Shephard and

Nakajima (2007).

In this study, we propose simple estimator for an SV(p) model based on the ARMA representation of the state-space model that defined in (2.14) and (2.15). Furthermore, we also extend the method of Dufour and Valéry (2006) in the context of SV(p) process.

3. Higher-order stochastic volatility

In this section, we discuss the econometric motivation for SV(p) models. It has been well documented that the volatility process is driven by at least two factors: one factor captures the salient properties of volatility such as randomness and persistence and a second one to deal with the shape of the conditional distribution of financial returns such as fat-tails; examples of these studies include Gallant, Hsu and Tauchen (1999), Meddahi (2001), Alizadeh, Brandt and Diebold (2002), Barndorff-Nielsen, Nicolato and Shephard (2002), Bollerslev and Zhou (2002), Chernov, Gallant, Ghysels and Tauchen (2003), and Durham (2006, 2007). Many of these studies also considered more than two factors and tried to fit the volatility process of asset returns. These types of factor models are important for capturing non-linearities in financial returns and improves the fit of data dramatically. However, these models need highly complex numerical optimization techniques, and they are not tractable analytically. Note that we can always transform the MFSV models to an SV model that has an ARMA representation in the log volatility process (SV-ARMA). This transformation is perfectly acceptable since we can recuperate the MFSV parameters form the estimates of SV-ARMA parameters. Further instead of an SV-ARMA model, we can estimate an SV(p) model and recuperate the SV-ARMA parameters from there.

Assumption 3.1 MULTI-FACTOR STOCHASTIC VOLATILITY MODEL. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *follows an MFSV*(m) *model of the type:*

$$y_t = \sigma_y \exp(\sum_{i=1}^{m} \frac{w_{it}}{2}) z_t,$$

$$w_{it} = \phi_{if} w_{it-1} + \sigma_{iv} v_{it}, \quad |\phi_{if}| < 1, \quad i = 1, \ldots, m,$$

where $\Theta_m^{MFSV} := (\sigma_y, \{\phi_{if}\}_{i=1}^m, \{\sigma_{iv}\}_{i=1}^m)$ *are fixed parameters and* (z_t, v_{it}) *are i.i.d. Gaussian such that* z_t *is N(0, 1) and* v_{it}*'s are* $N(0, I_m)$ *and* $\mathbb{E}[v_{it} z_t] = 0 \; \forall i$.

Lemma 3.1 SV-ARMA REPRESENTATION OF MFSV MODEL. *The model MFSV*(m) *defined by assumption 3.1 has the following SV-ARMA*$(m, m-1)$ *representation:*

$$y_t = \sigma_y \exp(\frac{w_t}{2}) z_t, \tag{3.1}$$

$$w_t = \sum_{j=1}^{m} \alpha_j w_{t-j} + \sigma_v v_t - \sigma_v \sum_{j=1}^{m-1} \beta_j v_{t-j}, \tag{3.2}$$

where $\beta_{m,m-1}^{SV-ARMA} := (\sigma_y, \{\alpha_j\}_{j=1}^m, \{\beta_j\}_{j=1}^{m-1}, \sigma_v)$ *are fixed parameters and* $(z_t, v_t)'$, $t \in \mathbb{N}_0$, *are i.i.d. according to a* $N[0, I_2]$ *distribution. Further, all the roots of characteristic equations* $[1 - \alpha_1 B - \cdots - \alpha_m B^m = 0$ *and* $1 - \beta_1 B - \cdots - \beta_{m-1} B^{m-1} = 0]$ *are lie outside the unit circle.*

To understand the Lemma 3.1, we illustrate the following example.

Example 3.1 THE SV-ARMA(2,1) REPRESENTATION OF THE MFSV(2) MODEL. Under the assumption 3.1, we have an MFSV(2) model where the volatility process is driven by the sum of two independent AR(1) process such that

$$w_{1t} - \phi_{1f} w_{1t-1} = (1-\phi_{1f}L) w_{1t} = \sigma_{1v} v_{1t}$$
$$w_{2t} - \phi_{2f} w_{2t-1} = (1-\phi_{2f}L) w_{2t} = \sigma_{2v} v_{2t}$$

Using the aggregation principle of autoregressive process [see Granger and Morris (1976)], i.e., $AR(p) + AR(q) = ARMA(p+q, max(p,q))$ and in particular if we add two independent AR(1) processes then we can obtain an ARMA(2,1) process. This could be view from following:

As $w_t = w_{1t} + w_{2t}$ and v_{1t} and v_{2t} are two independent white noise, it follows that

$$(1-\phi_{1f}L)(1-\phi_{2f}L)w_t = (1-\phi_{2f}L)\sigma_{1v}v_{1t} + (1-\phi_{1f}L)\sigma_{2v}v_{2t},$$

or,

$$(1-\alpha_1 L - \alpha_2 L^2)w_t = (1-\beta_1 L)\sigma_v v_t.$$

This is an ARMA(2,1) process with AR parameters, $\alpha_1 = \phi_{1f} + \phi_{2f}$ and $\alpha_2 = -\phi_{1f}\phi_{2f}$. Solving will yields

$$\phi_{1f} = \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2}}{2}, \quad \phi_{2f} = \frac{\alpha_1 + \sqrt{\alpha_1^2 + 4\alpha_2}}{2}$$

and R.H.S. is an invertible $MA(1)$ process follows from $MA(p) + MA(q) = MA(max(p,q))$ with variance is $[(1+\phi_{2f}^2)\sigma_{1v}^2 + (1+\phi_{1f}^2)\sigma_{2v}^2]$ and covariance at lag 1 is $[-(\phi_{2f}\sigma_{1v}^2 + \phi_{1f}\sigma_{2v}^2)]$. Thus we can have an SV-ARMA(2,1) with $\alpha_1 = \phi_{1f} + \phi_{2f}$, $\alpha_2 = -\phi_{1f}\phi_{2f}$, $[(1+\phi_{2f}^2)\sigma_{1v}^2 + (1+\phi_{1f}^2)\sigma_{2v}^2] = (1+\beta_1^2)\sigma_v^2$, and $[-(\phi_{2f}\sigma_{1v}^2 + \phi_{1f}\sigma_{2v}^2)] = -\beta_1\sigma_v^2$.

SV-ARMA models are more parsimonious but difficult to estimate whereas SV(p) models are not parsimonious but easy to estimate and use. We can estimate an SV(p) model instead of an SV-ARMA model and recuperate the parameters of SV-ARMA model from the estimates of SV(p) parameters. This process is based on the AR approximation of ARMA-type latent volatility process. Some estimators, based on autoregressive approximation, have been proposed for general ARMA models. These methods derive estimated coefficients from an approximating autoregressive process, and use a linear regression or other technique to extract information from the full set of autoregressive coefficients; see for example Hannan and Rissanen (1982), Saikkonen (1986), Koreisha and Pukkila (1990) and Galbraith and Zinde-Walsh (1997). In the context of VARMA models, this type of method used by Dufour and Pelletier (2005) and Dufour and Jouini (2014).

Lemma 3.2 INFINITE-ORDER SV REPRESENTATION OF SV-ARMA PROCESS. *The model SV-ARMA(p,q) defined by lemma 3.1 [where $p = m$ and $q = m-1$] has the following SV(∞) representation:*

$$y_t = \sigma_y \exp(\frac{w_t}{2}) z_t, \tag{3.3}$$

$$\sum_{j=0}^{\infty} (-\phi_j) w_{t-j} = \sigma_v v_t, \tag{3.4}$$

where $\Theta_{\infty}^{SV} := (\sigma_y, \{\phi_j\}_{j=1}^{\infty}, \sigma_v)$ are fixed parameters and $(z_t, v_t)'$, $t \in \mathbb{N}_0$, are i.i.d. according to a $N[0, I_2]$ distribution.

The SV(∞) model, given in Lemma 3.2, can be replaced by a truncated SV(p) model and we can recuperate the SV-ARMA(p,q) parameters from it. Using standard results on representation of an ARMA(p,q) process [see Fuller (1996), Ch. 2, page 74], we have following expressions that relate the parameters of SV(∞) and SV-ARMA(p,q) model:

$$\sum_{j=0}^{\infty}(-\phi_j)w_{t-j} = \sigma_v v_t, \tag{3.5}$$

where

$$\begin{aligned}
&\phi_0 = -1,\\
&\phi_1 = -\beta_1 + \alpha_1,\\
&\phi_2 = -\beta_1\phi_1 + \beta_2 + \alpha_2,\\
&\vdots\\
&\phi_j = -\sum_{i=1}^{min(j,q)} \beta_i\phi_{j-i} + \alpha_j, \quad (j \leq p)\\
&\phi_l = -\sum_{i=1}^{min(l,q)} \beta_i\phi_{l-i}, \quad (l > p).
\end{aligned}$$

Given the above equations, we can identify the parameters of an SV-ARMA(p,q) model from the parameters of an SV(k) model. The identification requires $k \geq p+q$. To understand the whole identification process, we illustrate the following example.

Example 3.2 SV-ARMA(2,1) FROM SV(3) MODEL. Under the lemma 3.1, we have SV-ARMA$(2,1)$ model where the volatility process is driven by an ARMA(2,1). To identify an SV-ARMA$(2,1)$ model from an SV(3) model, we use following equations:

$$\phi_1 = -\beta_1 + \alpha_1, \quad \phi_2 = -\beta_1\phi_1 + \alpha_2, \quad \text{and} \quad \phi_3 = -\beta_1\phi_2.$$

Solving the above equations yields the parameters of SV-ARMA(2,1) model with $\alpha_1 = \phi_1 + \frac{\phi_3}{\phi_2}$, $\alpha_2 = \phi_2 + \frac{\phi_1\phi_3}{\phi_2}$, and $\beta_1 = -\frac{\phi_3}{\phi_2}$ in terms of SV(3) parameter. The whole identification process:

1. The MFSV(2) model, where the volatility process is driven by two independent AR(1) process, has an SV-ARMA(2,1) representation by aggregation.
2. The SV(∞) representation by using the invertibility of the MA part of SV-ARMA(2,1) model.
3. Estimate an SV(3) model [instead of an SV(∞)] and recuperate the SV-ARMA(2,1) parameters.
4. From SV-ARMA(2,1) parameters, we can identify the AR factor polynomials of the MFSV(2) model.

From most of the empirical studies, it is prominent that researchers try to fit a distribution that provides best fits for the volatility of asset return. In this section, we point out that an SV(p) model may be served better for that respect since it is not only a natural extension SV(1) model but also an approximated representation of the MFSV or the SV-ARMA model.

4. Stationarity, ergodicity and mixing properties

In case of SV models, the mutual independence of the noise (z_t) and the volatility sequence (w_t) allow for a much simpler probabilistic structure than that of a GARCH process. This independence is one of the attractive probabilistic features of SV models. It is difficult to establish a necessary and sufficient condition for stationarity of GARCH-type models. Nelson (1990) established a solution for the GARCH(1,1) case and Bougerol and Picard (1992) for the general GARCH(p,q) case. For a review of the stationarity of GARCH processes, one may refer to Straumann (2005) or Francq and Zakoïan (2011). To establish the large sample property of SV(p) models we need (w_t, y_t) to be strictly stationary and ergodic. From Carrasco and Chen (2002), the following results ensure the stationarity, ergodicity and mixing condition of SV(p) models.

Result 4.1 STATIONARITY AND ERGODICITY. Assume that z_t and v_t are mutually independent and $\{z_t\}$ is a sequence of i.i.d. real-valued random variables, independent of w_0, with $\mathbb{E}(z_t) = 0$ and $\mathbb{E}(z_t)^2 = 1$. The probability distribution of z_t has a continuous density with respect to Lebesgue measure on real line, and its density is positive on $(-\infty, +\infty)$. Also assume that all the roots of the characteristic equation of the volatility process $[\phi(B) = 0]$ are lie outside the unit circle, i.e., $(1 - \phi_1 B - \cdots - \phi_p B^p) = 0 \Leftrightarrow |B_i| > 1$ for $i = 1, \ldots, p$ ensures stationarity of the volatility process and there is an integer $s \geq 1$ such that

$$\mathbb{E}\,|v_t|^s < \infty, \qquad \left|\sum_{j=1}^{p} \phi_j B^j\right|^s < 1. \tag{4.1}$$

Then

1. $\mathbb{E}[|w_t|]^s < \infty$. *The term* $\{w_t\}$ *is Markov geometrically ergodic. If* $\{w_t\}$ *is initialized from its stationary distribution, then* $\{w_t\}$ *and* $\{y_t\}$ *are strictly stationary and exponential* β *- mixing and this property is preserved by any continuous transformation of* $\{w_t\}$*, i.e.,* $\{\exp(w_t/2)\}$. *The condition is also necessary when* $s = 2$*, i.e., the existence of second moments.*

2. *If* $\mathbb{E}[|\ln|z_t||^s] < \infty$*, then* $\mathbb{E}[|\ln|y_t||^s] < \infty$.

Stochastic volatility model $\{y_t\}$ is a hidden Markov model since it includes a latent Markov chain $\{w_t\}$ and $\{w_t\}$ is independent of the i.i.d. noise process $\{z_t\}$. Proposition 2.1 of Genon-Catalot, Jeantheau and Laredo (2000) show that a hidden Markov model $\{y_t\}$ is ergodic and strong mixing if the hidden chain $\{w_t\}$ is ergodic and strong mixing. In case of SV models, we can conclude a similar result using the Proposition 4 of Carrasco and Chen (2002).

Result 4.2 BETA MIXING. Let $\{y_t\}$ be a generalized hidden Markov model with a hidden chain $\{w_t\}$. Then

1. Since $y_t := \exp(\frac{w_t}{2})\sigma_y z_t \Leftrightarrow \ln|y_t| = w_t/2 + \ln|\sigma_y| + \ln|z_t|$. Thus if $\{w_t\}$ is geometrically ergodic, then $\{(w_t, \ln|y_t|)\}$ is Markov geometrically ergodic.

2. If $\{w_t\}$ is stationary β-mixing, then $\{\ln|y_t|\}$ is stationary β-mixing with a decay rate at least as fast as that of $\{w_t\}$.

5. Simple estimation methods

In this section, we propose simple estimators for SV(p) models and recursive algorithms to obtain these estimators. The moment-based estimator is the extension of Dufour and Valéry (2006, 2009) and the ARMA-based estimators are the extension of Ahsan and Dufour (2015).

5.1. Simple moment-based estimation

We propose a simple method of moments estimator for SV(p) models based on the moments of the following identity that can be obtained from substituting (2.2) into (2.1).

$$y_t := \sigma_y \exp(\frac{\sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t}{2}) z_t, \quad \forall t \tag{5.1}$$

The moments and cross-moments of y_t [$y_t = y_t(\theta)$ where $\theta := (\phi_1, \dots, \phi_p, \sigma_y, \sigma_v)'$] are given in the following Lemma which is a generalization of Lemma 3.1 of Dufour and Valéry (2006).

Lemma 5.1 MOMENTS AND CROSS-MOMENTS OF THE VOLATILITY PROCESS. *Under the assumptions 2.1 − 2.2, and if* $U \sim N(0,1)$, *then* $\mathbb{E}(U^{2p+1}) = 0, \ \forall p \in \mathbb{N}$ *and* $\mathbb{E}(U^{2p}) = \frac{2p!}{2^p p!}, \ \forall p \in \mathbb{N}$*; then the moments and cross-moments of* $y_t = \exp((\sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t)/2)\sigma_y z_t$ *are given by the following formulas: For k, l and m* $\in \mathbb{N}$, *we have:*

$$\begin{aligned} \mu_k(\theta) := \mathbb{E}(y_t^k) &= \sigma_y^k \frac{k!}{2^{k/2}(k/2)!} \exp\left[\frac{k^2}{8}\frac{\sigma_v^2}{1-\sum_{j=1}^{p}\phi_j\rho_j}\right], \ \text{if } k \text{ is even} \\ &= 0, \ \text{if } k \text{ is odd} \end{aligned} \tag{5.2}$$

$$\begin{aligned} \mu_{k,l}(m|\theta) := \mathbb{E}(y_t^k y_{t+m}^l) &= \sigma_y^{k+l} \frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!} \exp\left[\frac{1}{8}\frac{\sigma_v^2(k^2+l^2+2kl\rho_m)}{1-\sum_{j=1}^{p}\phi_j\rho_j}\right], \ \text{if } k \,\&\, l \text{ are even} \\ &= 0, \ \text{if } k \,\&\, l \text{ are odd} \end{aligned} \tag{5.3}$$

where $\rho_j := \text{corr}(w_t, w_{t+j})$.

Dufour and Valéry (2006) derived a closed-form solution for an SV(1) model by exploiting Lemma 5.1. In line with Dufour and Valéry (2006), we can derive a closed-form solution for the higher-order SV process by using Lemma 5.1. In following Lemma, we show it for an SV(p) model where $p = 2$:

Lemma 5.2 CLOSED-FORM MOMENT EQUATIONS SOLUTION FOR THE SV(2) MODEL. *Using Lemma 5.1, we have following moment equations solution:*

$$\phi_1 = \frac{-(log(\frac{\mu_{2,2}(1|\theta)}{(\mu_2(\theta))^2}))(log(\frac{3\mu_{2,2}(2|\theta)}{\mu_4(\theta)}))}{(log(\frac{\mu_4(\theta)}{3(\mu_2(\theta))^2}))^2 - (log(\frac{\mu_{2,2}(1|\theta)}{(\mu_2(\theta))^2}))^2}, \tag{5.4}$$

$$\phi_2 = \frac{-(log(\frac{\mu_{2,2}(1|\theta)}{(\mu_2(\theta))^2}))^2 + (log(\frac{\mu_{2,2}(2|\theta)}{(\mu_2(\theta))^2}))(log(\frac{\mu_4(\theta)}{3(\mu_2(\theta))^2}))}{(log(\frac{\mu_4(\theta)}{3(\mu_2(\theta))^2}))^2 - (log(\frac{\mu_{2,2}(1|\theta)}{(\mu_2(\theta))^2}))^2}, \tag{5.5}$$

$$\sigma_y = \frac{3^{1/4}\mu_2(\theta)}{(\mu_4(\theta))^{1/4}} \tag{5.6}$$

$$\sigma_v = [log(\frac{\mu_4(\theta)}{3(\mu_2(\theta))^2}) - \phi_1[log(\frac{\mu_{2,2}(1|\theta)}{(\mu_2(\theta))^2})] - \phi_2[log(\frac{\mu_{2,2}(2|\theta)}{(\mu_2(\theta))^2})]]^{1/2}. \tag{5.7}$$

From Lemma 5.1, it is easy to derive higher-order autocovariance functions of y_t^2, y_t^2, $y_t^2 y_{t-1}^2$ and $y_t^2 y_{t-2}^2$.

In the following Lemma, we derive these higher-order autocovariance functions. In particular, we need these higher-order autocovariance functions for the asymptotic properties of an SV(2) estimator that is based on Lemma 5.2.

Lemma 5.3 HIGHER-ORDER AUTOCOVARIANCE FUNCTIONS. *Under the assumptions 2.1-2.2 and Lemma 5.1, let* $X_t = (X_{1t}, X_{2t}, X_{3t}, X_{4t})'$ *with*

$$X_{1t} = y_t^2 - \mu_2(\theta), \quad X_{2t} = y_t^4 - \mu_4(\theta),$$

$$X_{3t} = y_t^2 y_{t-1}^2 - \mu_{2,2}(1\,|\,\theta\,), \quad X_{4t} = y_t^2 y_{t-2}^2 - \mu_{2,2}(2\,|\,\theta\,).$$

Then the auto-covariances $\zeta_i(\tau) = Cov(X_{i,t}, X_{i,t+\tau})$, $i = 1,2,3,4$ *are given by:*

$$\zeta_1(\tau) = \mu_2^2(\theta)[\exp(\gamma_\tau) - 1] \tag{5.8}$$

$$\zeta_2(\tau) = \mu_4^2(\theta)[\exp(4\gamma_\tau) - 1], \forall \tau \geq 1 \tag{5.9}$$

$$\zeta_3(\tau) = \mu_{2,2}^2(1\,|\,\theta\,)[\exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}) - 1], \forall \tau \geq 2 \tag{5.10}$$

$$\zeta_4(\tau) = \mu_{2,2}^2(2\,|\,\theta\,)[\exp(\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}) - 1], \forall \tau \geq 3 \tag{5.11}$$

where $\gamma_j := \mathrm{cov}(w_t, w_{t+j})$.

Now it is natural to estimate $\mu_2(\theta)$, $\mu_4(\theta)$, $\mu_{2,2}(1\,|\,\theta\,)$, and $\mu_{2,2}(2\,|\,\theta\,)$ by the corresponding empirical moments:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2, \;\; \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} y_t^4, \;\; \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T} y_t^2 y_{t-1}^2, \;\; \hat{\mu}_2(2) = \frac{1}{T}\sum_{t=1}^{T} y_t^2 y_{t-2}^2 \tag{5.12}$$

This yields the following estimators of the SV coefficients:

$$\hat{\phi}_1 = \frac{-(log(\frac{\hat{\mu}_2(1)}{(\hat{\mu}_2)^2}))(log(\frac{3\hat{\mu}_2(2)}{\hat{\mu}_4})}{(log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}))^2 - (log(\frac{\hat{\mu}_2(1)}{(\hat{\mu}_2)^2}))^2}, \tag{5.13}$$

$$\hat{\phi}_2 = \frac{-(log(\frac{\hat{\mu}_2(1)}{(\hat{\mu}_2)^2}))^2 + (log(\frac{\hat{\mu}_2(2))}{(\hat{\mu}_2))^2}))(log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}))}{(log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}))^2 - (log(\frac{\hat{\mu}_2(1)}{(\hat{\mu}_2)^2}))^2}, \tag{5.14}$$

$$\hat{\sigma}_y = \frac{3^{1/4}\hat{\mu}_2}{(\hat{\mu}_4)^{1/4}}, \tag{5.15}$$

$$\hat{\sigma}_v = [log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}) - \hat{\phi}_1[log(\frac{\hat{\mu}_2(1)}{(\hat{\mu}_2)^2})] - \hat{\phi}_2[log(\frac{\hat{\mu}_2(2)}{(\hat{\mu}_2)^2})]]^{1/2}. \tag{5.16}$$

From the above analysis, we observe that the procedure of Dufour and Valéry (2006) can be easily extended to an SV(2) process. This estimator is computationally simpler than those that are based on sophisticated numerical optimization technique. Similarly, one can compute other higher-order SV models. The expressions of SV(3) or SV(4) estimators are lengthier, so we do not include those equations in the text. However, using this moment-based estimator, we propose recursive estimation algorithm for SV(p) models in Section 5.4.

5.2. ARMA-based estimation

In this subsection, we propose simple estimators for SV(p) models by exploiting the autocovariance structure of y_t^*. In Ahsan and Dufour (2015), we proposed a simple estimator for the SV(1) model where we consider a set of moments based on $y_t^* = (\log(y_t^2) - \mu)$. The ARMA representation of the observed process $\{y_t^*\}$ is given in the following proposition.

Proposition 5.4 ARMA REPRESENTATION OF SV(P) MODELS. *Under the assumptions 2.1 - 2.2, the process y_t^* defined in* (2.12) *has the following ARMA(p, p) representation:*

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} \tag{5.17}$$

with $\eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} = v_t + \varepsilon_t - \sum_{j=1}^{p} \phi_j \varepsilon_{t-j}$, where the error processes $\{v_t\}$ and $\{\varepsilon_t\}$ are mutually independent, the errors v_t are i.i.d. $N(0, \sigma_v^2)$, and the errors ε_t are i.i.d. according to the distribution of a $\log(\chi_1^2)$ *random variable.*

From the above proposition, we have simple expressions for the autocovariances and parameters of the SV(p) model and these are stated in following corollaries.

Corollary 5.5 AUTOCOVARIANCES OF THE OBSERVED PROCESS. *Under the assumptions of Proposition 5.4, the autocovariances of the observed process y_t^* defined in* (2.12) *satisfy the following equations:*

$$\operatorname{cov}(y_t^*, y_{t-k}^*) := \gamma_{y^*}(k) = \begin{cases} \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) + \sigma_v^2 + \sigma_\varepsilon^2; & \text{if } k = 0, \\ \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) - \phi_k \sigma_\varepsilon^2; & \text{if } 1 \le k \le p, \\ \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p); & \text{if } k > p. \end{cases} \tag{5.18}$$

Corollary 5.6 CLOSED-FORM EXPRESSIONS FOR SV(P) PARAMETERS. *Under the assumptions of Proposition 5.4, we have:*

$$\phi_{\boldsymbol{p}} = \boldsymbol{\Gamma}_{(\boldsymbol{k+j-1,p})}^{-1} \gamma_{(\boldsymbol{k+j-1,p})}, \quad j \ge 1 \tag{5.19}$$

$$\sigma_y = [\exp(\mu + 1.27)]^{1/2}, \tag{5.20}$$

$$\sigma_v = [\gamma_{y^*}(0) - \phi_{\boldsymbol{p}}^{'} \gamma_{(\boldsymbol{k,p})} - \pi^2/2]^{1/2}, \tag{5.21}$$

where $\phi_{\boldsymbol{p}} = (\phi_1, \ldots, \phi_p)'$, $\gamma_{(\boldsymbol{k+j-1,p})} = (\gamma_{y^*}(k+j), \ldots, \gamma_{y^*}(k+j+p-1))'$ are vectors and $\boldsymbol{\Gamma}_{(\boldsymbol{k+j-1,p})}$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}_{(\boldsymbol{k,p})} = \begin{bmatrix} \gamma_{y^*}(k+j-1) & \gamma_{y^*}(k-1) & \cdots & \gamma_{y^*}(j) \\ \gamma_{y^*}(k+j) & \gamma_{y^*}(k) & \cdots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(k+j+p-2) & \gamma_{y^*}(k+j+p-3) & \cdots & \gamma_{y^*}(k+j-1) \end{bmatrix}$$

where $k > p$, $\gamma_{y^*}(k) = \operatorname{cov}(y_t^*, y_{t-k}^*)$, with y_t^* and μ defined in (2.12).

Now, it is natural to estimate $\gamma_{y^*}(k)$ and μ by the corresponding empirical moments:

$$\hat{\gamma}_{y^*}(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} y_t^* y_{t+k}^*, \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} \log(y_t^2) \tag{5.22}$$

where by construction y_t^* is a mean corrected process.

Setting $j=1$ in (5.19) and replacing theoretical moments by their corresponding empirical moments yield the following simple estimators of the SV(p) coefficients:

$$\hat{\phi}_{\boldsymbol{p}} = \hat{\boldsymbol{\Gamma}}^{-1}_{(\boldsymbol{k},\boldsymbol{p})}\hat{\gamma}_{(\boldsymbol{k},\boldsymbol{p})}, \tag{5.23}$$

$$\hat{\sigma}_y = [\exp(\hat{\mu}+1.27)]^{1/2}, \tag{5.24}$$

$$\hat{\sigma}_v = [\hat{\gamma}_{y^*}(0) - \hat{\phi}_{\boldsymbol{p}}^{'}\hat{\gamma}_{(\boldsymbol{k},\boldsymbol{p})} - \pi^2/2]^{1/2}. \tag{5.25}$$

5.3. ARMA-based winsorized estimation

One limitation of the above simple ARMA-based estimator is that it may yield a solution outside the admissible area, i.e., some of the eigenvalues of the latent AR process may lie outside the unit circle, and this makes $\hat{\sigma}_v^2 < 0$. This issue can arise especially in small samples or in the presence of outliers.

To deal with a similar problem, Kristensen and Linton (2006) proposed to use “winsorization” (censoring) method which ensures stationarity. From (5.19), it is easy to see that:

$$\phi_{\boldsymbol{p}} = \sum_{j=1}^{\infty} \omega_j \boldsymbol{\Gamma}^{-1}_{(\boldsymbol{k}+\boldsymbol{j}-\boldsymbol{1},\boldsymbol{p})}\gamma_{(\boldsymbol{k}+\boldsymbol{j}-\boldsymbol{1},\boldsymbol{p})} \tag{5.26}$$

for any ω_j sequence with $\sum_{j=1}^{\infty}\omega_j = 1$ so that a more general class of estimators can be defined based on this relationship. It can be expected that for a sufficiently general class of weights one may obtain the same efficiency as the Whittle estimator of Giraitis and Robinson (2001). An estimator of $\phi_{\boldsymbol{p}}$ based on (5.26) is as follows:

$$\tilde{\phi}_{\boldsymbol{p}} = \sum_{j=1}^{J} \omega_j \hat{\boldsymbol{\Gamma}}^{-1}_{(\boldsymbol{k}+\boldsymbol{j}-\boldsymbol{1},\boldsymbol{p})}\hat{\gamma}_{(\boldsymbol{k}+\boldsymbol{j}-\boldsymbol{1},\boldsymbol{p})}, \tag{5.27}$$

where $1 \le J \le T-p$ with $\sum_{j=1}^{J}\omega_j = 1$ and T is the length of time series. Using (5.27), we can propose alternative estimators of $\phi_{\boldsymbol{p}}$.

In the simulation section, we consider four types of winsorized or censored estimators based on the expression given in (5.27). These estimators are also considered by Hafner and Linton (2017) in the context of closed-form estimation of the EGARCH(1,1) model. The first, $\hat{\phi}_{\boldsymbol{p}}^{m}$, is a simple mean of ratios, i.e., it is the estimator in (5.27) with $\omega_j = 1/J$; the second, $\hat{\phi}_{\boldsymbol{p}}^{ld}$, is a mean of ratios with linearly declining weights, i.e., it is the estimator in (5.27) with $\omega_j = 2(1-j/(J+1))/J$; the third is the median: $\hat{\phi}_{\boldsymbol{p}}^{med} = med\{\hat{\boldsymbol{\Gamma}}^{-1}_{(\boldsymbol{k}+\boldsymbol{j}-\boldsymbol{1},\boldsymbol{p})}\hat{\gamma}_{(\boldsymbol{k}+\boldsymbol{j}-\boldsymbol{1},\boldsymbol{p})}\}$, and the fourth is the OLS without intercept regression estimator given by (5.28)

$$\hat{\phi}_{\boldsymbol{p}}^{ols} = (\overline{a}'\overline{a})^{-1}\overline{a}'\overline{e}, \tag{5.28}$$

where $\overline{a} = (\hat{\Gamma}^{-1}_{(k,p)}\omega_1^{1/2}, \ldots, \hat{\Gamma}^{-1}_{(k+J-1,p)}\omega_J^{1/2})'$ and $\overline{e} = (\hat{\gamma}_{(\boldsymbol{k},\boldsymbol{p})}\omega_1^{1/2}, \ldots, \hat{\gamma}_{(\boldsymbol{k}+\boldsymbol{J}-\boldsymbol{1},\boldsymbol{p})}\omega_J^{1/2})'$. All these estimators are depend on J and for $J=1$, they are equivalent to the simple ARMA estimator that given by (5.23).

5.4. Recursive estimation for SV(p) models

Previously we had shown that it is possible to derive higher-order closed-form solution for SV(p) models. In this section, we propose recursive estimation algorithms for SV(p) models. We use an alternative method provided by Durbin (1960) that avoids the matrix inversion in the Yule-Walker equations. This method is called the Durbin-Levinson (DL) Algorithm, and it is a prediction algorithm. One important feature of DL Algorithm is that we will automatically get partial autocorrelations and mean-squared errors associated with our predictions. For notational convenience, we use a different indexation for the autoregressive parameters of the volatility process. For example, the SV(p) parameters are now denoted by $\Theta_p^{SV} := (\{\phi_{p,j}\}_{j=1}^{p}, \sigma_{pv}, \sigma_y)$.

Under the assumptions 2.1-2.2, the latent volatility process is an AR(p) process, satisfy the Yule-Walker equations. Thus we can apply DL algorithm that is designed for a pure autoregressive process, and we obtain parameters of higher-order SV models recursively. We obtain the extension of the Dufour and Valéry (2006) estimator for SV(p) models by using the following recursive formulae:

$$\phi_{p,p} = \frac{\rho_p - \sum_{j=1}^{p-1} \phi_{p-1,j}\, \rho_{p-j}}{1 - \sum_{j=1}^{p-1} a_{p-1,j}\, \rho_j}, \tag{5.29}$$

$$\phi_{p,j} = \phi_{p-1,j} - \phi_{p,p}\phi_{p-1,p-j}, \quad \forall j = 1,2,3,\ldots,p-1, \tag{5.30}$$

where ρ_j is the auto-correlation of the autoregressive process at lag j.

Using the following Lemma, we can get the solution of an $SV(p)$ model from an $SV(p-1)$ model:

Lemma 5.7 RECURSIVE MOMENT EQUATION SOLUTION. *Under the assumptions 2.1-2.2, we can obtain parameters of an $SV(p)$ model, i.e., $\Theta_p^{SV} := (\{\phi_{p,j}\}_{j=1}^{p}, \sigma_{pv}, \sigma_y)$, from parameters of an $SV(p-1)$ model, recursively by using the following algorithm:*

$$\hat{\sigma}_y = \frac{3^{1/4}\hat{\mu}_2}{(\hat{\mu}_4)^{1/4}}, \tag{5.31}$$

$$\hat{\phi}_{p,p} = \frac{\hat{\rho}_p - \sum_{j=1}^{p-1} \hat{\phi}_{p-1,j}\hat{\rho}_{p-j}}{1 - \sum_{j=1}^{p-1} \hat{\phi}_{p-1,j}\hat{\rho}_j}, \tag{5.32}$$

$$\hat{\phi}_{p,j} = \hat{\phi}_{p-1,j} - \hat{\phi}_{p,p}\hat{\phi}_{p-1,p-j} \quad \forall j = 1,2,3,\ldots,p-1, \tag{5.33}$$

$$\hat{\sigma}_v = \left[(1 - \sum_{j=1}^{k} \hat{\phi}_{p-1,j}\hat{\rho}_j) \left[log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}) \right] \right]^{1/2}, \tag{5.34}$$

where

$$\hat{\rho}_j = log(\frac{\hat{\mu}_2(j)}{(\hat{\mu}_2)^2}) / log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}), \quad for \;\; j = 1,2,\ldots \tag{5.35}$$

and the sample auto-correlation of volatility process at lag j is denoted by $\hat{\rho}_j$ and we can calculate these auto-correlations by using (5.35).

Note that after calculating the sample auto-correlations, we can estimate the parameters of the model in the second stage with the help of a DL algorithm. In the proof, we show that how one can recursively obtain the solution of an SV(2) model.

The recursive estimation of the ARMA-based estimator exploits extended Yule-Walker (EYW) equations of the observed process. When the MA order is fixed, the system of the EYW equations constitutes a nested Toeplitz system. A *Generalized Durbin-Levinson* algorithm for the ARMA-based estimator for SV(p) model is useful when neither the AR order nor the MA order is known. We consider the case $i = q$, i.e., the MA order is q.

For $i = 0$, use the Durbin-Levinson algorithm to calculate

$$\{\hat{\phi}_{p,j}^{(0)} \mid p \geq 1, \ \ j = 1, \dots, p\}.$$

For $i \geq 1$, calculate

$$\hat{\phi}_{p,0}^{(i-1)} = -1,$$

and

$$\hat{\phi}_{p,j}^{(i)} = \hat{\phi}_{p+1,j}^{(i-1)} - \frac{\hat{\phi}_{p+1,p+1}^{(i-1)}}{\hat{\phi}_{p,p}^{(i-1)}} \hat{\phi}_{p,j-1}^{(i-1)}, \ \text{where } p \geq 1, \ \ j = 1, \dots, p,$$

$$\hat{\sigma}_y = [\exp(\hat{\mu} + 1.27)]^{1/2},$$

$$\hat{\sigma}_{pv} = [\hat{\gamma}_{y^*}(0) - \sum_{j=1}^{p} \hat{\phi}_{p,j} \hat{\gamma}_{y^*}(j) - \pi^2/2]^{1/2}.$$

This algorithm is the same as Tsay and Tiao (1984) algorithm [except for equations involving $\hat{\sigma}_y$ and $\hat{\sigma}_{pv}$] for calculating the extended sample autocorrelation function under the stationarity assumption.

6. Andersen-Sørensen type GMM estimation

The simple estimators proposed in the previous section can be considered specific cases of a GMM-type estimator where we used a few moments. In this section, we propose GMM estimators for SV(p) models with many moments in line with Andersen and Sørensen (1996). In literature, Andersen and Sørensen (1996) proposed GMM estimators for the SV(1) model but the GMM estimator for SV(p) models remains to be discussed. The GMM estimation was formalized by Hansen (1982), and since then it has become one of the most popular methods of estimation for many models in economics and finance. Unlike the MLE, GMM does not require complete knowledge of the distribution of the data. The GMM estimator of SV(p) model is a natural extension of our simple closed-form moment-based estimator. Following the general methodology of GMM, our goal is to minimize the quadratic form with respect to the parameter vector:

$$M_T = [\bar{g}_T(Y_T) - \mu(\theta)]' \hat{\Omega}_T [\bar{g}_T(Y_T) - \mu(\theta)] \tag{6.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(Y_T)$ the corresponding vector of empirical moments based on the vector $Y_T = (y_1, \dots, y_T)'$, and $\hat{\Omega}_T$ a positive-definite (possibly random) matrix. We compute the corresponding sample averages such that

$$g_T(\theta) := \bar{g}_T(Y_T) - \mu(\theta) = \sum_{t=1}^{T} [\bar{g}_t(Y_t) - \mu(\theta)].$$

Now under the standard regularity assumptions,

$$\sqrt{T}[\hat{\theta}_T(\Omega) - \theta_0] \xrightarrow{D} N[0, V(\theta_0 \,|\, \Omega)], \tag{6.2}$$

where

$$V(\theta_0 \,|\, \Omega) = [J(\theta)\Omega J(\theta)']^{-1} J(\theta)\Omega\Omega_* \Omega J(\theta)' [J(\theta)\Omega J(\theta)']^{-1}, \tag{6.3}$$

and $J(\theta) = \frac{\partial \mu'}{\partial \theta}$. Furthermore, if (i) $J(\theta)$ is a square matrix, or (ii) Ω_* is non-singular and $\Omega = \Omega_*^{-1}$, then

$$V(\theta_0|\Omega) = [J(\theta)\Omega_*^{-1}J(\theta)']^{-1} := V_*(\theta). \tag{6.4}$$

The $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$. The latter, in particular, is reached when the dimensions of μ and θ are the same, in which case the estimator is obtained by solving the equation

$$\bar{g}_T(Y_T) = \mu(\hat{\theta}_T)$$

Consistent estimators $V(\theta_0|\Omega)$ and $V_0(\theta_0)$ can be obtained by replacing θ_0 and Ω_* with their consistent estimators. The sample analog of Ω_* is given by

$$\hat{\Omega}_* = \hat{\Gamma}_0 + \sum_{i=1}^{\infty}(\hat{\Gamma}_i + \hat{\Gamma}_i') \tag{6.5}$$

Given this structure, it is natural to estimate $\hat{\Omega}_*$ by truncating this infinite sum and using the sample auto-covariances, where

$$\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}[g_{t-j}(\hat{y}) - \mu(\theta)][g_{t-j}(\hat{y}) - \mu(\theta)]'$$

with θ replaced by a consistent estimator of it.[2] However for Ω_*, we need to use the heteroskedasticity and autocorrelation covariance (HAC) matrices to avoid any potential inconsistency caused by inappropriate assumptions about the dynamic specification of $[g_t(\hat{y}) - \mu(\theta)]$. This estimator is consistent under relatively weak assumptions on the dependence structure of the process, and this class consists of estimators of the form:

$$\hat{\Omega}_{*,HAC} = \hat{\Gamma}_0 + \sum_{i=1}^{T-1}\omega_{i,T}(\hat{\Gamma}_i + \hat{\Gamma}_i') \tag{6.6}$$

where $\omega_{i,T}$ is known as the kernel (or weight), and it must be chosen to ensure: (i) consistency and (ii) positive semi-definiteness of $\hat{\Omega}_*$. In literature, there have been few proposed kernel functions that can fit into the above equation.[3] Thus a consistent estimator of $\hat{V}_*(\theta_0)$ is given by

$$\hat{V}_* = [J(\hat{\theta}_T)\hat{\Omega}_*^{-1}J(\hat{\theta}_T)']^{-1}.$$

Andersen and Sørensen (1996), based on a Monte Carlo simulations study, address several issues related to GMM estimation of SV(1) model. One issue of GMM estimation is the choice of the number of moment conditions. If weighted appropriately, by increasing the number of moment conditions (using additional information), one cannot make the parameter estimates worse. However the weighting matrix, Ω, must itself be estimated, and with q moment conditions, we need to estimate $q(q+1)/2$ elements of Ω, and a larger number of moment conditions could lead to poorer estimates of Ω and worse estimates of the parameters. Another issue with GMM is that there is not much guidance on which moment conditions to use. For SV models, one can construct moment conditions based on infinitely many functions of returns; see Melino and Turnbull (1990).

[2] The truncation parameter l_T is allowed to grow with the sample size such that: $l_T \to \infty$ as $T \to \infty$ and $l_T = 0(T^{1/3})$ see White and Domowitz (1984).

[3] (i) Bartlett kernel by Newey and West (1987), (ii) Parzen kernel by (Gallant (1987) page 533), and (iii) Quadratic spectral kernel by Andrews (1991).

7. Asymptotic distributional theory

In this section, we derive asymptotic distributions for our simple estimators. For the asymptotic distribution of GMM estimators; see Hansen (1982).

7.1. Moment-based estimators

Dufour and Valéry (2006) derived an asymptotic distributional theory for the SV(1) estimator. In line with that we establish the asymptotic distributional theory for the moment-based SV(p) estimators. Our approach for constructing an MM estimator is to minimize the quadratic form:

$$M_T = [\bar{g}_T(Y_T) - \mu(\theta)]' \hat{\Omega}_T [\bar{g}_T(Y_T) - \mu(\theta)] \tag{7.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(y_T)$ the corresponding vector of empirical moments based on the vector $Y_T = (y_1, \ldots, y_T)'$, and $\hat{\Omega}_T$ a positive-definite (possibly random) matrix. Of course, this estimator belongs to the general family of moment estimators, for which a number of general asymptotic results do exist; see Hansen (1982), Gouriéroux and Monfort (1995) (Volume 1, Chapter 9) and Newey and McFadden (1994).

It is worth noting at this stage that Andersen and Sørensen (1996) did refer to the asymptotic distribution of the usual GMM estimator as derived in Hansen (1982) for the SV(1) model, but without checking the suitable regularity conditions. We want to find the estimator $\hat{\theta}_T(\hat{\Omega}_T)$ by minimizing $M_T(\theta)$, and for that we will consider the following assumptions, where θ_0 denotes the “true" value of the parameter vector θ.

Assumption 7.1 ASYMPTOTIC NORMALITY OF EMPIRICAL MOMENTS.

$$\sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{D} N[0, \Omega_*]$$

where $Y_T := (y_1, \ldots, y_T)'$ *and*

$$\Omega_* = \lim_{T\to\infty} \mathbb{E}\{T[\bar{g}_T(Y_T) - \mu(\theta_0)][\bar{g}_T(Y_T) - \mu(\theta_0)]'\}$$

Assumption 7.2 ASYMPTOTIC NON-SINGULARITY OF WEIGHT MATRIX.

$$\underset{T\to\infty}{\text{plim}}(\hat{\Omega}_T) = \Omega,$$

where $det(\Omega) \neq 0$.

Assumption 7.3 DIFFERENTIABILITY OF WEIGHT MATRIX. $\mu(\theta_0)$ *is twice continuously differentiable in an open neighborhood of* θ_0 *and the Jacobian matrix* $J(\theta_0)$ *has full rank, where* $J(\theta) = \frac{\partial \mu'}{\partial \theta}$.

Given these assumptions, the asymptotic distribution of $\hat{\theta}_T$ is determined by a standard argument on method-of-moments estimation.

Lemma 7.1 ASYMPTOTIC DISTRIBUTION OF METHOD-OF-MOMENTS ESTIMATOR. *Under the assumptions 2.1 - 2.2 and 7.1 - 7.3,*

$$\sqrt{T}[\hat{\theta}_T - \theta_0] \xrightarrow{D} N[0, V(\theta_0 | \Omega)] \tag{7.2}$$

where

$$V(\theta_0 | \Omega) = [J(\theta)\Omega J(\theta)']^{-1} J(\theta)\Omega\Omega_*\Omega J(\theta)'[J(\theta)\Omega J(\theta)']^{-1} \tag{7.3}$$

where $J(\theta) = \frac{\partial \mu'}{\partial \theta}$. *If, furthermore, (i)* $J(\theta)$ *is a square matrix, or (ii)* Ω_* *is non-singular and* $\Omega = \Omega_*^{-1}$, *then*

$$V(\theta_0 | \Omega) = [J(\theta)\Omega_*^{-1}J(\theta)']^{-1} := V_*(\theta). \quad (7.4)$$

Here, $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$, and a consistent estimator of $\hat{V}_*(\theta_0)$ is given by

$$\hat{V}_* = [J(\hat{\theta}_T)\hat{\Omega}_*^{-1}J(\hat{\theta}_T)']^{-1}.$$

Since we are using a number of moments equal to the number of parameters, the moment estimator can be obtained by taking $\hat{\Omega}_T$ equal to an identity matrix so that Assumption 7.2 automatically holds. Thus we only need to show that the Assumption 7.1 holds.

Lemma 7.2 ASYMPTOTIC DISTRIBUTION FOR EMPIRICAL MOMENTS. *Under the assumptions 2.1 - 2.2 with* $p > 1$, *we have:*

$$\sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{D} N[0, \Omega_*] \quad (7.5)$$

where $\bar{g}_T(Y_T) = \sum_{t=1}^{T} g_t$, $g_t = [y_t^2, y_t^4, y_t^2 y_{t-1}^2, \ldots, y_t^2 y_{t-p}^2]'$, *and*

$$\Omega_* = V[g_t] = \mathbb{E}[g_t g_t'] - \mu(\theta_0)\mu(\theta_0)'.$$

7.2. ARMA-based estimators

We derive the asymptotic properties of the ARMA-based estimator $\hat{\theta} := (\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v)'$ under the following set of assumptions:

Assumption 7.4 DISTRIBUTION OF THE ERROR PROCESSES. *The error processes* z_t *and* v_t *are mutually independent and* $\{z_t\}$ *is a sequence of i.i.d. real-valued random variables, independent of* w_0. *The probability distribution of* z_t *has a continuous density with respect to Lebesgue measure on real line, and its density is positive on* $(-\infty, +\infty)$. *The transformed error* ε_t *satisfies* $\mathbb{E}(|\varepsilon_t|)^s < \infty$, *where s is an integer such that* $s \geq 1$.

Assumption 7.5 STATIONARITY OF THE LATENT PROCESS. *The latent process* $\{w_t\}$ *is strictly stationary with* $\mathbb{E}[|w_t|]^s < \infty$ *and there is an integer* $s \geq 1$ *such that*

$$\mathbb{E}|v_t|^s < \infty, \quad \left|\sum_{j=1}^{p} \phi_j B^j\right|^s < 1.. \quad (7.6)$$

Under the assumptions 7.4 and 7.5 with $s = 2$, the observed process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with exponential β-mixing (see results 4.1 and 4.2) with finite second moment, *i.e.*, $\mathbb{E}\left[y_t^{*2}\right] < \infty$. In the following Lemma, using Ergodic theorem, we prove the consistency of the empirical moments that defined in (5.22).

Lemma 7.3 CONSISTENCY OF EMPIRICAL MOMENTS. *Under the assumptions 7.4 and 7.5 with* $s = 2$, *the estimators* $\hat{\Gamma}(m) = (\hat{\gamma}_{y^*}(k))_{k=0,\ldots,m}$ *and* $\hat{\mu}$ *in* (5.22) *satisfy:*

$$\hat{\Gamma}(m) \xrightarrow{p} \Gamma(m) = (\gamma_{y^*}(k))_{k=0,\ldots,m} \text{ and } \hat{\mu} \xrightarrow{p} \mu, \quad m \geq 1. \quad (7.7)$$

The assumptions 7.4 and 7.5 with $s = 4$ are necessary and sufficient for the SV model to have a strictly stationary solution with a finite fourth moment of y_t^*, *i.e.*, $\mathbb{E}\left[y_t^{*^4}\right] < \infty$. Note that the fourth moment of y_t^* translates into the eighth moment of y_t. This solution will be β-mixing with geometrically decreasing mixing coefficients. In the following Lemma, using a Central Limit theorem for the stationary and ergodic process (Lindeberg-Levy theorem for the dependent process), we derive the asymptotic distribution of the empirical moments that defined in (5.22).

Lemma 7.4 ASYMPTOTIC DISTRIBUTION OF EMPIRICAL MOMENTS. *Under the assumptions 7.4 and 7.5 with $s = 4$, the estimators $\hat{\Gamma}(m) = (\hat{\gamma}_{y^*}(k))_{k=0,\dots,m}$ and $\hat{\mu}$ in* (5.22) *satisfy:*

$$\sqrt{T}\begin{bmatrix}\hat{\mu}-\mu \\ \hat{\Gamma}(m)-\Gamma(m)\end{bmatrix} \xrightarrow{d} N\left(0,\ \begin{bmatrix} V_\mu & C_{\mu,\Gamma(m)}^{'} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)}\end{bmatrix}\right), \tag{7.8}$$

where

$$V_\mu = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau), \quad V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t) + 2\sum_{\tau=1}^{\infty}\mathrm{cov}(\Lambda_t, \Lambda_{t+\tau}), \quad C_{\mu,\Gamma(m)} = (\bar{c},\ 0_{[1\times m]})', \tag{7.9}$$

$$\Lambda_t = (\Lambda_{t,0},\dots,\Lambda_{t,m})', \quad \Lambda_{t,k} = y_t^* y_{t+k}^* = [\log(y_t^2)-\mu][\log(y_{t+k}^2)-\mu], \quad k = 0,\dots,m, \tag{7.10}$$

$$\bar{c} := C_{\mu,\Gamma(0)} = 2\sum_{t=1}^{\infty}\mathbb{E}[y_t^{*3}] = 2\sum_{t=1}^{\infty}(\mathbb{E}[w_t^3]+\mathbb{E}[\varepsilon_t^3]) = 2\sum_{t=1}^{\infty}\mathbb{E}[\varepsilon_t^3]. \tag{7.11}$$

Theorem 7.5 ASYMPTOTIC DISTRIBUTION OF THE SIMPLE ARMA-BASED ESTIMATOR. *Under the assumptions 7.4 and 7.5 with $s = 4$, and if* (7.7) - (7.8) *hold, then the estimator $\hat{\theta} := (\hat{\phi}_1,\dots,\hat{\phi}_p,\hat{\sigma}_y,\hat{\sigma}_v)'$ given in* (5.23) - (5.25) *is consistent, i.e., $\hat{\theta} \xrightarrow{p} \theta$ and*

$$\sqrt{T}(\hat{\theta}-\theta) \xrightarrow{d} \mathrm{N}[0,V], \tag{7.12}$$

where

$$V = \frac{\partial D_p(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\dots,\gamma_{y^*}(2p))}{\partial(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\dots,\gamma_{y^*}(2p))}\begin{bmatrix} V_\mu & C_{\mu,\Gamma(2p)}^{'} \\ C_{\mu,\Gamma(2p)} & V_{\Gamma(2p)}\end{bmatrix}\left(\frac{\partial D_p(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\dots,\gamma_{y^*}(2p))}{\partial(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\dots,\gamma_{y^*}(2p))}\right)^{'},$$

where the function $D_p = (D_{\phi_{\boldsymbol{p}}}, D_{\sigma_y}, D_{\sigma_v})'$ is given by

$$D_{\phi_{\boldsymbol{p}}} = \boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})}^{-1}\gamma_{(\boldsymbol{k},\boldsymbol{p})}, \quad D_{\sigma_y} = \exp(\mu+1.27)^{1/2}, \quad D_{\sigma_v} = [\gamma_{y^*}(0) - \phi_{\boldsymbol{p}}^{'}\gamma_{(\boldsymbol{k},\boldsymbol{p})} - \pi^2/2]^{1/2},$$

where $\phi_{\boldsymbol{p}} = (\phi_1,\dots,\phi_p)'$, $\gamma_{(\boldsymbol{k},\boldsymbol{p})} = (\gamma_{y^*}(k+1),\dots,\gamma_{y^*}(k+p))'$ are vectors and $\boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})}$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})} = \begin{bmatrix} \gamma_{y^*}(k) & \gamma_{y^*}(k-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(k+1) & \gamma_{y^*}(k) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(k+p-1) & \gamma_{y^*}(k+p-2) & \cdots & \gamma_{y^*}(k)\end{bmatrix}$$

with $k > p$, $\gamma_{y^*}(k) = \mathrm{cov}(y_t^*, y_{t-k}^*)$, $y_t^* = (\log y_t^2 - \mu)$, and μ is equal to the mean of $\log(y_t^2)$ *and with $\hat{\theta} =$*

$D_p(\hat{\mu}, \hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \ldots, \hat{\gamma}_{y^*}(2p))$. *In the proof, we gave explicit form of the analytical moment derivatives of* $D_p(\hat{\mu}, \hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \ldots, \hat{\gamma}_{y^*}(2p))$.

An estimator of the covariance matrix V can be obtained by first estimating V_μ and $V_\Gamma(2p)$ using heteroskedasticity and autocorrelation consistent variance estimators and then substituting $\hat{\mu}, \hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \ldots, \hat{\gamma}_{y^*}(2p)$ into $\partial D_p(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(2p))/\partial(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(2p))$. One can alternatively use the analytic expressions of $\gamma_{y^*}(k)$ to obtain an estimator of V_μ. The ARMA-based estimator can view as a GMM-type estimator, so alternatively one can also use GMM standard errors.

Theorem 7.5 covers the simplest ARMA-based estimator. It is easy to see that the asymptotic distribution of more general winsorized estimators can be derived in the same way upon using Lemmas 7.3 - 7.4.

8. Monte Carlo tests

In this section, we discuss simulation-based inference procedures for SV(p) models. The simulation-based methods are more attainable in the context of this study for two reasons: (1) the SV(p) model is a parametric model, and we can easily simulate this model; (2) we can simulate the test statistic of SV(p) parameters that based on a computationally inexpensive estimator. However, if the estimator is computationally expensive, then we cannot simulate the test statistic easily, and the simulation will run forever. Using our proposed computationally simple estimators, one can construct more reliable finite-sample inference.

We now examine the usefulness of our simple estimators in the context of simulation-based inference, i.e., Monte Carlo test technique. The technique of Monte Carlo tests was originally proposed by Dwass (1957) for implementing permutation tests and did not involve nuisance parameters. This technique was also independently proposed by Barnard (1963); for a review, see Dufour and Khalaf (2001) and for a general discussion and proofs, see Dufour (2006). It has the great attraction of providing exact (randomized) tests based on any statistic whose finite-sample distribution may be intractable but can be simulated. One can replace the unknown or intractable theoretical distribution $F(S|\theta)$, where $\theta := (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v)'$, by its sample analog based on the statistics $S_1(\theta), \ldots, S_N(\theta)$ simulated under the null hypothesis.

Let us first consider the pivotal statistics case, i.e., the case where the distribution of the test statistic under the null hypothesis does not depend on nuisance parameters. We can then proceed as follows to obtain an exact critical region.

1. Let S_0 be the observed test statistic (based on data).

2. By Monte Carlo methods, draw N i.i.d. replications of S, denoted by $S(N) = (S_1, \ldots, S_N)$ under H_0, independently of S_0, i.e., $S_0, S_1, \ldots, S_N$ be exchangeable.

3. From the simulated samples compute the MC p-value $\hat{p}_N[S] := p_N[S_0; S(N)]$, where

$$p_N[x, S(N)] := \frac{NG_N[x; S(N)] + 1}{N+1} \tag{8.1}$$

$$G_N[x; S(N)] := \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty)}(S_i - x), \qquad I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \in [0,\infty), \\ 0 & \text{if } x \notin [0,\infty). \end{cases} \tag{8.2}$$

In other words, $p_N[S_0; S(N)] = (NG_N[S_0; S(N)] + 1)/(N+1)$ where $NG_N[S_0; S(N)]$ is the number of simulated values which are greater than or equal to S_0 . When $S_0, S_1, \ldots, S_N$ are all distinct [an event with probability one when the vector $(S_0, S_1, \ldots, S_N)'$ has an absolutely continuous distribution], $\hat{R}_N(S_0) = N + 1 - NG_N[S_0; S(N)]$ is the rank of S_0 in the series $S_0, S_1, \ldots, S_N$.

4. The MC critical region is: $\hat{p}_N[S] \leq \alpha$, $0 < \alpha < 1$. If $\alpha(N+1)$ is an integer and the distribution of S is continuous under the null hypothesis, then under null,

$$P[\hat{p}_N[S] \leq \alpha] = \alpha; \tag{8.3}$$

see Dufour (2006).

We will now study the case where the distribution of the test statistic depends on nuisance parameters. In other words, we consider a model $\{(\Xi, \mathbb{A}_{\Xi}, P_\theta) : \theta \in \Omega\}$ where we assume that the distribution of S is determined by $P_{\bar{\theta}}$, where $\bar{\theta}$ represents the true parameter vector. To deal with this complication, the MC test procedure can be modified as follows.

1. To test the null hypothesis
$$H_0 : \bar{\theta} \in \Omega_0,$$
where $\Omega_0 \subset \Omega$, we calculate the relevant test statisticS_0 based on data.

2. For each $\theta \in \theta_0$, by Monte Carlo methods, we generate N i.i.d. replications of $S : S(N, \theta) = [(S_1(\theta), \ldots, S_N(\theta)]$.

3. Using these simulated test statistics, we compute the MC p-value $\hat{p}_N[S|\theta] := p_N[S_0; S(N, \theta)]$, where
$$p_N[x; S(N, \theta)] := \frac{NG_N[x; S(N, \theta)] + 1}{N+1}. \tag{8.4}$$

4. The p-value function $\hat{p}_N[S|\theta]$ as a function of θ is maximized over the parameter values compatible with the Ω_0, i.e., the null hypothesis, and H_0 is rejected if
$$\sup_{\theta \in \Omega_0} \hat{p}_N[S|\theta] \leq \alpha. \tag{8.5}$$
If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:
$$P[\sup_{\theta \in \Omega_0} \{\hat{p}_N[S|\theta]\} \leq \alpha] \leq \alpha, \tag{8.6}$$
The test defined by $\hat{p}_N[S|\theta] \leq \alpha$ has size α for known θ . Treating θ as a nuisance parameter and Ω_0 is a nuisance parameter set consistent with null, the test is *exact at level* α; for a proof, see Dufour (2006).

Because of the maximization in the critical region (8.5) the test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as almost-unidentified models or time series processes involving unit roots. In particular, even though the moment conditions defining the estimator are derived under the stationarity assumption, this does not question in any way the validity of maximized MC tests, unlike the parametric bootstrap whose distributional theory is based on strong regularity conditions. Only the power of MMC tests may be affected. However, the simulated p-value function is not continuous, so standard gradient-based methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, e.g. simulated annealing [see Goffe, Ferrier and Rogers (1994)].

A simplified approximate version of the MMC procedure can alleviate its computational load whenever a consistent point or set estimate of θ is available. To do this, we reformulate the setup in order to allow for an increasing sample size, i.e., now the test statistic depends on a sample of size T, $S = S_T$.

1. Let S_{T0} be the observed test statistic (based on data) and the distribution of S involves nuisance parameters under the null and $\bar{\theta} \in \Omega_0$ with $\Omega_0 \subset \Omega$ and $\Omega_0 \neq \emptyset$.

2. we have a consistent set estimator C_T of θ (under H_0) such that
$$\lim_{T\to\infty} P[\bar{\theta} \in C_T] = 1 \text{ under } H_0. \tag{8.7}$$

3. For each $\theta \in C_T$, by Monte Carlo methods, we generate N i.i.d. replications of $S : S_T(N, \theta) = [(S_{T1}(\theta), \ldots, S_{TN}(\theta)]$.

4. Using these simulations we compute the MC p-value $\hat{p}_{TN}[S_T|\theta] := p_{TN}[S_{T0}; S_T(N, \theta)]$, where
$$p_{TN}[x; S_T(N, \theta)] := \frac{NG_{TN}[x; S_T(N, \theta)] + 1}{N+1}. \tag{8.8}$$

5. The p-value function $\hat{p}_{TN}[S_T|\theta]$ as a function of θ is maximized with respect to θ in C_T, and H_0 is rejected if
$$\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha. \tag{8.9}$$
If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:
$$\lim_{T\to\infty} P[\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha] \leq \alpha, \tag{8.10}$$
i.e., we control for the level asymptotically.

In practice, it is easy to find a consistent set estimate of $\bar{\theta}$, whenever a *consistent* point estimate $\hat{\theta}_T$ of $\bar{\theta}$ available (e.g. a GMM estimator). For instance, any set of the form
$$C_T = \{\theta : \left\|\hat{\theta}_T - \theta\right\| < d\} \tag{8.11}$$
with d a fixed positive constant independent of T, satisfies (8.7). The consistent set estimate MMC (CSEMMC) method is especially useful when the distribution of the test statistic is highly sensitive to nuisance parameters. Here, possible discontinuities in the asymptotic distribution are automatically overcome through a numerical maximization over a set that contains the true value of the nuisance parameter with probability one asymptotically (while there is no guarantee for the point estimate to converge sufficiently fast to overcome the discontinuity). It is worth noting that there is no need to maximize the p-value function with respect to unidentified parameters under the null hypothesis. Thus, parameters which are unidentified under the null hypothesis can be set to any fixed value and the maximization be performed only over the remaining identified nuisance parameters. When there are several nuisance parameters, one can use simulated annealing, an optimization algorithm which does not require differentiability. Indeed the simulated p-value function is not continuous, so standard gradient based methods cannot be used to maximize it. For an example where this is done on a VAR model involving a large number of nuisance parameters, see Dufour and Jouini (2006).

In Dufour and Khalaf (2002) call the test based on simulations using a point nuisance parameter estimate a *local* MC (LMC) test. The term local reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. Here if the set C_T in (8.9) is reduced to a single point estimate $\hat{\theta}_T$, *i.e.* $C_T = \{\hat{\theta}_T\}$, we get a LMC test

$$\hat{p}_{TN}[S_T|\hat{\theta}_T] \leq \alpha, \tag{8.12}$$

which can be interpreted as a parametric bootstrap test. Note that no asymptotic argument on the number N of MC replications is required to obtain this result; this is the fundamental difference between the latter procedure and the parametric bootstrap method.

Even if $\hat{\theta}_T$ is a consistent estimate of θ (under the null hypothesis), the condition (8.7) is not usually satisfied in this case, so additional assumptions are needed to show that the parametric bootstrap procedure yields an asymptotically valid test. It is computationally less costly but clearly less robust to violations of regularity conditions than the MMC procedure; for further discussion, see Dufour (2006).

Furthermore, the LMC non-rejections are *exactly* conclusive in the following sense: if $\hat{p}_N[S|\hat{\theta}_0] > \alpha$, then the exact *Maximized Monte Carlo* (MMC) test is clearly not significant at level α.

9. Simulation study

In this section, we investigate the finite-sample properties of the proposed winsorized estimators through a simulation study. We generate an SV(2) processes with $(\phi_1, \phi_2, \sigma_y, \sigma_v) = (0.45, 0.45, 0.25, 2.5)$. We consider different sample sizes $T = (100, 1000, 10000)$ and use 1000 replications. All four censored estimators depend on J, so we use different values of $J = (1, 5, 10, 20, 30, 40, 50, 100)$. Simulation results are reported in Table 1. It is remarkable that both the weighted and unweighted means underperform for higher J, due to the high variability of estimated ACF. Further, these mean estimators produce inadmissible parameter values even in large samples. However, the number of unacceptable parameter values decline as the sample size increases. This fact also tells us that the variability of estimated ACF is also going down as the sample size increases. Median and OLS estimates are robust to these, and standard errors are reasonably small. However, OLS estimates outperformed the other three estimators in terms of bias and standard error, across different sample sizes particularly in small samples. Further, it is also robust to different values of J. From the reported results; there may be a bias-variance trade-off for higher values of J. Finally, we suggest to use OLS for winsorizing and use small values of J for large samples or vice versa.

Now we explore the statistical performance of our proposed estimators, these include the moment estimator, the simple ARMA-based estimator, the winsorized ARMA-based (W-ARMA) estimator [it is the no intercept regression with $J = 10$] and GMM estimators, in terms of bias and root mean square error (RMSE). Globally, there is no uniform ranking between the different estimators, but the performance of the Bayesian estimator remain superior among the competing methods in the context of SV(1) model. Under the following simulation designs, we compare our proposed estimators to the Bayesian estimator. We use the Matlab code of Chan and Grant (2016) for the Bayesian estimation with their specified prior.

We simulate four SV(2) models where parameter values of $(\phi_1, \phi_2, \sigma_y, \sigma_v)$ are $M1 = (0.30, 0.60, 0.025, 2.5)$, $M2 = (0.90, -0.90, 0.5, 2.5)$, $M3 = (0.45, 0.45, 0.25, 2.5)$ and $M4 = (0.0, 0.90, 0.025, 2.5)$. The parameters have been selected arbitrarily since empirical applications of SV(p) models are rare in the literature. The variance of the returns process determined by σ_y and the magnitude of σ_y tends to vary depending on whether returns are measured, e.g., daily, monthly. The estimation of σ_y, will typically not have much of an effect on the estimation of the other parameters $(\phi_1, \phi_2, \sigma_v)$. The simulations use 1000 replications and we consider two different sample sizes, T = 500 and 2000. The choice of samples is adequate in the sense that in case of low-frequency financial data, the sample size of T = 1200 observations is corresponding with roughly five years of daily returns, whereas for high-frequency financial data, the sample size of T = 1200 observations is corresponding with fifteen days of five-minute intraday returns [for one trading day corresponds to 78 five-minute intraday returns].

In our GMM setting, we consider two sets of moments. One set contains the 24 moment conditions similar to the set of moments that consider by Andersen and Sørensen (1996) in the context of SV(1) estimation. They recommended using moment conditions for GMM estimation based on lower-order moments, since higher-order moments tend to exhibit erratic finite-sample behavior. The other set consider 6 moment conditions. The large and small sets are denoted by M_L and M_S and given by

$$M_L = \begin{pmatrix} |y_t|^j - \mu_j(\theta) \;\; for \;\; j = 1,\ldots,4 \\ |y_t||y_{t-j}| - \mu_{1,1}(j\,|\theta) \;\; for \;\; j = 1,\ldots,10 \\ y_t^2 y_{t-j}^2 - \mu_{2,2}(j\,|\theta) \;\; for \;\; j = 1,\ldots,10 \end{pmatrix} \quad \text{and} \quad M_S = \begin{pmatrix} |y_t|^j - \mu_j(\theta) \;\; for \;\; j = 1,\ldots,4 \\ y_t^2 y_{t-j}^2 - \mu_{2,2}(j\,|\theta) \;\; for \;\; j = 1,2 \end{pmatrix},$$

where $\mu_1(\theta) := \mathbb{E}(|y_t|) = \sigma_y(2/\pi)^{1/2}\exp[\gamma_0/8]$, $\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2\exp[\gamma_0/2]$, $\mu_3(\theta) := \mathbb{E}(|y_t|^3) = 2\sigma_y^3(2/\pi)^{1/2}\exp[9\gamma_0/8]$, $\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4\exp[2\gamma_0]$, $\mu_{1,1}(j\,|\theta) := \mathbb{E}(|y_t||y_{t-j}|) = \sigma_y^2(2/\pi)\exp\Big[\gamma_0(1+\rho_j)/4\Big]$, $j = 1,\ldots,10$ $\mu_{2,2}(j\,|\theta) := \mathbb{E}(y_t^2 y_{t-j}^2) = \sigma_y^4\exp\Big[\gamma_0(1+\rho_j)\Big]$, $j = 1,\ldots,10$, and $\gamma_0 = \sigma_v^2/(1-\sum_{j=1}^{2}\phi_j\rho_j)$. Furthermore, we consider two types of GMM estimators based on the choice of weighting matrix (inverse of the asymptotic covariance matrix and HAC covariance matrix using Bartlett Kernel).

Table 2 reports the estimation results for model $M1$. The results suggest that W-ARMA estimates are superior in terms of bias and RMSE whereas GMM, EDV [simple moment estimator that is the extension of Dufour and Valéry (2006)] and MCMC estimator performed poorly. The simple ARMA estimates also performed well. For each parameter, the smallest and the second smallest bias and RMSE has associated with W-ARMA and ARMA estimates. The third smallest RMSE for ϕ_1, ϕ_2, σ_y are associated with the efficient GMM estimator with 24-moments (GMM-24M-E) and for σ_v^2 is associated with EDV. We also find that GMM, EDV, and MCMC estimators are highly biased and the magnitude of the bias is substantial. Finally, for the larger samples, T=2000, we have almost identical results for all the parameter estimates as in the case of T = 500. Again W-ARMA outperforms all other estimators in terms of bias and RMSE. The RMSE of ARMA and W-ARMA estimates decreases as the sample size increases, shows the consistency of these estimators.

The results of $M2$, $M3$ and $M4$ models are reported in Tables 4-5 and results are very similar to the ones reported in Table 2. Again, the performance of simple ARMA and W-ARMA stands out in these designs. However, there are some important points to be noted down. In $M2$, simple ARMA estimator produces inadmissible parameter values (5 when T=500 and 1 when T=2000) for σ_v. In $M3$, we find that some of EDV and MCMC estimates are exploded with substantial bias and RMSE and this problem persists even when T=2000. Among GMM estimators, GMM-24M-E performs quite well in several cases, especially for ϕ_1 and ϕ_2. The Bayesian estimator did not perform well and ranked poorly compared to other estimators. However, in some cases (particularly for σ_v), the 6-moments GMM estimator performed better than the 24-moments GMM estimator.

Our results cast doubt on the advice that one should use a large number of moments. In this respect, one should not include too many moments thereby the chance of including irrelevant ones in the estimation procedure. This assertion is documented in the literature on asymptotic theory; see Buse (1992), Chao and Swanson (2007), and Dufour and Valéry (2006). In particular, overidentification results in biased GMM estimators in finite-samples. Concurring evidence based on finite-sample optimality results and Monte Carlo simulations is also available in Dufour and Taamouti (2003).

We also encounter the non-convergence problem with the Bayesian estimation. Furthermore, the EDV estimator of σ_v often produces a negative value. In these cases, we discarded those simulations from the calculation. So the bias and RMSE of EDV estimator are not comparable with other estimator. Note that in each simulations the W-ARMA estimator yields a solution. Several conclusions may be drawn from these simulation results for the W-ARMA and the simple ARMA estimator. First, when T = 500, these estimators provide accurate estimates since it outperforms all other estimators in terms of bias. Second, the W-ARMA is more efficient than other estimators in terms of RMSE for simulation design $M1$, $M2$ and $M3$. In the case of $M4$, the simple ARMA estimator is more efficient than the W-ARMA for some parameters. Third, when T = 2000, ARMA estimators uniformly outperformed all other estimators in terms of bias and RMSE, and these include the Bayesian estimator. Furthermore, from Table 6, the simple estimators are highly time efficient and the margin of time efficiency is huge compared to other estimators.

Our simulation results also show that the Bayesian method is very fragile and the convergence of this method depends on the choice of prior distribution. The specified prior [for the SV(2) model] of Chan and Grant (2016) produces a substantial bias for all four parameter estimates, indicating that their choice of prior distribution is terrible. Another well calibrated prior may reduce the magnitude of these biases, but it requires a considerable amount of work. Furthermore, the Bayesian method requires different prior distributions as the order of SV model changes, specifying an additional limitation of the Bayesian approach.

10. Empirical applications

In this section, we demonstrate two empirical applications of SV(p) models. First, we examine the fit of SV(p) models with real data to see the empirical evidence of this type of parametric models. Second, we further extend our analysis and compare the forecast performance of three typical volatility models in the daily out-of-sample experiments, models include the GARCH-type, SV-type and realized volatility based models that exploit high-frequency data.

10.1. Empirical evidence

The SV(p) models fitted to daily observations of the Standard and Poor's (S&P) Composite Price Index. The raw series is converted to returns by the transformation $r_t = 100[\log(P_t) - \log(P_{t-1})]$ and the returns are converted to residual returns by $y_t = r_t - \hat{\mu}_r$ where $\hat{\mu}_r$ is the sample average of returns. The sample period is from January 3, 1928 to September 27, 2016 and the number of observations is T = 23372. This data is obtained from Wharton Research Data Services (WRDS). The sample includes many volatile periods that cover the Great Depression (1929), the second world war (1937-45), the OPEC oil price shock (1973), the Black Monday (1987), the Asian financial crisis (1997), the early 2000s recession (Dot-com bubble), the late-2000s Financial Crisis (subprime mortgage crisis / United States housing bubble) and the recent Russian financial crisis (2014).

Table 7 reports summary statistics of the daily residual returns (y_t) and its several transformed series ($y_t^2, \log|y_t|, y_t^*$). We observe that the skewness and kurtosis of y_t and y_t^2 show the evidence of non-normal distribution, while the distributions of log-transformed residual returns are close to normal. This result is consistent with most empirical studies. Table 8 shows the parameter estimates of the SV(p) models using our ARMA-based winsorized estimator. To estimate ϕ_p, we used (5.28) with $J = 7$. Since we have a long financial time series, small values of J is more appropriate [according to simulation results]. Our result shows that there is some persistence in the volatility process during the period 1928-2016 and this is statistically significant. We also found that parameters of SV(p) models where $p = 2$ and $p = 3$ are statistically significant. This finding suggests that the latent volatility process can be treated as an autoregressive process of order more than one. This result is also consistent with Asai (2008). We also estimated an SV(4) model and several parameters of SV(4) model turned out to be insignificant. Again from Table 8, we can see that the ARMA-based estimator is extremely time efficient. We also have similar empirical estimates of SV(p) models for a smaller sample size (the sample period is from January 1, 1996 to September 27, 2016, and the number of observations is T = 5222).

In Table 8, we also reported the p-values that are based on the usual large-sample approximation based on HAC estimator. The variance-covariance $\hat{V}$ is estimated by a Bartlett kernel estimator with the bandwidth varying with the sample size, i.e. $m = [1.14T^{1/3}]$, where $[\cdot]$ denotes the integer part of the enclosed number; see Newey and West (1994). Note that, the asymptotic standard error can be markedly different and may be quite unreliable in finite-samples. To construct more reliable finite-sample inference, we can compute the Monte Carlo tests [discussed thoroughly in section 8] since our estimator is convenient for use in the context of computationally costly inference techniques.

We implemented parametric Bootstrap or Local Monte Carlo (LMC) tests as discussed in section 8 where we replace the nuisance parameters by their point estimate and simulate the test statistic under the null hypothesis. Results of LMC tests are reported in Table 9. Except for ϕ_1 and ϕ_2 parameters of the SV(3) model and σ_y in all models, we test each coefficient is zero against a right-sided alternative using a t-type test statistic. Note that, we cannot test $\phi_1 = 0$ and $\phi_2 = 0$ in SV(3) model since under each of these restrictions, we cannot simulate an SV(3) model since each of these restrictions leads to the latent volatility process non-stationary. In these cases, the ARMA-based estimation is infeasible. So we test $\phi_1 = 0.2$ and $\phi_2 = -0.4$ against a right-sided and a left-sided alternative, respectively. Further, we test $\sigma_y = 0.01$ against a right-sided alternative since when $\sigma_y = 0$, SV models are unidentified.

From the Table 9, we can see that the estimate of ϕ_4 of SV(4) model is not statistically significant in both asymptotic and finite-sample parametric Bootstrap tests, and this entails that an SV(3) model could be more suitable for the volatility dynamics of this sample periods. Note that, one can easily exploit these simple estimators and construct exact tests based on MMC procedure as discussed in Section 8. To summarize, the results presented here indicate that SV models with additional lag terms in the volatility process may be appropriate to model the S&P 500 index.

10.2. Volatility forecast performance

We evaluate the volatility forecast performance amongst GARCH, SV, and realized volatility based models.[4] Volatility has long been modeled and forecasted using GARCH models because of the earlier discussed complexity of SV models. We considered several popular GARCH-type models in our experiments, these include: GARCH models of Bollerslev (1986), Exponential GARCH (EGARCH) models of Nelson (1991) and GJR models of Glosten, Jagannathan and Runkle (1993). For the details of these models and their forecast equations, see appendix A.4.

We also consider the realized volatility based models, specially Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models of Corsi (2009). In this study, we use a logarithmic version of HAR-RV model since the logarithmic transformation of RV appears approximately Gaussian [see Andersen, Bollerslev, Diebold and Ebens (2001), Andersen, Bollerslev, Diebold and Labys (2001)]. The HAR-RV model takes into account the long memory feature of realized volatility, and among the models proposed to forecast realized volatility, it stands out because of its simplicity (for details, see appendix A.6).

For SV(p) models, we exploit the state-space representation which is given by (2.14-2.15) and calculate forecasts based on the Kalman filter. The SV(p) parameters are computed using our simple method where we used (5.28) with $J = 50$ and fixed the value of J before any estimations. Given the simple estimates, we computed the forecasts of SV(p) models through the Kalman filter. For the details of this forecasting

[4] Realized volatility (RV), is a model free volatility, received much attention among the financial economists and econometricians as an accurate measure of the true latent volatility under the ideal market assumption [Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2001)] and it can be used as a proxy for true latent volatility (for details, see appendix A.5).

procedure and out-of-sample forecasting equations, see appendix A.3.

We use three loss measures to evaluate the forecast accuracy, these include MSE, MAE, and R2LOG. MSE and MAE are the mean squared error and mean absolute error, respectively, and R2LOG is the logarithmic loss function proposed by Pagan and Schwert (1990) and can penalize volatility forecast asymmetry in high and low level of volatility. These loss measures are defined by

$$MSE : l_t = (\hat{\sigma}_t^2 - h_{t|t-k}^2)^2, \quad MAE : l_t = \mid \hat{\sigma}_t^2 - h_{t|t-k}^2 \mid, \quad R2LOG : l_t = (\log \hat{\sigma}_t^2 - \log h_{t|t-k})^2,$$

where $\hat{\sigma}_t^2$ is an unbiased *ex-post* proxy of conditional variance (such as squared return or realized volatility) and $h_{t|t-k}$ is a volatility forecast based on $t-k$ information set where $k > 0$.

Using the above loss functions, we also compute the model confidence set (MCS) procedure proposed by Hansen et al. (2011). The model confidence set involves a sequence of tests for equal predictive ability (EPA) hypothesis. Given a model set M_0, which contains m competing forecast models, the null hypothesis is that all models in M_0 have equal predictive accuracy. If the null hypothesis is rejected at a given confidence level α, then the worst performing model in M_0 is eliminated. After that, the EPA test is repeated until the null hypothesis is accepted. When the null hypothesis is accepted, the remainder composes $1-\alpha$ confidence set, $\hat{M}^*_{1-\alpha}$.

We now briefly discuss how it is implemented. Define the relative loss differential between models by

$$d_{i,j,t} = l_{i,t} - l_{j,t} \quad \text{for all} \;\; i, j \in M, \;\; t = 1, \ldots, T,$$

be the simple loss of model i relative to any other model j at time t. Using the loss differential between competing models, the MCS procedure tests the EPA hypothesis in two alternative ways.

$$H_0 : \mu_{ij} = 0 \quad \text{for all} \;\; i, j \in M \quad \text{and} \quad H_A : \mu_{ij} \neq 0 \quad \text{for some} \;\; i, j \in M \tag{10.1}$$

or

$$H_0 : \mu_{i,\cdot} = 0 \quad \text{for all} \;\; i \in M \quad \text{and} \quad H_A : \mu_{i,\cdot} \neq 0 \quad \text{for some} \;\; i \in M \tag{10.2}$$

where $\mu_{ij} = \mathbb{E}(d_{i,j})$ and $\mu_{i,\cdot} = \mathbb{E}(d_{i,\cdot})$. The two statistics, used in the model confidence set test, are expressed as follows:

$$MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid \quad \text{and} \quad MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}, \tag{10.3}$$

where $t_{i,j} = \frac{\bar{d}_{i,j}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{i,j})}}$, $t_{i,\cdot} = \frac{\bar{d}_{i,\cdot}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{i,\cdot})}}$, $\bar{d}_{i,\cdot} = m^{-1} \sum_{j \in M} \bar{d}_{i,j}$, and $\bar{d}_{i,j} = T^{-1} \sum_{t=1}^{T} d_{i,j,t}$ for $i, j \in M$, while $\widehat{\text{Var}}(\bar{d}_{i,\cdot})$ and $\widehat{\text{Var}}(\bar{d}_{i,j})$ are bootstrapped estimates of $\text{Var}(\bar{d}_{i,\cdot})$ and $\text{Var}(\bar{d}_{i,j})$, respectively. In our calculations, we perform a block–bootstrap [block length 12 is used] procedure of 10000 re-samples. The first statistic, $t_{i,j}$, is used in the well-known test for comparing two forecasts; see Diebold and Mariano (2002) and West (1996), while the second one, $t_{i,\cdot}$, is used in Hansen, Lunde and Nason (2003), Hansen (2005), and Hansen et al. (2011).

We conduct two out-of-sample forecast experiments using different volatility proxy:

1. **Design 1 (Moderate volatility regimes)**: In this setting, we consider a sample period, from September 01, 2005 to August 31, 2010. The in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. In this setting, we forecast a moderately volatile period but the in-sample contains the most volatile part of the late-2000s Financial Crisis.

2. **Design 2 (High volatility regimes)**: In this design, we consider a sample period, from January 01, 2005 to December 31, 2009. The in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. The out-of-sample include the late-2000s Financial Crisis (Subprime mortgage crisis / United States housing bubble). In this setting, we forecast a highly volatile period.

In both designs, we consider a sample of five years that divided into three years span of in-sample and two years span of out-of-sample. Three years span for the in-sample window is adequate for finding the most accurate volatility forecasts; see Kambouroudis and McMillan (2015).

Within the SV and GARCH framework, the key element is the specification for conditional variance. Parametric SV and GARCH models utilize daily returns (typically squared returns) to extract information about the current level of volatility, and this information is used to form expectations about the next period's volatility. The daily squared return is a conditionally unbiased estimator of the daily conditional variance. However, as pointed out by Andersen and Bollerslev (1998), although the use squared returns is justified because it is a conditionally unbiased estimator of the daily conditional variance, it provides a noisy measure. They suggest that realized volatility based on cumulative intraday squared returns is a more accurate proxy for true latent volatility. The literature on constructing consistent volatility proxy using realized volatility measures is considerable. The proposed methods include Maximum likelihood estimator [Aït-Sahalia, Mykland and Zhang (2005)], Quasi-maximum likelihood estimator [Xiu (2010)], Two Scales Realized Volatility (TSRV) [Zhang, Mykland and Aït-Sahalia (2005)], Multi-Scale Realized Volatility (MSRV) [Zhang (2006)], Realized Kernels (RK) [Hansen and Lunde (2006), Barndorff-Nielsen, Hansen, Lunde and Shephard (2008, 2011)] and Pre-Averaging volatility estimator (PAV) [Jacod, Li, Mykland, Podolskij and Vetter (2009)]. Other relevant references include Bandi and Russell (2006), Fan and Wang (2007), Gatheral and Oomen (2010), Kalnina and Linton (2008), Li and Mykland (2007), Aït-Sahalia, Mykland and Zhang (2011). So, we assessed out-of-sample volatility forecasts across competing models using different loss functions as well as MCS procedure with squared return and realized volatility proxies over a range of forecast horizons.

We computed out-of-sample forecasts using Rolling (moving) window method and computed for forecast horizon $1-$day, $2-$day, $1-$week, $2-$week, $3-$week and $1-$month. In this rolling forecasts setup, an initial sample using data from $t = 1,\dots,T$ is used to determine a window width T, to estimate the models, and to form $h-$step ahead out-of-sample forecasts starting at time T. Then the window is moved ahead one time period, the models are re-estimated using data from $t = 2,\dots,T+1$, and $h-$step ahead out-of-sample forecasts are produced starting at time $T+1$. This process is repeated until no more $h-$step ahead forecasts can be computed.

10.2.1. Forecasting squared return

Now we consider the daily squared return as volatility proxy and evaluate the volatility forecast performance amongst GARCH, SV, and HAR-RV models using the S&P 500 index. The high-frequency RV estimates and prices of S&P 500 index are obtained from the Oxford-Man Institute's Realized Library [Heber, Lunde, Shephard and Sheppard (2009)]. The raw prices are converted to returns by the transformation $r_t = 100[\log(P_t) - \log(P_{t-1})]$. The returns are converted to residual returns by $y_t = r_t - \hat{\mu}_r$ where $\hat{\mu}_r$ is the sample average of returns. Note that, y_t^2 is our volatility proxy at time t.

We consider a modified version of the HAR-RV model [that defined in (A.1)], where the daily squared return is the dependent variable. However, there are problems in measuring daily realized volatility measures from high-frequency returns. First, the *discretization error* in the estimates of volatility due to the fact that we only observe prices at intermittent and discrete points in time. Second, and more importantly, the *market microstructure noise* due to bid/ask bounces, the different price impact of different types of trades, limited liquidity, or other types of market frictions. So the choice of RV estimator is important and we consider other estimators of RV and these include: realized bipower variation (BV) [Barndorff-Nielsen and Shephard (2006)], realized Semi-variance (RSV) [Barndorff-Nielsen, Kinnebrock and Shephard (2010)], Realized Kernel (RK) [Barndorff-Nielsen, Hansen, Lunde and Shephard (2008, 2011)]. We also consider subsampled versions of all the estimators of RV (except for the RK, since it is using tick-by-tick data, it cannot be subsampled). Subsampling, introduced by Zhang et al. (2005), is a simple way to improve the efficiency of some sparse-sampled estimators.[5]

For each forecast experiments, we obtain forecasts from three SV models, eight GARCH-type models, and nine HAR-RV type models. The eight GARCH-type models are: GARCH$(1,1)$, GARCH$(1,2)$, GARCH$(2,1)$, GARCH$(2,2)$, EGARCH$(1,1)$, EGARCH$(2,2)$, GJR$(1,1)$ and GIR$(2,2)$.

For S&P 500, in design 1, the sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1257$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 504$). In design 2, the sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1258$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 504$).

Tables 10-11 report summary statistics of daily variables, high-frequency RV estimators, and their logarithmic transformations. Using our out-of-sample forecasts, we calculated forecast evaluation measures, i.e., MSE, MAE and R2LOG. Tables 12-17 presents the main results of our forecasting experiments. For easy comparison, we report the relative MSE, the relative MAE and the relative R2LOG of forecast error. These are relative to the reference model HAR-RV5, and hence, values smaller than unity indicate better forecast performance than the HAR-RV5 model. Furthermore, we also reported the MCS p-value for the corresponding model.

In design 1, Tables 12-14, when we forecast a moderately unstable period after the core Financial Crisis, the forecasting performance of higher-order SV models [especially the SV(3) model] are superior to all other volatility models. This result is consistent across different forecast horizons, different evaluation measures and based on MCS. According to MCS, the SV(3) model dominates all other competing models except for 1, 2, and 3-weeks horizon when using MSE loss function and 1-month when using MAE and R2LOG loss function. In this setting, the SV(2) model is the superior forecasting model by the MAE and R2LOG. Several HAR-RV models performed well according to MSE, but this result is undermined by their performance in terms of MAE and R2LOG. However, the forecasting performance of GARCH-type models is worst among all models according to all performance measures and across different forecast horizons.

In design 2, Tables 15-17, when we forecast a highly volatile period, i.e., the core Financial Crisis, the findings are following: in most cases, SV(p) models perform better than other competing models [this holds across different forecast horizons and different evaluation measure] while HAR-RV models performed better than GARCH models. The SV(3) model produces the superior forecast in terms of MSE criteria in horizon 1-day, 2-day, 3-weeks, and 1-month. Furthermore, SV(p) models [SV(3) or SV(2)] are the top forecasting models based on MCS p-value when using MAE and R2LOG. The forecasting performance of GARCH and HAR-RV models is poor according to R2LOG. These models may produce asymmetric forecast errors because R2LOG heavily penalized asymmetry in a high and low level of volatility. However, HAR-RV models performed better than GARCH models according to RMSE, but GARCH models outperformed HAR-RV models according to MAE and R2LOG, this implies that GARCH models produce large forecast errors because MSE heavily penalized any outlier.

[5]Subsampling involves using a variety of “grids" of prices sampled at a given frequency to obtain a collection of realized measures, which are then averaged to yield the “subsampled" version of the estimator. For example, 5-minute RV can be computed using prices sampled at 10:30, 10:35, etc. and can also be computed using prices sampled at 10:31, 10:36, etc.

In both settings, among HAR models, those that based on the subsampled version of RV estimators produces identical forecasts, implies that subsampling is not improving any forecast performance. Further, the performance of HAR models is inferior among all models in long-horizon. Note that in both designs, the Financial Crisis is included either in-sample or out-of-sample. During this time, the financial market is unstable, and the high-frequency RV estimators are affected by the large market microstructure noise. The forecasting performance of HAR-RV models may be affected by these noisy RV estimators.

From Tables 12-17, we can see that except for a few instances, higher-order SV models perform better than GARCH-type and HAR-RV type models not only in all evaluation measures but also across different volatility regimes and horizons. In both of our out-of-sample experiments, higher-order SV models also outperform the first-order SV model. This finding suggests that adding additional lag terms in the latent volatility equation are essential for forecasting volatility.

10.2.2. Forecasting realized volatility

In previous forecast experiments, we forecast the daily squared return which is a noisy proxy for the true latent volatility. Now, we consider realized volatility as a volatility proxy since it is an accurate measure of the true latent volatility [see Andersen and Bollerslev (1998), Barndorff-Nielsen and Shephard (2001)]. In this section, we compare the performance of three SV models to the HAR-RV model. In realized volatility literature, several RV estimates have been proposed, but following the results of Liu, Patton and Sheppard (2015), we use the 5-minute RV which is constructed from five-minute intraday returns. In case of SV(p) models, we replace the squared return by realized volatility [squared return is considered as the observed process in SV models] and then estimate the models by our ARMA-based estimator.

We consider five assets; these include S&P 500, FTSE100, NASDAQ100, N225, SSMI20 indices and their 5-minute realized volatilities are sourced from the Oxford-Man Institute's Realized Library. Main results of these forecast experiments are reported in Tables 18-23. For easy comparison, we report the relative MSE, the relative MAE, and the relative R2LOG of forecast error. These are relative to the HAR-RV model, and hence, values smaller than unity indicate better forecast performance than the HAR-RV model. Furthermore, we also reported the MCS p-value for the corresponding model, and the best model is highlighted by boldface color font.

In design 1, Tables 18-20, when we forecast a moderately volatile period after the financial crisis, in most cases, higher-order SV models [SV(2) or SV(3)] provide superior forecasts. This finding is consistent across different evaluation measures. Out of 30 cases [across five assets and six forecast horizons], SV models delivered the best forecast performance in almost in all cases [according to MSE, MAE, R2LOG], except for the 1-week ahead forecasting of SSMI20 volatility in terms of MSE. The forecasting performances of higher-order SV models are ranked top in 80%, 97%, and 97% of cases according to MSE, MAE, R2LOG, respectively. If we consider only longer horizons [2-week, 3-week, 1-month] then these winning percentages of SV(p) models are increased to 87%, 100% and 100%, while for short horizons [1-day, 2-day, 1-week] these percentages are 73%, 93%, and 93%. One out of ninety cases [across five assets, six forecast horizons, and three loss measures], the HAR-RV model produces the best forecasting performance.

Between the SV(2) and SV(3) model, the performance of the SV(2) model is better in short forecast horizons but in long horizons, the performance SV(3) model is better. This finding tells us that additional lag term is essential for forecasting realized volatility in long horizons. Furthermore, compared to all other models, the forecasting performance of the SV(3) model is getting better as the forecast horizon increases. Note that the MCS p-values of other competing models declined significantly in long horizons. The performance of HAR-RV is pointedly poor compared to SV(p) models in long horizons since the relative loss measures of SV(p) models are now lower.

In design 2, Tables 21-23, when we forecast highly volatile periods such as financial crisis or expansions, the ranking of models are similar to design 1. Out of ninety cases [across five assets, six forecast horizons, and three loss measures], HAR-RV model produced the best forecasts in 3% of cases whereas SV(p) models delivered best forecasts in 86% of cases.

In both designs, our findings suggest that SV(p) models are better in forecasting realized volatility. So fitting non-parametric volatility measure in traditional parametric models can provide better forecasting performance. This also tells us that the HAR-RV model is not capturing the proper mean dynamics that comes from the moving average part of the market microstructure noise during the financial crisis. As pointed out by Meddahi (2003), that if several factors influence the dynamics of RV, then RV follows an ARMA-type process. In this study, within a parametric SV framework, we modeled realized volatility as a non-Gaussian ARMA process [see Lemma 5.4].

11. Conclusion

In this paper, we propose several estimators for higher-order SV models, and these include computationally simple estimators and GMM-type estimators. The motivation, as well as the stationarity, ergodicity and mixing properties of SV(p) models, are thoroughly discussed. Furthermore, the study develop recursive estimation procedures for SV(p) models using simple estimators and derive asymptotic distributions of these simple estimators. We show that simple estimators are especially convenient for use in the context of simulation-based inference techniques, i.e., Bootstrap or Monte Carlo tests.

In simulations, we compare our proposed estimators to the Bayesian estimator. The simple winsorized ARMA-based estimator uniformly outperformed all other estimators in terms of bias and statistical efficiency. This conclusion holds across different simulation designs. Furthermore, the simple estimators are highly time efficient compared to other estimators. We also find that both GMM estimators and the Bayesian estimator have the non-convergence problem while the EDV estimator sometimes produces a negative value for the variance of the volatility innovation.

Our results cast doubt on the use of a large number of moments. In this respect, one should not include too many instruments since it can increase the chance of including irrelevant ones in the estimation procedure. In particular, over-identification increases bias of GMM estimators in finite-samples. Concurring evidence based on finite-sample optimality results and Monte Carlo simulations is also available in Dufour and Taamouti (2003). In an optimal GMM setting, the number of moment equations should be equal to the number of parameters, provided that these moments are well selected. These type of GMM estimators are efficient and good at forecasting. In that sense, the ARMA-based estimator, based on a few moments, is nearly optimal and can be view as a parsimonious moment-based (or GMM) estimator.

We also present empirical applications relating to SV(p) models and the simple ARMA-based estimator. First, we find that S&P 500 returns can be better modeled as an SV(3) model and this result confirms by both asymptotic and finite-sample tests. Second, we conduct out-of-sample forecasting experiments to forecast the daily squared return as well as the realized volatility. We compare the forecasting performance among SV, GARCH, and HAR-RV models. Our results suggest that higher-order SV models perform better than GARCH and HAR-RV type models in most cases. This finding holds even if a high volatility period (such as Financial Crisis) is included in the estimation sample or the forecasted sample. These inferences are not only based on a standard forecasting precision assessment [such as using MSE and MAE statistics] but also based on formal prediction tests, i.e., the MCS procedure of Hansen et al. (2011). These findings highlight the usefulness of higher-order SV models for volatility forecasting.

A. Appendix

A.1. Proofs

PROOF OF LEMMA 3.1 Under the assumption 3.1, we have an MFSV model where the volatility process is driven by the sum of m independent AR(1) process. Granger and Morris (1976) shown that the sum of m independent AR(1) processes is an ARMA$(m,m-1)$ process. The proof follows from there. Note that Meddahi (2003) derived ARMA representation of integrated and realized variances when the spot variance depends linearly on two autoregressive factors. This class of processes includes affine, GARCH diffusion, as well as the eigenfunction stochastic volatility and the positive Ornstein-Uhlenbeck models. □

PROOF OF LEMMA 3.2 We consider an SV-ARMA(p,q) model defined by Lemma 3.1 with the latent volatility process driven by an ARMA(p,q) such that

$$\alpha(L)w_t = \beta(L)\sigma_v v_t$$

where

$$\alpha(L) = 1-\alpha_1 L-\alpha_2 L^2-\ldots-\alpha_p L^p, \quad \beta(L) = 1-\beta_1 L-\beta_2 L^2-\ldots-\beta_q L^q$$

and where the innovations $\{v_t\}$ form a stationary, ergodic sequence such that, for the $\sigma-$algebra $\boldsymbol{F}_{t-1}$ generated by $\{v_\tau, \tau \leq t-1\}$, $\mathbb{E}(v_t\,|\boldsymbol{F}_{t-1}) = 0$ almost surely, $\mathbb{E}(v_t^2\,|\boldsymbol{F}_{t-1}) = 1 > 0$ almost surely, and $\mathbb{E}(\varepsilon_t^4) < \infty$. Since the roots of the moving average polynomial are lie outside the unit circle [the model is invertible], and there exists an infinite-order autoregressive representation of the latent volatility process such that

$$\sum_{j=0}^{\infty}(-\phi_j)w_{t-j} = \sigma_v v_t. \tag{A.1}$$

□

PROOF OF LEMMA 5.1 Under the assumptions 2.1 − 2.2, and if $U \sim N(0,1)$, then $\mathbb{E}(U^{2p+1}) = 0,\ \forall p \in \mathbb{N}$ and $\mathbb{E}(U^{2p}) = \frac{2p!}{2^p p!},\ \forall p \in \mathbb{N}$. Then

$$\begin{aligned}
\mu_k(\theta) := \mathbb{E}(y_t^k) &= \sigma_y^k \mathbb{E}(z_t^k)\mathbb{E}\left[\exp(\frac{kw_t}{2})\right] \\
&= \sigma_y^k \frac{k!}{2^{k/2}(k/2)!}\exp\left[\frac{k^2}{8}\mathrm{Var}(w_t)\right] \\
&= \sigma_y^k \frac{k!}{2^{k/2}(k/2)!}\exp\left[\frac{k^2}{8}\frac{\sigma_v^2}{1-\sum_{j=1}^{p}\phi_j\rho_j}\right], \quad if\ k\ is\ even \\
&= 0, \quad if\ k\ is\ odd
\end{aligned} \tag{A.2}$$

where $\rho_j := \text{corr}(w_t, w_{t+j})$ and for the cross-moment we have:

$$\begin{aligned}
\mu_{k,l}(m|\theta) &:= \mathbb{E}(y_t^k y_{t+m}^l) = \sigma_y^{k+l}\mathbb{E}(z_t^k)\mathbb{E}(z_{t+m}^l)\mathbb{E}\left[\exp(\frac{kw_t}{2}+\frac{lw_{t+m}}{2})\right] \\
&= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\left[\exp(\frac{k^2}{8}\text{Var}(w_t)+\frac{l^2}{8}\text{Var}(w_{t+m})+\frac{2kl}{8}\text{cov}(w_t,w_{t+m}))\right] \\
&= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\left[\exp(\frac{k^2}{8}\gamma_0+\frac{l^2}{8}\gamma_0+\frac{2kl}{8}\gamma_0\rho_m)\right] \\
&= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\exp\left[\frac{1}{8}\gamma_0(k^2+l^2+2kl\rho_m)\right] \\
&= \sigma_y^{k+l}\frac{k!}{2^{k/2}(k/2)!}\frac{l!}{2^{l/2}(l/2)!}\exp\left[\frac{1}{8}\frac{\sigma_v^2}{1-\sum_{j=1}^p\phi_j\rho_j}(k^2+l^2+2kl\rho_m)\right], \\
&\quad if\ k\ and\ l\ are\ even \\
&= 0, \quad if\ k\ or\ l\ are\ odd.
\end{aligned} \tag{A.3}$$

□

PROOF OF LEMMA 5.2 Using Lemma 5.1 and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$ and $k = l = 2$ & $m = 2$, we get:

$$\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2\exp\left[\frac{1}{2}\gamma_0\right] \tag{A.4}$$

$$\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4\exp[2\gamma_0] \tag{A.5}$$

$$\mu_{2,2}(1|\theta) := \mathbb{E}(y_t^2y_{t-1}^2) = \sigma_y^4\exp[\gamma_0(1+\rho_1)] \tag{A.6}$$

$$\mu_{2,2}(2|\theta) := \mathbb{E}(y_t^2y_{t-2}^2) = \sigma_y^4\exp[\gamma_0(1+\rho_2)] \tag{A.7}$$

where $\gamma_0 = \sigma_v^2/(1-\sum_{j=1}^2\phi_j\rho_j)$. From (A.4) and (A.5), we get:

$$\frac{\mathbb{E}(y_t^4)}{(\mathbb{E}(y_t^2))^2} = 3\exp(\gamma_0)$$

or,

$$\gamma_0 = log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2})$$

From (A.4),

$$\sigma_y^2 = \frac{\mathbb{E}(y_t^2)}{\exp(\gamma_0/2)}$$

or,

$$\sigma_y = \frac{(\mathbb{E}(y_t^2))^{1/2}}{(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2})^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{(\mathbb{E}(y_t^4))^{1/4}} \tag{A.8}$$

From (A.4) and (A.6), we get:

$$\gamma_0\rho_1 = log(\frac{\mathbb{E}(y_t^2y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})$$

$$\gamma_1 = log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}) = log(\mathbb{E}(y_t^2 y_{t-1}^2)) - 2log(\mathbb{E}(y_t^2)).$$

Similarly from (A.4) and (A.7), we get:

$$\gamma_2 = log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2}) = log(\mathbb{E}(y_t^2 y_{t-2}^2)) - 2log(\mathbb{E}(y_t^2)).$$

Now under the assumptions 2.2, the latent volatility process satisfy the Yule-Walker equations, see Fuller (1996) or Box, Jenkins and Reinsel (2013). So the auto-covariance and autocorrelation of the volatility process satisfy the following equations respectively:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + (\sigma_v)^2$$

$$\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1$$

$$\gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0$$

Solving for ϕ_1 and ϕ_2 as functions of the autocovariances:

$$\phi_1 = \frac{-\gamma_1(\gamma_2-\gamma_0)}{(\gamma_0)^2-(\gamma_1)^2}, \quad \phi_2 = \frac{-(\gamma_1)^2+\gamma_2\gamma_0}{(\gamma_0)^2-(\gamma_1)^2}$$

Substitute the values of γ_0, γ_1, and γ_2 into the Yule-Walker equations, we get:

$$\phi_1 = \frac{-(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))(log(\frac{3\mathbb{E}(y_t^2 y_{t-2}^2)}{\mathbb{E}(y_t^4)}))}{(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2 - (log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))^2}, \tag{A.9}$$

$$\phi_2 = \frac{-(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))^2 + (log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2}))(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))}{(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2 - (log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))^2} \tag{A.10}$$

$$\sigma_v = [log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) - \phi_1(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})) - \phi_2(log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2}))]^{1/2}. \tag{A.11}$$

□

PROOF OF LEMMA 5.3 Here we derive the covariances of the components of

$$\begin{aligned}\zeta_1(\tau) &= Cov(X_{1,t}, X_{1,t+\tau}) = \mathbb{E}([y_t^2 - \mu_2(\theta)][y_{t+\tau}^2 - \mu_2(\theta)]) \\ &= \mathbb{E}(y_t^2 y_{t+\tau}^2) - \mu_2^2(\theta) = \sigma_y^4 \exp[\gamma_0(1+\rho_\tau)] - \mu_2^2(\theta) = \mu_2^2(\theta)[\exp(\gamma_\tau) - 1]\end{aligned} \tag{A.12}$$

where $\gamma_j := \text{cov}(w_t, w_{t+j})$. Similarly,

$$\begin{aligned}\zeta_2(\tau) &= Cov(X_{2,t}, X_{2,t+\tau}) = \mathbb{E}[y_t^4 - \mu_4(\theta)][y_{t+\tau}^4 - \mu_4(\theta)] = \mathbb{E}(y_t^4 y_{t+\tau}^4) - \mu_4^2(\theta) \\ &= 9\sigma_y^8 \exp[4\gamma_0(1+\rho_\tau)] - \mu_4^2(\theta) = \mu_4^2(\theta)[\exp(4\gamma_\tau) - 1], \quad \forall \tau \geq 1\end{aligned} \tag{A.13}$$

and

$$\begin{aligned}\zeta_3(\tau) &= Cov(X_{3,t}, X_{3,t+\tau}) = \mathbb{E}([y_t^2 y_{t-1}^2 - \mu_{2,2}(1|\theta)][y_{t+\tau}^2 y_{t+\tau-1}^2 - \mu_{2,2}(1|\theta)]) \\ &= \mathbb{E}[y_t^2 y_{t-1}^2 y_{t+\tau}^2 y_{t+\tau-1}^2] - \mu_{2,2}^2(1|\theta) \\ &= \sigma_y^8 \mathbb{E}[\exp(w_{t-1} + w_{t+\tau} + w_t + w_{t+\tau-1})] - \mu_{2,2}^2(1|\theta) \\ &= \sigma_y^8 \exp[2(\gamma_0+\gamma_1) + \gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}] - \mu_{2,2}^2(1|\theta) \\ &= \sigma_y^8 \exp[2(\gamma_0+\gamma_1)]\exp[\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}] - \mu_{2,2}^2(1|\theta) \\ &= \mu_{2,2}^2(1|\theta)[\exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}) - 1], \quad \forall \tau \geq 2.\end{aligned} \tag{A.14}$$

Finally,

$$\begin{aligned}\zeta_4(\tau) &= Cov(X_{4,t}, X_{4,t+\tau}) = \mathbb{E}([y_t^2 y_{t-2}^2 - \mu_{2,2}(2|\theta)][y_{t+\tau}^2 y_{t+\tau-2}^2 - \mu_{2,2}(2|\theta)]) \\ &= \mathbb{E}[y_t^2 y_{t-2}^2 y_{t+\tau}^2 y_{t+\tau-2}^2] - \mu_{2,2}^2(2|\theta) \\ &= \sigma_y^8 \mathbb{E}[\exp(w_{t-2} + w_{t+\tau} + w_t + w_{t+\tau-2})] - \mu_{2,2}^2(2|\theta) \\ &= \sigma_y^8 \exp[2(\gamma_0+\gamma_2) + \gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}] - \mu_{2,2}^2(2|\theta) \\ &= \sigma_y^8 \exp[2(\gamma_0+\gamma_2)]\exp[\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}] - \mu_{2,2}^2(2|\theta) \\ &= \mu_{2,2}^2(2|\theta)[\exp(\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}) - 1], \quad \forall \tau \geq 3.\end{aligned} \tag{A.15}$$

□

PROOF OF PROPOSITION 5.4 From (2.14) - (2.15), we have

$$\phi(B) w_t = v_t,$$

and

$$y_t^* = w_t + \varepsilon_t,$$

where $\phi(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$. The error processes v_t's and ε_t's are i.i.d. $N(0, \sigma_v^2)$ and $i.i.d.\ \log\chi^2_{(1)}$ random variables, respectively. Furthermore, assumption 2.1 implies that v_t's and ε_t's are independent. on applying $\phi(B)$ to both sides of (2.15) yields

$$\phi(B) y_t^* = \phi(B) w_t + \phi(B)\varepsilon_t = v_t + \phi(B)\varepsilon_t. \tag{A.16}$$

The right hand side of (A.16) is clearly a covariance stationary process. By the Wold decomposition theorem it must have a moving average representation. Since the autocovariance function cuts off for lags $k > p$ it must be an $MA(p)$ process, say $\theta(B)\eta_t = (1 - \theta_1 B - \cdots - \theta_p B^p)\eta_t$. Hence, y_t^* must be an $ARMA(p, p)$ process [see equation (2.1) of Granger and Morris (1976)].
The moving average parameters $\theta_1, \theta_2, \ldots, \theta_p$ and the white noise variance σ_η^2 of this $ARMA(p, p)$ process can be found by equating the autocovariance function of the right hand side of (A.16) with that of $\theta(B)\eta_t$ for lags $k = 0, 1, \ldots, p$ and solving the $p+1$ resulting non-linear equations

$$\begin{aligned}(1 + \theta_1^2 + \cdots + \theta_p^2)\sigma_\eta^2 &= \sigma_v^2 + (1 + \phi_1^2 + \cdots + \phi_p^2)\sigma_\varepsilon^2 \\ (-\theta_1 + \theta_1\theta_2 + \cdots + \theta_{p-1}\theta_p)\sigma_\eta^2 &= (-\phi_1 + \phi_1\phi_2 + \cdots + \phi_{p-1}\phi_p)\sigma_\varepsilon^2 \\ &\vdots\end{aligned}$$

$$(-\theta_{p-1}+\theta_1\theta_p)\sigma_\eta^2 = (-\phi_{p-1}+\phi_1\phi_p)\sigma_\varepsilon^2$$
$$-\theta_p\sigma_\eta^2 = -\phi_p\sigma_\varepsilon^2.$$

Note that there may be multiple solutions, only some of which result in an invertible process. □

PROOF OF COROLLARY 5.5 From Proposition 5.4, the observed process $\{y_t^*\}$ satisfies the following equation:

$$y_t^* = \sum_{j=1}^{p}\phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{p}\theta_j\eta_{t-j}, \tag{A.17}$$

or,

$$y_t^* = \sum_{j=1}^{p}\phi_j y_{t-j}^* + v_t + \varepsilon_t - \sum_{j=1}^{p}\phi_j\varepsilon_{t-j}. \tag{A.18}$$

Multiply both sides of (A.18) by y_{t-k}^* and taking expectation we get:

$$\gamma_{y^*}(k) = \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-1) + \mathbb{E}[v_t y_{t-k}^*] + \mathbb{E}[\varepsilon_t y_{t-k}^*] - \sum_{j=1}^{p}\phi_j\mathbb{E}[\varepsilon_{t-j}y_{t-k}^*].$$

Setting $k=0$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + \mathbb{E}[v_t y_t^*] + \mathbb{E}[\varepsilon_t y_t^*] - \sum_{j=1}^{p}\phi_j\mathbb{E}[\varepsilon_{t-j}y_t^*] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2 - \sum_{j=1}^{p}\phi_j\mathbb{E}[\varepsilon_{t-j}(\phi_j y_{t-j}^* - \phi_j\varepsilon_{t-j})] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2 - \sum_{j=1}^{p}\phi_j^2\mathbb{E}[\varepsilon_{t-j}y_{t-j}^* - \varepsilon_{t-j}^2] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2 - \sum_{j=1}^{p}\phi_j^2[\sigma_\varepsilon^2 - \sigma_\varepsilon^2] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2.
\end{aligned} \tag{A.19}$$

Setting $1 \le k \le p$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + \mathbb{E}[v_t y_{t-k}^*] + \mathbb{E}[\varepsilon_t y_{t-k}^*] - \sum_{j=1}^{p}\phi_j\mathbb{E}[\varepsilon_{t-j}y_{t-k}^*] \\
&= \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) + 0 + 0 - \phi_k\mathbb{E}[\varepsilon_{t-k}y_{t-k}^*] = \sum_{j=1}^{p}\phi_j\gamma_{y^*}(k-j) - \phi_k\sigma_\varepsilon^2.
\end{aligned} \tag{A.20}$$

Setting $k > p$, we get

$$\begin{aligned}\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \mathbb{E}[v_t y^*_{t-k}] + \mathbb{E}[\varepsilon_t y^*_{t-k}] - \sum_{j=1}^{p} \phi_j \mathbb{E}[\varepsilon_{t-1} y^*_{t-k}] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + 0 + 0 - 0 = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j).\end{aligned} \tag{A.21}$$

Combining (A.19), (A.20), and (A.21), we get the autocovariance structure of the observed process that stated in the Corollary. □

PROOF OF COROLLARY 5.6 The estimator of $\phi_{\boldsymbol{p}}$ is based on the autocovariance structure of the process y^*_t. This is the solution of p-system of equations from (5.18) with $k = p+1, \ldots, 2p$. So

$$\begin{bmatrix} \gamma_{y^*}(k) & \gamma_{y^*}(k-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(k+1) & \gamma_{y^*}(k) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(k+p-1) & \gamma_{y^*}(k+p-2) & \cdots & \gamma_{y^*}(k) \end{bmatrix} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma_{y^*}(k+1) \\ \gamma_{y^*}(k+2) \\ \vdots \\ \gamma_{y^*}(k+p) \end{bmatrix},$$

or,

$$\boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})} . \phi_{\boldsymbol{p}} = \gamma_{(\boldsymbol{k},\boldsymbol{p})}$$

or,

$$\phi_{\boldsymbol{p}} = \boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})}{}^{-1} \gamma_{(\boldsymbol{k},\boldsymbol{p})}, \tag{A.22}$$

where $\phi_{\boldsymbol{p}} = (\phi_1, \ldots, \phi_p)'$, $\gamma_{(\boldsymbol{k},\boldsymbol{p})} = (\gamma_{y^*}(k+1), \ldots, \gamma_{y^*}(k+p))'$ are vectors and $\boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})}$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}_{(\boldsymbol{k},\boldsymbol{p})} = \begin{bmatrix} \gamma_{y^*}(k) & \gamma_{y^*}(k-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(k+1) & \gamma_{y^*}(k) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(k+p-1) & \gamma_{y^*}(k+p-2) & \cdots & \gamma_{y^*}(k) \end{bmatrix}.$$

Note that (A.22) is also valid for any $j \geq 1$ such that

$$\phi_{\boldsymbol{p}} = \boldsymbol{\Gamma}_{(\boldsymbol{k}+\boldsymbol{j}-\mathbf{1},\boldsymbol{p})}{}^{-1} \gamma_{(\boldsymbol{k}+\boldsymbol{j}-\mathbf{1},\boldsymbol{p})}. \tag{A.23}$$

Now from (5.18) with $k = 0$ we have

$$\gamma_{y^*}(0) = \phi_1 \gamma_{y^*}(k-1) + \cdots + \phi_p \gamma_{y^*}(k-p) + \sigma_v^2 + \sigma_\varepsilon^2,$$

or,

$$\sigma_v = [\gamma_{y^*}(0) - \sum_{j=1}^{p} \phi_j \gamma_{y^*}(j) - \pi^2/2]^{1/2},$$

or, equivalently

$$\sigma_v = [\gamma_{y^*}(0) - \phi_{\boldsymbol{p}}' \gamma_{(\boldsymbol{k},\boldsymbol{p})} - \pi^2/2]^{1/2}, \tag{A.24}$$

where $\sum_{j=1}^{p}\phi_j\gamma_{y^*}(j) = \phi_{\boldsymbol{p}}^{'}\gamma_{(\boldsymbol{k},\boldsymbol{p})}$ and $\sigma_{\varepsilon}^2 = \psi^{(1)}(1/2) = \pi^2/2$. Now by construction,

$$\mu = \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)] = \log(\sigma_y^2) - 1.27, \tag{A.25}$$

or, equivalently

$$\sigma_y^2 = \exp(\mu + 1.27). \tag{A.26}$$

□

PROOF OF LEMMA 5.7 We are using an alternative method provided by Durbin (1960) that avoids the matrix inversion in the Yule-Walker equations. We derive the solution set of SV(2) parameters recursively from the solution set of SV(1) parameters, and the results of Lemma 5.7 can easily identify from there.
First, we solve for the moment equations of an SV(1) model as follows. We want to find closed-form moment equations solution for $\Theta_1^{SV} := (\phi_{11}, \sigma_{1v}, \sigma_y)$. Using Lemma 5.1, and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$, we get:

$$\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2 \exp\left[\frac{1}{2}\gamma_0\right] \tag{A.27}$$

$$\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4 \exp\left[2\gamma_0\right] \tag{A.28}$$

$$\mu_{2,2}(1|\theta) := \mathbb{E}(y_t^2 y_{t-1}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_1)\right] \tag{A.29}$$

where $\gamma_0 = \sigma_{1v}^2/(1-\phi_{11}\rho_1)$. Solving above equations yields:

$$\gamma_0 = log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}), \tag{A.30}$$

$$\sigma_y = \frac{(\mathbb{E}(y_t^2))^{1/2}}{(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2})^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{(\mathbb{E}(y_t^4))^{1/4}}, \tag{A.31}$$

$$\gamma_0\rho_1 = log(\frac{\mathbb{E}(y_t^2 u_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}), \tag{A.32}$$

hence

$$\rho_1 = log(\frac{\mathbb{E}(y_t^2 u_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}). \tag{A.33}$$

Since

$$\gamma_1 = \phi_{11}\gamma_0 \Leftrightarrow \phi_{11} = \gamma_1/\gamma_0 = \rho_1 \tag{A.34}$$

$$\Leftrightarrow \phi_{11} = log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}), \tag{A.35}$$

we get, using the Yule-Walker equation:

$$\gamma_0 = \phi_{11}\gamma_1 + \sigma_v^2 \Leftrightarrow 1 = \phi_{11}\rho_1 + \sigma_v^2/\gamma_0 = \phi_{11}^2 + \sigma_v^2\gamma_0, \tag{A.36}$$

or,

$$\sigma_v = [(1-\phi_{11}^2)\gamma_0]^{1/2} = [(1-(log(\frac{\mathbb{E}(y_t^2u_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2)(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))]^{1/2} \tag{A.37}$$

So the moment equations solution of SV(1) are:

$$\sigma_y = \frac{(\mathbb{E}(y_t^2))^{1/2}}{(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2})^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{(\mathbb{E}(y_t^4))^{1/4}}, \tag{A.38}$$

$$\phi_{11} = log(\frac{\mathbb{E}(y_t^2y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}), \tag{A.39}$$

$$\sigma_{1v} = [(1-(log(\frac{\mathbb{E}(y_t^2y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2)(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))]^{1/2}. \tag{A.40}$$

Let us now consider the SV(2) model. We wish to find closed-form estimator for $\theta := (\phi_{21}, \phi_{22}, \sigma_{2v}, \sigma_y)$. Using Lemma 5.1, and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$ and $k = l = 2$ & $m = 2$, we get:

$$\mu_2(\theta) := \mathbb{E}(y_t^2) = \sigma_y^2 \exp\left[\frac{1}{2}\gamma_0\right] \tag{A.41}$$

$$\mu_4(\theta) := \mathbb{E}(y_t^4) = 3\sigma_y^4 \exp\left[2\gamma_0\right] \tag{A.42}$$

$$\mu_{2,2}(1\,|\theta\,) := \mathbb{E}(y_t^2y_{t-1}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_1)\right] \tag{A.43}$$

$$\mu_{2,2}(2\,|\theta\,) := \mathbb{E}(y_t^2y_{t-2}^2) = \sigma_y^4 \exp\left[\gamma_0(1+\rho_2)\right] \tag{A.44}$$

where $\gamma_0 = \sigma_{2v}^2/(1-\sum_{j=1}^2 \phi_{2j}\rho_j)$. Solving the above equations yield:

$$\gamma_0 = log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) \tag{A.45}$$

$$\sigma_y = \frac{(\mathbb{E}(y_t^2))^{1/2}}{(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2})^{1/4}} = \frac{3^{1/4}\mathbb{E}(y_t^2)}{(\mathbb{E}(y_t^4))^{1/4}} \tag{A.46}$$

$$\rho_1 = log(\frac{\mathbb{E}(y_t^2y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) \tag{A.47}$$

$$\rho_2 = log(\frac{\mathbb{E}(y_t^2y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) \tag{A.48}$$

One interesting observation is

$$\rho_j = log(\frac{\mathbb{E}(y_t^2y_{t-j}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}), \;\; where \;\; j = 1,2 \tag{A.49}$$

Now using Durbin-Levinson recurrence formula,

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1-\phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\rho_2 - \phi_{11}^2}{1-\phi_{11}^2} \tag{A.50}$$

and

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \frac{\rho_1 - \rho_1\rho_2}{1-\rho_1^2} \tag{A.51}$$

Substitute $\phi_{11}, \rho_1, \rho_2$ to get:

$$\phi_{22} = \frac{log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) - (log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2}{1-(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2}, \tag{A.52}$$

or,

$$\phi_{22} = \frac{-(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))^2 + (log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2}))(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))}{(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2 - (log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))^2}, \tag{A.53}$$

and

$$\phi_{21} = \frac{log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) - (log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))(log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))}{1-(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})/log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2}, \tag{A.54}$$

or,

$$\phi_{21} = \frac{-(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))(log(\frac{3\mathbb{E}(y_t^2 y_{t-2}^2)}{\mathbb{E}(y_t^4)}))}{(log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}))^2 - (log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2}))^2}, \tag{A.55}$$

and

$$\sigma_{2v} = [(1-\phi_{21}\rho_1 - \phi_{22}\rho_2)\gamma_0]^{1/2} \tag{A.56}$$

$$\sigma_{2v} = [log(\frac{\mathbb{E}(y_t^4)}{3(\mathbb{E}(y_t^2))^2}) - \phi_1(log(\frac{\mathbb{E}(y_t^2 y_{t-1}^2)}{(\mathbb{E}(y_t^2))^2})) - \phi_2(log(\frac{\mathbb{E}(y_t^2 y_{t-2}^2)}{(\mathbb{E}(y_t^2))^2}))]^{1/2}. \tag{A.57}$$

We obtain the same results that requires matrix inversion using Durbin-Levinson algorithm. □

PROOF OF LEMMA 7.1 The method-of-moments estimator $\hat{\theta}_T$ is solution of the following optimization problem:

$$\min_{\theta} M_T = [\bar{g}_T(Y_T) - \mu(\theta)]' \hat{\Omega}_T [\bar{g}_T(Y_T) - \mu(\theta)].$$

Under assumptions 7.2 The score condition associated with this problem is:

$$J(\theta)\hat{\Omega}_T[\mu(\theta) - \bar{g}_T(Y_T)] = 0$$

A Taylor series expansion of the score condition around the true value of θ yields

$$0 = J(\theta)\hat{\Omega}_T[\mu(\theta) - \bar{g}_T(Y_T)] = J(\theta)\hat{\Omega}_T[\mu(\theta) - \bar{g}_T(Y_T)] + J(\theta)\hat{\Omega}_T J(\theta)'(\hat{\theta}_T - \theta) = O_p(T^{-1})$$

where, after rearranging the equation and using assumption 7.2,

$$\sqrt{T}[\hat{\theta}_T - \theta] = [J(\theta)\Omega J(\theta)^{'}]^{-1}J(\theta)\Omega\sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta)] + O_p(T^{-1/2})$$

Using assumptions 7.1, we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $V(\Omega)$ as specified in Lemma 7.1. □

PROOF OF LEMMA 7.2 To establish the asymptotic normality of $[\bar{g}_T(Y_T) - \mu(\theta)]$; we need to use a central limit theorem (CLT) for dependent processes (see Davidson (1994), Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this CLT holds. We workout this proof where $p = 2$. Setting

$$X_t := \begin{pmatrix} y_t^2 - \mu_2(\theta) \\ y_t^4 - \mu_4(\theta) \\ y_t^2 y_{t-1}^2 - \mu_{2,2}(1\,|\,\theta) \\ y_t^2 y_{t-2}^2 - \mu_{2,2}(2\,|\,\theta) \end{pmatrix} = g_t(\theta) - \mu(\theta)$$

$$S_T = \sum_{t=1}^{T} X_t = \sum_{t=1}^{T}[g_t(\theta) - \mu(\theta)]$$

and the subfields $F_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)^{'}$, we need to check the following conditions in order to get that $T^{-1/2}S_T = \sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{D} N(0, \Omega_*)$.

(i) $\{X_t, F_t\}$ is stationary and ergodic,

(ii) $\{X_t, F_t\}$ is a L_1-mixingale of size -1, and

(iii) $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\,|S_T| < \infty$.

(i) This follows from results 4.1 and 4.2.
(ii)-(1) A mixing zero-mean process is an adapted L_1- mixingale with respect to the sub-fields F_t provided it is bounded in the L_1-norm. To see that $\{X_t\}$ is bounded in the L_1-norm, using Theorem 14.2 of Davidson (1994), we have

$$\mathbb{E}\left|y_t^{2k} - \mu_{2k}(\theta)\right| \leq \mathbb{E}(\left|y_t^{2k}\right| + |\mu_{2k}(\theta)|) = 2\mu_{2k}(\theta) < \infty, \quad for \;\; k = 1, 2, \ldots,$$

$$\mathbb{E}\left|y_t^2 y_{t-k}^2 - \mu_{2,2}(k\,|\,\theta)\right| \leq \mathbb{E}(\left|y_t^2 y_{t-k}^2\right| + \left|\mu_{2,2}(k\,|\,\theta)\right|) = 2\mu_{2,2}(k\,|\,\theta) < \infty, \quad for \;\; k = 1, 2, \ldots \; .$$

(ii)-(2) We now need to show that the $\{X_t, F_t\}$ is a L_1-mixingale of size -1. From the discussion in section 4, we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi\rho^T$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\varphi})$, with $\varphi > 1$ (see Davidson (1994), Definition 16.1, p. 247). To see that

$$\lim_{T\to\infty}\frac{\rho^T}{T^{-\varphi}} = \lim_{T\to\infty} T^{\varphi}\exp(T\log\rho) = \lim_{T\to\infty}\exp(\varphi\log T)\exp(T\log\rho) = \lim_{T\to\infty}\exp(\varphi\log T + T\log\rho) = 0.$$

This holds in particular for $\varphi > 1$; see Rudin (1976) [Theorem 3.20(d), page 57].
(iii) We need to show that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\,|S_T| < \infty$ and using Cauchy-Schwarz inequality, we have

$\mathbb{E}\left|T^{-1/2}S_T\right| \leq T^{-1/2}\|S_T\|_2$. Now we can prove it by showing that

$$\limsup_{T\to\infty} T^{-1}\mathbb{E}(S_T S_T^{'}) < \infty \Leftrightarrow \limsup_{T\to\infty} \mathrm{Var}(T^{-1/2}S_T) < \infty$$

(iii)-(1) First and second components of S_T. Define $S_{T1} = \sum_{t=1}^{T} X_{1,t}$ where $X_{1,t} := y_t^2 - \mu_2(\theta)$ and compute:

$$\begin{aligned}\mathrm{Var}(T^{-1/2}S_{T1}) =& \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(X_{1,t}) + \sum_{t\neq s}\mathrm{cov}(X_{1,s}, X_{1,t})\right] = \frac{1}{T}\left[T\zeta_1(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_1(\tau)\right] \\ =& \zeta_1(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_1(\tau)\end{aligned} \tag{A.58}$$

Now we must prove that $\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_1(\tau)$ converges as $T\to\infty$. By Lemma 3.1.5 in Fuller (1976, p. 112), it is sufficient to show that $\sum_{\tau=1}^{\infty}\zeta_1(\tau)$ converge. Using Lemma 5.3, we have

$$\begin{aligned}\zeta_1(\tau) =& \mu_2^2(\theta)[\exp(\gamma_\tau) - 1] = \mu_2^2(\theta)\left[1 + \sum_{k=1}^{\infty}\frac{\gamma_\tau^k}{k!} - 1\right] = \mu_2^2(\theta)\left[\gamma_\tau\sum_{k=1}^{\infty}\frac{\gamma_\tau^{k-1}}{k!}\right] \\ =& \mu_2^2(\theta)\left[\gamma_\tau\sum_{k=0}^{\infty}\frac{\gamma_\tau^k}{(k+1)!}\right] \leq \mu_2^2(\theta)\left[\gamma_\tau\sum_{k=0}^{\infty}\frac{\gamma_\tau^k}{(k)!}\right] = \mu_2^2(\theta)\gamma_\tau\exp(\gamma_\tau)\end{aligned} \tag{A.59}$$

Therefore, the series

$$\sum_{\tau=1}^{\infty}\zeta_1(\tau) \leq \mu_2^2(\theta)\sum_{\tau=1}^{\infty}\gamma_\tau\exp(\gamma_\tau) \leq \mu_2^2(\theta)\exp(\gamma_1)\sum_{\tau=1}^{\infty}\gamma_\tau \leq \mu_2^2(\theta)\exp(\gamma_1)\underbrace{\sum_{\tau=1}^{\infty}|\gamma_\tau|}_{<\infty} < \infty \tag{A.60}$$

converges and by the Cauchy-Schwarz inequality we deduce that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T1}| < \infty$. The proof is very similar for S_{T2}.

(iii)-(2) Third and fourth components of S_T. we just have to show that $\sum_{\tau=1}^{\infty}\zeta_3(\tau) < \infty$. By Lemma 5.3, we have for all $\tau \geq 2$:

$$\begin{aligned}\zeta_3(\tau) &= \mu_{2,2}^2(1|\theta)[\exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}) - 1] = \mu_{2,2}^2(1|\theta)\left[1 + \sum_{k=1}^{\infty}\frac{(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})^k}{k!} - 1\right] \\ &= \mu_{2,2}^2(1|\theta)\left[(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})\sum_{k=1}^{\infty}\frac{(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})^{k-1}}{k!}\right] \\ &= \mu_{2,2}^2(1|\theta)\left[(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})\sum_{k=0}^{\infty}\frac{(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})^{k}}{(k+1)!}\right] \\ &\leq \mu_{2,2}^2(1|\theta)\left[(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})\sum_{k=0}^{\infty}\frac{(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})^{k}}{(k)!}\right] \\ &= \mu_{2,2}^2(1|\theta)(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})\exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1})\end{aligned} \tag{A.61}$$

Therefore, the series

$$
\begin{aligned}
\sum_{\tau=1}^{\infty}\zeta_3(\tau) &\le \zeta_3(1)+\mu_{2,2}^2(1|\theta)\sum_{\tau=2}^{\infty}(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1}) \\
&\le \zeta_3(1)+\mu_{2,2}^2(1|\theta)\exp(\gamma_1+2\gamma_2+\gamma_3)\sum_{\tau=2}^{\infty}(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1}) \\
&\le \zeta_3(1)+\mu_{2,2}^2(1|\theta)\exp(\gamma_1+2\gamma_2+\gamma_3)\left[\underbrace{\sum_{\tau=2}^{\infty}|\gamma_{\tau-1}|}_{<\infty}+2\underbrace{\sum_{\tau=2}^{\infty}|\gamma_\tau|}_{<\infty}+\underbrace{\sum_{\tau=2}^{\infty}|\gamma_{\tau+1}|}_{<\infty}\right]<\infty
\end{aligned}
\tag{A.62}
$$

converges and by the Cauchy-Schwarz inequality we deduce that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T3}|<\infty$. The proof for S_{T4} is similar. Thus we can apply Theorem 24.5 of Davidson (1994) to each component S_{Ti}, $i=1,2,3,4$ of S_T to state that: $T^{-1/2}S_{Ti}\xrightarrow{D}N(0,\lambda_i)$ and then by Cramér-Wold theorem we can establish the limiting result that stated in Lemma 7.2 . □

PROOF OF LEMMA 7.3 Under the assumptions 7.4 and 7.5 with $s=2$, the observed process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with $\mathbb{E}[y_t^*]<\infty$ and $\mathbb{E}[y_t^*y_{t+k}^*]<\infty$. So the consistency is a simple application of the Law of Large Numbers for stationary and ergodic processes, *i.e.*, the Ergodic theorem; see Theorem 13.12 and Corollary 13.14 of Davidson (1994). □

PROOF OF LEMMA 7.4 To establish the asymptotic normality of empirical moments; we shall use a CLT for dependent processes (see Davidson (1994), Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this CLT holds. Setting

$$
X_t := \begin{pmatrix} \log(y_t^2)-\mu \\ y_t^*y_{t+k}^*-\gamma_{y^*}(k)\end{pmatrix} = \begin{bmatrix}\Psi_t \\ \Lambda_{t,k}\end{bmatrix},
$$

$$
S_T=\sum_{t=1}^{T}X_t=\begin{bmatrix}\sum_{t=1}^{T}\Psi_t \\ \sum_{t=1}^{T}\Lambda_{t,k}\end{bmatrix},\ \ k=0,1,\ldots m,
$$

and the subfields $F_t=\sigma(s_t,s_{t-1},\ldots)$ where $s_t=(y_t,w_t)'$, we need to check the following conditions in order to get that

$$
T^{-1/2}S_T=\sqrt{T}\begin{bmatrix}\hat{\mu}-\mu \\ \hat{\Gamma}(m)-\Gamma(m)\end{bmatrix}\xrightarrow{d}N\left(0,\ \begin{bmatrix}V_\mu & C_{\mu,\Gamma(m)}' \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)}\end{bmatrix}\right),
$$

where $k=0,1,\ldots m$.

(i) $\{X_t,F_t\}$ is stationary and ergodic,

(ii) $\{X_t,F_t\}$ is a L_1-mixingale of size -1, and

(iii) $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_T|<\infty$.

(i) This follows from results 4.1 and 4.2.

(ii)-(1) A mixing zero-mean process is an adapted L_1- mixingale with respect to the sub-fields F_t provided it is bounded in the L_1-norm (see Davidson (1994), Theorem 14.2, p. 211). To see that $\{X_t\}$ is bounded in the L_1-norm, we note that:

$$\mathbb{E}|\log(y_t^2)-\mu| \leq \mathbb{E}(|\log(y_t^2)|+|\mu|) = 2\mu < \infty,$$

$$\mathbb{E}|y_t^* y_{t+k}^* - \gamma_{y^*}(k)| \leq \mathbb{E}(|y_t^* y_{t+k}^*| + |\gamma_{y^*}(k)|) = 2\gamma_{y^*}(k) < \infty, \quad for \;\; k = 0,1,\dots m.$$

(ii)-(2) We now need to show that the $\{X_t, F_t\}$ is a L_1-mixingale of size -1. From the discussion in section 4, we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi\rho^T$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\varphi})$, with $\varphi > 1$ (see Davidson (1994), Definition 16.1, p. 247). To see that

$$\lim_{T\to\infty}\frac{\rho^T}{T^{-\varphi}} = \lim_{T\to\infty} T^{\varphi}\exp(T\log\rho) = \lim_{T\to\infty}\exp(\varphi\log T)\exp(T\log\rho) = \lim_{T\to\infty}\exp(\varphi\log T + T\log\rho) = 0. \tag{A.63}$$

This holds in particular for $\varphi > 1$; see Rudin (1976) [Theorem 3.20(d), page 57].

(iii) We need to show that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_T| < \infty$ and using Cauchy-Schwarz inequality, we have $\mathbb{E}\left|T^{-1/2}S_T\right| \leq T^{-1/2}\left\|S_T\right\|_2$. Now we can prove it by showing that

$$\limsup_{T\to\infty} T^{-1}\mathbb{E}(S_T S_T^{'}) < \infty \Leftrightarrow \limsup_{T\to\infty}\mathrm{Var}(T^{-1/2}S_T) < \infty.$$

(iii)-(1) First component of S_T. Define $S_{T1} = \sum_{t=1}^T \Psi_t$ where $\psi_t := \log(y_t^2) - \mu$ and compute:

$$\begin{aligned}\zeta_\Psi(\tau) := \mathrm{cov}(\Psi_t, \Psi_{t+\tau}) =& \mathbb{E}[(\log(y_t^2)-\mu)(\log(y_{t+\tau}^2)-\mu)] \\ =& \mathbb{E}[y_t^* y_{t+\tau}^*] \\ =& \gamma_{y^*}(\tau),\end{aligned} \tag{A.64}$$

$$\begin{aligned}\mathrm{Var}(T^{-1/2}S_{T1}) =& \frac{1}{T}\left[\sum_{t=1}^T \mathrm{Var}(\Psi_t) + \sum_{t\neq s}\mathrm{cov}(\Psi_t,\Psi_s)\right] = \frac{1}{T}\left[T\zeta_\Psi(0) + 2\sum_{\tau=1}^T (T-\tau)\zeta_\Psi(\tau)\right] \\ =& \zeta_\Psi(0) + 2\sum_{\tau=1}^T(1-\frac{\tau}{T})\zeta_\Psi(\tau) = \gamma_{y^*}(0) + 2\sum_{\tau=1}^T(1-\frac{\tau}{T})\gamma_{y^*}(\tau).\end{aligned}$$

Now we have

$$\begin{aligned}\limsup_{T\to\infty}\mathrm{Var}(T^{-1/2}S_{T1}) =& \limsup_{T\to\infty}[\gamma_{y^*}(0) + 2\sum_{\tau=1}^T(1-\frac{\tau}{T})\gamma_{y^*}(\tau)] \\ =& \gamma_{y^*}(0) + 2\sum_{\tau=1}^\infty \gamma_{y^*}(\tau) \\ =& \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau) \leq \sum_{\tau=-\infty}^{\infty}|\gamma_{y^*}(\tau)| < \infty.\end{aligned} \tag{A.65}$$

This convergence is due to the fact that y_t^* follows a stationary $ARMA(p,p)$ process. So y_t^* can be viewed as an MA(∞) process with absolute summable coefficients and this implies absolute summability of auto-covariances (see Hamilton (1994), chapter 3, page 52). We deduce that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T1}| < \infty$ by the

Cauchy-Schwarz inequality.
(iii)-(2) Second component of $S_{T2,k}$. Define $S_{T2,k} = \sum_{t=1}^{T} \Lambda_{t,k}$ where $\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k)$, where $k = 0, 1, \ldots m$ and compute:

$$\begin{aligned}
\zeta_{\Lambda_k}(\tau) := & \ \mathrm{cov}(\Lambda_{t,k}, \Lambda_{t+\tau,k}) \\
&= \mathbb{E}[(y_t^* y_{t+k}^* - \gamma_{y^*}(k))(y_{t+\tau}^* y_{t+\tau+k}^* - \gamma_{y^*}(k))] \\
&= \mathbb{E}[y_t^* y_{t+k}^* y_{t+\tau}^* y_{t+\tau+k}^*] - \gamma_{y^*}^2(k) \\
&= \mathbb{E}[y_t^* y_{t+k}^*]\mathbb{E}[y_{t+\tau}^* y_{t+\tau+k}^*] + \mathrm{cov}(y_t^*, y_{t+\tau}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau+k}^*) \\
&+ \mathrm{cov}(y_t^*, y_{t+\tau+k}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau}^*) - \gamma_{y^*}^2(k) \\
&= \gamma_{y^*}^2(k) + \gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) - \gamma_{y^*}^2(k) \\
&= \gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) \quad \forall \tau \geq k,
\end{aligned} \tag{A.66}$$

$$\begin{aligned}
\mathrm{Var}(T^{-1/2} S_{T2,k}) &= \frac{1}{T}\left[\sum_{t=1}^{T} \mathrm{Var}(\Lambda_{t,k}) + \sum_{t \neq s} \mathrm{cov}(\Lambda_{t,k}, \Lambda_{s,k})\right] = \frac{1}{T}\left[T\zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\Lambda_k}(\tau)\right] \\
&= \zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\Lambda_k}(\tau) \\
&= \gamma_{y^*}^2(0) + \gamma_{y^*}(k)\gamma_{y^*}(-k) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)].
\end{aligned} \tag{A.67}$$

Now we have

$$\begin{aligned}
\limsup_{T\to\infty} \mathrm{Var}(T^{-1/2} S_{T2,k}) &= \gamma_{y^*}^2(0) + \gamma_{y^*}(k)\gamma_{y^*}(-k) \\
&+ \limsup_{T\to\infty}[2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)]] \\
&= \sum_{\tau=-\infty}^{\infty}[\gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)] \\
&= \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau) + \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) \\
&= \underbrace{\sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau)}_{<\infty} + \underbrace{\sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau+k)}_{<\infty} < \infty.
\end{aligned} \tag{A.68}$$

The sums are converged because absolute summability implies square-summability. We deduce that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T2,k}| < \infty$ by the Cauchy-Schwarz inequality. Thus we can apply Theorem 24.5 of Davidson (1994) to each component of S_T to state that:

$$T^{-1/2} S_{Ti} \xrightarrow{d} N(0, \lambda_i)$$

and then by the Cramér-Wold theorem we can establish the following limiting result for the $(m+2) \times 1$-

vector S_T:

$$T^{-1/2}S_T = T^{-1/2}\sum_{t=1}^{T} X_t = \sqrt{T}\begin{bmatrix}\hat{\mu}-\mu\\ \hat{\Gamma}(m)-\Gamma(m)\end{bmatrix} \xrightarrow{d} N\left(0,V\right),$$

where

$$V = \lim_{T\to\infty}\mathbb{E}[(T^{-1/2}S_T)^2] = \lim_{T\to\infty}\mathbb{E}\{T[M_T][M_T]^{'}\},$$

where $M_T = \left(\hat{\mu}_T-\mu,\ \hat{\Gamma}_T(m)-\Gamma(m)\right)^{'}$ is $(m+2)\times 1$-vector. Furthermore,

$$V = \begin{bmatrix} V_\mu & C^{'}_{\mu,\Gamma(m)}\\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)}\end{bmatrix},$$

since from (A.65) and (A.67) we have

$$V_\mu = \gamma_{y^*}(0)+2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau),$$

and $V_{\Gamma(m)}$ is a $(m+1)\times(m+1)$ matrix given by

$$V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t)+2\sum_{\tau=1}^{\infty}\mathrm{cov}(\Lambda_t,\Lambda_{t+\tau}),$$

where Λ_t is an $(m+1)\times 1$ vector with $\Lambda_{t,k} = y^*_t y^*_{t+k} = (\log(y_t^2)-\mu)(\log(y_{t+k}^2)-\mu)$, $k=0,\dots,m$ and

$$\begin{aligned} C_{\mu,\Gamma(m)} &= \sum_t \mathrm{cov}(\Psi_t,\Lambda_{t,k}) = 2\sum_{t=1}^{\infty}\mathbb{E}[(log(y_t^2)-\mu)(y^*_t y^*_{t+k}-\gamma_{y^*})] = 2\sum_{t=1}^{\infty}\mathbb{E}[y^*_t(y^*_t y^*_{t+k}-\gamma_{y^*})]\\ &= 2\sum_{t=1}^{\infty}(\mathbb{E}[y_t^{*2}y^*_{t+k}]-\underbrace{\mathbb{E}[y^*_t]}_{=0}\gamma_{y^*}) = 2\sum_{t=1}^{\infty}\mathbb{E}[y_t^{*2}y^*_{t+k}];\quad k=0,1,2,\dots,m. \end{aligned} \tag{A.69}$$

Now for $k=0$, we substitute $y^*_t = w_t+\varepsilon_t$, to get

$$\bar{c} := C_{\mu,\Gamma(0)} = 2\sum_{t=1}^{\infty}\mathbb{E}[y_t^{*3}] = 2\sum_{t=1}^{\infty}(\mathbb{E}[w_t^3]+\mathbb{E}[\varepsilon_t^3]) = 2\sum_{t=1}^{\infty}\mathbb{E}[\varepsilon_t^3]. \tag{A.70}$$

Note that if we assume z_t is an i.i.d. normal random variable then $\mathbb{E}(\varepsilon_t^3) = \psi^{(2)}(\frac{1}{2})$ from (2.6) and it is equal to $-14\mathscr{Z}(3)$ where $\mathscr{Z}(\cdot)$ is Riemann's Zeta function with $\mathscr{Z}(3) = 1.20205$.[6] For $k=1,\dots,m$, it is easily seen that $C_{\mu,\Gamma(m)} = 0$ from the MA(∞) representation of w_t. So $C_{\mu,\Gamma(m)}$ is a $(m+1)\times 1$ vector given by $(\bar{c},\ 0_{[1\times m]})'$, with $\bar{c}$ is defined in (A.70). □

PROOF OF THEOREM 7.5 It is easily seen that D is a continuously differentiable mapping of $(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\dots,\gamma_{y^*}(2p))$. The convergence result stated in (7.12) follows from the standard result for differentiable transformations of asymptotically normally distributed variables together with the application of the multivariate delta method.

[6] The Riemann Zeta function for $s\in\mathbb{C}$ with $\mathrm{Re}(s)>1$ is defined as $\mathscr{Z}(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$.

In case of an SV(1) model, the function $D_1 = (D_{\phi_1}, D_{\sigma_y}, D_{\sigma_v})'$ is given by

$$D_{\phi_1} = \gamma_{y^*}(2)/\gamma_{y^*}(1), \quad D_{\sigma_y} = \exp(\mu + 1.27)^{1/2}, \quad D_{\sigma_v} = \kappa_1 \kappa_2, \tag{A.71}$$

with $\hat{\theta} = D_1(\hat{\mu}, \hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \hat{\gamma}_{y^*}(2))$ and the analytical moment derivatives of D_1, i.e.,

$$\frac{\partial D_1(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2))}{\partial(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2))} = \begin{pmatrix} 0 & 0 & -\gamma_{y^*}(2)/\gamma_{y^*}(1)^2 & 1/\gamma_{y^*}(1) \\ \sigma_y/2 & 0 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{\frac{\kappa_1}{\kappa_2}} & \frac{\gamma_{y^*}(2)^2}{\gamma_{y^*}(1)^3}\sqrt{\frac{\kappa_2}{\kappa_1}} & -\frac{\gamma_{y^*}(2)}{\gamma_{y^*}(1)^2}\sqrt{\frac{\kappa_2}{\kappa_1}} \end{pmatrix}$$

where $\sigma_y = \sqrt{\exp(\mu + 1.27)}$, $\kappa_1 = [1 - (\gamma_{y^*}(2)/\gamma_{y^*}(1))^2]$, $\kappa_2 = [\gamma_{y^*}(0) - \pi^2/2]$.
In case of an SV(2) model,, the function $D_2 = (D_{\phi_1}, D_{\phi_2}, D_{\sigma_y}, D_{\sigma_v})'$ is given by

$$\begin{aligned} D_{\phi_1} &= \frac{\gamma_{y^*}(2)\gamma_{y^*}(3) - \gamma_{y^*}(1)\gamma_{y^*}(4)}{\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}, \qquad & D_{\phi_2} &= \frac{\gamma_{y^*}(2)\gamma_{y^*}(4) - \gamma_{y^*}^2(3)}{\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}, \\ D_{\sigma_y} &= \exp(\mu + 1.27)^{1/2}, & D_{\sigma_v} &= [\gamma_{y^*}(0) - \pi^2/2 - \phi_1\gamma_{y^*}(1) - \phi_2\gamma_{y^*}(2)]^{1/2}, \end{aligned} \tag{A.72}$$

with $\hat{\theta} = D_2(\hat{\mu}, \hat{\gamma}_{y^*}(0), \ldots, \hat{\gamma}_{y^*}(4))$ and the analytical moment derivatives of D_2, i.e.,

$$D_2' = \begin{pmatrix} 0 & 0 & \kappa_3(\gamma_{y^*}(4) - \phi_1\gamma_{y^*}(3)) & \kappa_3(2\phi_1\gamma_{y^*}(2) - \gamma_{y^*}(3)) & -\kappa_3(\phi_1\gamma_{y^*}(1) + \gamma_{y^*}(2)) & \kappa_3\gamma_{y^*}(1) \\ 0 & 0 & \kappa_3\phi_2\gamma_{y^*}(3) & -\kappa_3(2\phi_2\gamma_{y^*}(2) + \gamma_{y^*}(4)) & -\kappa_3(\phi_2\gamma_{y^*}(1) - 2\gamma_{y^*}(3)) & -\kappa_3\gamma_{y^*}(2) \\ \frac{\sigma_y}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2\sigma_v} & \frac{\kappa_3\kappa_4 - \phi_1}{2\sigma_v} & \frac{\kappa_3\kappa_5 - \phi_2}{2\sigma_v} & \frac{\kappa_3\kappa_6}{2\sigma_v} & \frac{\kappa_3(\gamma_{y^*}^2(2) - \gamma_{y^*}^2(1))}{2\sigma_v} \end{pmatrix}$$

where $\sigma_y = \sqrt{\exp(\mu + 1.27)}$, $\phi_1 = \frac{\gamma_{y^*}(2)\gamma_{y^*}(3) - \gamma_{y^*}(1)\gamma_{y^*}(4)}{\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}$, $\phi_2 = \frac{\gamma_{y^*}(2)\gamma_{y^*}(4) - \gamma_{y^*}^2(3)}{\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}$, $\sigma_v = [\gamma_{y^*}(0) - \pi^2/2 - \phi_1\gamma_{y^*}(1) - \phi_2\gamma_{y^*}(2)]^{1/2}$, $\kappa_3 = [\gamma_{y^*}^2(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)]^{-1}$, $\kappa_4 = [\phi_1\gamma_{y^*}(1)\gamma_{y^*}(3) + \phi_2\gamma_{y^*}(2)\gamma_{y^*}(3) - \gamma_{y^*}(1)\gamma_{y^*}(4)]$, $\kappa_5 = [\gamma_{y^*}(1)\gamma_{y^*}(2) + \gamma_{y^*}(2)\gamma_{y^*}(4) - 2\phi_1\gamma_{y^*}(1)\gamma_{y^*}(2) - 2\phi_2\gamma_{y^*}^2(2)]$, $\kappa_6 = [\phi_1\gamma_{y^*}^2(1) + (1 + \phi_2)\gamma_{y^*}(1)\gamma_{y^*}(2) - 2\gamma_{y^*}(2)\gamma_{y^*}(3)]$. □

A.2. Tables

Table 1: Comparison of different winsorized estimators using simulated data.

	$T = 100$							$T = 1000$							$T = 10000$						
J	$\hat{\phi}_1$	SE	$\hat{\phi}_2$	SE	$\hat{\sigma}_v$	SE	NIV	$\hat{\phi}_1$	SE	$\hat{\phi}_2$	SE	$\hat{\sigma}_v$	SE	NIV	$\hat{\phi}_1$	SE	$\hat{\phi}_2$	SE	$\hat{\sigma}_v$	SE	NIV
Mean																					
1	0.17	28.39	0.47	29.15	2.79	1.54	51	0.48	0.87	0.41	0.84	2.50	0.22	0	0.45	0.09	0.44	0.08	2.50	0.04	0
5	0.05	11.39	0.57	11.30	2.89	1.37	73	-0.25	7.10	1.10	6.59	2.60	0.74	28	-0.18	13.11	1.03	12.17	2.56	0.78	12
10	-0.03	7.05	0.26	7.56	3.08	1.98	112	-1.55	27.00	2.30	25.29	2.69	1.56	39	-3.95	115.97	4.55	108.09	2.67	2.26	27
20	-0.45	18.70	0.13	7.52	3.31	3.14	175	-0.95	14.10	1.72	13.20	2.72	1.23	51	-3.33	65.95	3.99	61.71	2.72	1.92	36
30	-0.25	12.90	0.01	6.74	3.34	2.93	198	-0.65	10.13	1.35	9.56	2.80	1.44	90	-3.21	49.15	3.84	45.76	2.77	1.90	39
40	-0.13	9.90	-0.03	5.53	3.33	2.71	210	-0.51	7.87	1.22	7.43	2.80	1.38	132	-2.50	36.99	3.00	34.73	2.86	2.42	40
50	-0.12	8.37	0.02	4.70	3.29	2.74	210	-0.44	6.54	1.11	6.11	2.85	1.57	151	-1.86	30.10	2.36	28.35	2.91	2.36	62
100								-0.08	4.60	0.61	4.70	2.96	1.95	202	-0.97	16.39	1.52	15.43	2.91	2.11	136
LD																					
1	0.17	28.39	0.47	29.15	2.79	1.54	51	0.48	0.87	0.41	0.84	2.50	0.22	0	0.45	0.09	0.44	0.08	2.50	0.04	0
5	0.04	11.80	0.61	11.90	2.86	1.30	73	-0.09	8.10	0.94	7.46	2.57	0.71	21	0.11	8.70	0.77	8.08	2.52	0.61	5
10	0.02	8.57	0.43	8.66	2.96	1.63	90	-1.22	22.57	2.00	21.23	2.66	1.40	36	-3.46	105.15	4.10	98.01	2.64	2.15	17
20	-0.46	16.20	0.36	7.41	3.19	2.84	142	-1.12	19.02	1.90	17.85	2.68	1.33	42	-3.37	84.40	4.01	78.71	2.68	2.02	33
30	-0.42	15.42	0.25	6.53	3.27	2.89	172	-0.87	14.78	1.64	13.87	2.70	1.27	54	-3.27	66.01	3.92	61.61	2.73	1.93	34
40	-0.26	13.50	0.10	6.00	3.28	2.82	188	-0.75	12.18	1.50	11.43	2.73	1.31	88	-3.03	54.42	3.64	50.79	2.79	1.99	35
50	-0.23	11.81	0.09	5.51	3.28	2.76	196	-0.65	10.34	1.38	9.72	2.77	1.35	112	-2.68	46.32	3.26	43.26	2.83	2.11	38
100								-0.19	6.36	0.82	6.08	2.86	1.71	158	-1.69	27.23	2.25	25.52	2.87	2.14	95
Median																					
1	0.17	28.39	0.47	29.15	2.79	1.54	51	0.48	0.87	0.41	0.84	2.50	0.22	0	0.45	0.09	0.44	0.08	2.50	0.04	0
5	0.27	0.55	0.48	0.52	2.65	0.57	5	0.39	0.34	0.50	0.32	2.52	0.18	0	0.45	0.18	0.45	0.16	2.50	0.06	0
10	0.26	0.39	0.45	0.38	2.73	0.59	5	0.33	0.26	0.55	0.24	2.54	0.16	0	0.38	0.18	0.51	0.17	2.52	0.07	0
20	0.31	0.30	0.44	0.32	2.65	0.63	5	0.28	0.21	0.59	0.19	2.58	0.23	0	0.32	0.17	0.57	0.15	2.54	0.08	0
30	0.32	0.27	0.47	0.29	2.55	0.60	11	0.27	0.18	0.60	0.17	2.59	0.32	0	0.29	0.15	0.60	0.13	2.56	0.11	0
40	0.32	0.24	0.48	0.28	2.52	0.60	14	0.28	0.15	0.60	0.15	2.55	0.35	1	0.27	0.14	0.61	0.13	2.57	0.17	0
50	0.32	0.23	0.49	0.27	2.47	0.60	9	0.28	0.15	0.61	0.14	2.50	0.38	0	0.27	0.13	0.61	0.12	2.56	0.21	0
100								0.29	0.11	0.62	0.12	2.42	0.35	1	0.28	0.10	0.63	0.11	2.45	0.27	0
OLS																					
1	0.17	28.39	0.47	29.15	2.79	1.54	51	0.48	0.87	0.41	0.84	2.50	0.22	0	0.45	0.09	0.44	0.08	2.50	0.04	0
5	0.33	0.35	0.43	0.33	2.65	0.46	0	0.44	0.22	0.45	0.21	2.52	0.14	0	0.45	0.09	0.45	0.08	2.50	0.04	0
10	0.30	0.28	0.44	0.28	2.69	0.44	0	0.41	0.19	0.48	0.18	2.53	0.13	0	0.45	0.09	0.45	0.08	2.50	0.04	0
20	0.30	0.23	0.46	0.24	2.67	0.44	0	0.39	0.17	0.50	0.16	2.53	0.13	0	0.44	0.08	0.46	0.08	2.50	0.04	0
30	0.30	0.22	0.47	0.22	2.63	0.44	0	0.38	0.16	0.50	0.16	2.53	0.13	0	0.44	0.08	0.46	0.08	2.50	0.04	0
40	0.30	0.20	0.49	0.21	2.60	0.44	0	0.38	0.16	0.51	0.15	2.53	0.13	0	0.44	0.08	0.46	0.08	2.50	0.04	0
50	0.30	0.19	0.49	0.21	2.58	0.44	0	0.38	0.16	0.51	0.15	2.53	0.13	0	0.44	0.08	0.46	0.08	2.50	0.04	0
100								0.37	0.15	0.52	0.14	2.51	0.13	0	0.44	0.08	0.46	0.08	2.50	0.04	0

Notes:
1. Simulation results: estimation of $(\phi_1, \phi_2, \sigma_v)'$ using simple mean, weighted mean (LD), median, and regression without intercept (OLS).
2. Simulated process: SV(2) with $(\phi_1, \phi_2, \sigma_y, \sigma_v) = (0.45, 0.45, 0.25, 2.5)$, and the number of replications is 1000.
3. NIV stands for the number of inadmissible parameter values produced by the individual estimators.

Table 2: Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data.

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.30	0.60	0.025	2.5	0.30	0.60	0.025	2.5
	Bias							
GMM-6M-E	-0.142	-0.563	0.054	3.287	-0.176	-0.621	0.052	2.794
GMM-6M-NW	-0.451	-0.464	1.716	0.401	-0.602	-0.508	2.710	0.141
GMM-24M-E	0.054	-0.343	0.055	3.947	0.057	-0.422	0.059	4.261
GMM-24M-NW	-0.503	-0.556	2.083	0.980	-0.587	-0.597	2.995	0.752
Bayesian-MCMC	0.917	-0.853	0.412	-2.242	0.839	-0.753	0.383	-2.277
EDV*	-0.284	0.021	0.719	-1.098	-0.222	0.051	0.692	-0.983
ARMA	0.008	-0.022	0.003	0.009	0.004	-0.007	0.001	0.001
W-ARMA	-0.009	-0.007	0.003	0.011	0.002	-0.006	0.001	0.001
	RMSE							
GMM-6M-E	0.712	0.668	0.117	4.632	0.687	0.747	0.094	4.349
GMM-6M-NW	0.661	0.552	2.667	2.059	0.738	0.577	3.588	1.676
GMM-24M-E	0.384	0.441	0.095	4.779	0.495	0.540	0.078	5.163
GMM-24M-NW	0.802	0.668	3.235	2.947	0.839	0.693	4.005	2.692
Bayesian-MCMC	0.932	0.872	0.737	2.242	0.840	0.754	0.499	2.277
EDV*	1.361	1.376	1.537	1.254	1.832	1.835	1.021	1.208
ARMA	0.198	0.193	0.016	0.184	0.084	0.081	0.007	0.091
W-ARMA	0.139	0.137	0.016	0.178	0.080	0.077	0.007	0.089

Notes:
1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 5.1.
6. ARMA is the simple ARMA-based estimator proposed in Section 5.2.
7. W-ARMA is the winsorized ARMA estimator based on OLS proposed in Section 5.3.
8. *EDV produces 159 and 128 inadmissible values out of 1000 simulations when $T = 500$ and $T = 2000$, respectively.

Table 3: Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data.

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.45	0.45	0.25	2.5	0.45	0.45	0.25	2.5
	Bias							
GMM-6M-E	-0.177	-0.639	0.796	3.098	-0.154	-0.623	0.864	2.462
GMM-6M-NW	-0.560	-0.599	4.143	0.768	-0.561	-0.671	3.814	0.740
GMM-24M-E	0.094	-0.408	0.855	3.471	0.041	-0.435	1.010	4.382
GMM-24M-NW	-0.585	-0.700	5.667	1.392	-0.548	-0.742	5.428	1.363
Bayesian-MCMC	0.760	-0.693	4.208	-2.246	0.691	-0.603	4.002	-2.277
EDV*	-0.679	0.415	9.065	-1.022	-1.203	1.042	8.685	-0.894
ARMA**	-0.017	0.002	0.035	0.015	0.022	-0.025	0.006	-0.004
W-ARMA	-0.079	0.057	0.035	0.042	-0.020	0.014	0.006	0.007
	RMSE							
GMM-6M-E	0.741	0.744	1.019	4.629	0.635	0.709	0.945	3.910
GMM-6M-NW	0.766	0.659	5.308	2.082	0.766	0.723	5.154	1.835
GMM-24M-E	0.453	0.513	1.175	4.482	0.529	0.558	1.310	5.274
GMM-24M-NW	0.884	0.808	6.700	3.359	0.853	0.838	6.717	3.158
Bayesian-MCMC	0.777	0.715	7.481	2.246	0.693	0.605	5.098	2.277
EDV*	3.723	3.709	20.36	1.340	20.45	20.44	13.27	1.827
ARMA**	1.032	0.985	0.162	0.327	0.214	0.204	0.074	0.102
W-ARMA	0.220	0.207	0.162	0.192	0.162	0.153	0.074	0.095

Notes:
1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 5.1.
6. ARMA is the simple ARMA-based estimator proposed in Section 5.2.
7. W-ARMA is the winsorized ARMA estimator based on OLS proposed in Section 5.3.
8. *EDV produces 160 and 159 inadmissible values out of 1000 simulations when $T = 500$ and $T = 2000$, respectively.
9. **ARMA produces 5 and 1 inadmissible values out of 1000 simulations, of ϕ when $T = 500$ and $T = 2000$, respectively.

Table 4: Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data.

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.90	-0.90	0.5	2.5	0.90	-0.90	0.5	2.5
	Bias							
GMM-6M-E	-0.734	-0.736	6.684	-0.412	-0.565	-1.296	7.682	-1.790
GMM-6M-NW	-0.639	-0.883	4.161	-0.807	-0.410	-0.551	2.608	-0.642
GMM-24M-E	-1.192	-0.600	0.885	4.700	-1.232	-0.627	0.838	5.182
GMM-24M-NW	-0.677	-0.965	5.045	-0.996	-0.443	-0.597	3.072	-0.759
Bayesian-MCMC	0.296	0.675	61.90	-2.248	0.234	0.758	76.06	-2.277
EDV*	137.86	172.04	1383.1	27.51	6.371	76.703	5368.3	31.58
ARMA	-0.001	0.002	0.001	-0.004	-0.001	0.000	-0.001	-0.003
W-ARMA	-0.002	0.002	0.001	-0.003	-0.001	0.001	-0.001	-0.002
	RMSE							
GMM-6M-E	0.940	1.071	7.175	2.646	0.823	1.363	7.898	2.108
GMM-6M-NW	0.829	1.015	5.450	1.349	0.613	0.774	4.279	1.139
GMM-24M-E	1.212	0.610	0.922	5.336	1.246	0.637	0.868	5.741
GMM-24M-NW	0.892	1.123	6.596	1.653	0.668	0.852	5.049	1.305
Bayesian-MCMC	0.334	0.698	83.650	2.249	0.239	0.760	109.6	2.277
EDV*	1166.5	1166.5	4926.9	55.3	145.3	144.6	19562	40.80
ARMA	0.026	0.031	0.037	0.185	0.012	0.014	0.019	0.093
W-ARMA	0.027	0.025	0.037	0.193	0.013	0.012	0.019	0.095

Notes:
1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 5.1.
6. ARMA is the simple ARMA-based estimator proposed in Section 5.2.
7. W-ARMA is the winsorized ARMA estimator based on OLS proposed in Section 5.3.
8. *EDV produces 922 and 971 inadmissible values out of 1000 simulations when $T = 500$ and $T = 2000$, respectively.

Table 5: Comparison of different estimation methods with respect to bias and RMSE for an SV(2) model using simulated data.

	$T = 500$				$T = 2000$			
	ϕ_1	ϕ_2	σ_y	σ_v	ϕ_1	ϕ_2	σ_y	σ_v
True value	0.00	0.90	0.025	2.5	0.00	0.90	0.025	2.5
	Bias							
GMM-6M-E	-0.613	-0.608	0.147	3.084	-0.559	-0.642	0.160	2.599
GMM-6M-NW	-0.725	-0.500	3.659	0.521	-0.743	-0.545	4.055	1.004
GMM-24M-E	-0.172	-0.378	0.129	4.772	-0.396	-0.378	0.169	5.168
GMM-24M-NW	-0.827	-0.575	4.367	1.487	-0.837	-0.592	5.285	1.763
Bayesian-MCMC	1.198	-1.131	1.528	-2.245	1.132	-1.043	1.493	-2.278
EDV*	0.508	0.415	4.007	-0.851	0.452	0.394	4.104	-0.793
ARMA	-0.003	-0.011	0.003	0.007	-0.001	-0.003	0.001	0.000
W-ARMA	-0.004	-0.014	0.003	0.019	-0.001	-0.003	0.001	0.002
	RMSE							
GMM-6M-E	0.851	0.731	0.352	4.549	0.786	0.763	0.203	4.118
GMM-6M-NW	0.825	0.554	4.778	1.929	0.835	0.588	5.182	2.045
GMM-24M-E	0.633	0.486	0.193	5.740	0.667	0.478	0.211	6.049
GMM-24M-NW	0.950	0.668	5.504	3.301	0.946	0.663	6.358	3.205
Bayesian-MCMC	1.208	1.144	6.164	2.246	1.133	1.044	2.033	2.278
EDV*	2.253	2.227	18.603	1.325	2.120	2.101	10.789	1.313
ARMA	0.031	0.033	0.016	0.188	0.014	0.014	0.007	0.093
W-ARMA	0.029	0.031	0.016	0.183	0.013	0.013	0.007	0.090

Notes:
1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 5.1.
6. ARMA is the simple ARMA-based estimator proposed in Section 5.2.
7. W-ARMA is the winsorized ARMA estimator based on OLS proposed in Section 5.3.
8. *EDV produces 241 and 209 inadmissible values out of 1000 simulations when $T = 500$ and $T = 2000$, respectively.

Table 6: Comparison of different estimation methods with respect to elapsed time (in seconds) for an SV(2) model using simulated data.

Elapsed time (in seconds)		
	$T = 500$	$T = 2000$
GMM-6M-E	469.03	1009.40
GMM-6M-NW	650.74	2064.03
GMM-24M-E	1118.43	2511.47
GMM-24M-NW	1973.47	5709.50
Bayesian-MCMC	35585.58	178738.09
EDV	0.63	1.40
ARMA	0.64	1.41
W-ARMA	1.08	1.91

Notes:
1. GMM-6M and GMM-24M are the generalized method of moment estimators with six moments and 24 moments, respectively.
2. E stands for the efficient GMM estimation where we used the inverse of the covariance matrix as the weighting matrix.
3. NW stands for the GMM estimation where we used the inverse of Newey West covariance matrix as the weighting matrix.
4. Bayesian-MCMC is the Bayesian estimator based on Markov Chain Monte Carlo methods.
5. EDV is the extension of Dufour and Valéry (2006) method proposed in Section 5.1.
6. ARMA is the simple ARMA-based estimator proposed in Section 5.2.
7. W-ARMA is the winsorized ARMA estimator based on OLS proposed in Section 5.3.

Table 7: Summary statistics.

S&P 500 index, 1928 - 2016, number of observations: 23372								
Series	Mean	SD	Kurtosis	Skewness	Range	Max	Min	LB(10)
y_t	0.00	0.50	21.98	-0.43	16.62	6.66	-9.95	104.4
y_t^2	0.25	1.14	2647.09	37.17	99.08	99.08	0.00	7338.5
$\log(\|y_t\|)$	-1.73	1.25	5.07	-0.96	13.60	2.30	-11.30	5180.9
$y_t^* = \log(y_t^2) - \mu$	0.00	2.49	5.07	-0.96	27.20	8.05	-19.15	5180.9

Notes:
1. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the squared of residual return and y_t^* is the residual of log squared of residual return.
2. LB(10) is the heteroskedasticity-corrected Ljung - Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 8: ARMA-based empirical estimates of SV(p) models.

S&P 500 index, 1928 - 2016, number of observations: 23372								
	$p=1$				$p=2$			
	Coefficient	Std. Error	t-stat	p-value	Coefficient	Std. Error	t-stat	p-value
ϕ_1	0.9938	(0.0357)	27.84	0.000	0.6887	(0.0719)	9.58	0.000
ϕ_2					0.2863	(0.0734)	3.90	0.000
σ_y	0.3356	(0.0167)	20.06	0.000	0.3356	(0.0167)	20.06	0.000
σ_v	0.6533	(0.0623)	10.48	0.000	0.6166	(0.3204)	1.92	0.027
Time (in seconds)	0.69				0.70			
	$p=3$				$p=4$			
	Coefficient	Std. Error	t-stat	p-value	Coefficient	Std. Error	t-stat	p-value
ϕ_1	0.5477	(0.1204)	4.55	0.000	0.3633	(0.2153)	1.69	0.05
ϕ_2	-0.4264	(0.0936)	-4.55	0.000	-0.0251	(0.2117)	-0.12	0.45
ϕ_3	0.8489	(0.0122)	69.67	0.000	0.6305	(0.0167)	37.68	0.00
ϕ_4					0.0005	(0.0162)	0.03	0.49
σ_y	0.3356	(0.0167)	20.06	0.000	0.3356	(0.0167)	20.06	0.00
σ_v	0.6211	(0.3993)	1.56	0.060	0.6133	(0.9210)	0.67	0.25
Time (in seconds)	0.79				0.97			
S&P 500 index, 1996 - 2016, number of observations: 5222								
	$p=1$				$p=2$			
	Coefficient	Std. Error	t-stat	p-value	Coefficient	Std. Error	t-stat	p-value
ϕ_1	0.9894	(0.0859)	11.52	0.00	0.5722	(0.1159)	4.94	0.00
ϕ_2					0.3836	(0.1091)	3.52	0.00
σ_y	0.3894	(0.0357)	10.91	0.00	0.3894	(0.0357)	10.91	0.00
σ_v	0.7543	(0.1061)	7.11	0.00	0.6760	(0.4557)	1.48	0.07
Time (in seconds)	0.66				0.68			
	$p=3$				$p=4$			
	Coefficient	Std. Error	t-stat	p-value	Coefficient	Std. Error	t-stat	p-value
ϕ_1	0.1033	(0.1585)	0.65	0.26	0.2751	(0.2320)	1.19	0.12
ϕ_2	0.5296	(0.1271)	4.17	0.00	0.5423	(0.2272)	2.39	0.01
ϕ_3	0.3560	(0.0218)	16.30	0.00	0.1917	(0.0262)	7.32	0.00
ϕ_4					0.0178	(0.0279)	0.64	0.26
σ_y	0.3894	(0.0357)	10.91	0.00	0.3894	(0.0357)	10.91	0.00
σ_v	0.5251	(0.6699)	0.78	0.22	0.5436	(1.1557)	0.47	0.32
Time (in seconds)	0.72				0.75			

Standard errors are in parenthesis.

Table 9: Finite sample inference of SV(p) models using ARMA-based estimates.

S&P 500 index, 1928 - 2016, number of observations: 23372						
	$p=1$					
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.9938	27.84	0.00	0.05	0.01	0.001
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6533	10.48	0.00	0.05	0.01	0.001
Time (in seconds)			0.69	1.5	4.5	38.2
	$p=2$					
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.6887	9.58	0.00	0.05	0.01	0.001
ϕ_2	0.2863	3.90	0.00	0.05	0.01	0.001
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6166	1.92	0.03	0.10	0.06	0.075
Time (in seconds)			0.70	2.6	10.3	95.1
	$p=3$					
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.5477	4.55	0.00	0.05	0.01	0.001
ϕ_2	-0.4264	-4.55	0.00	0.05	0.01	0.001
ϕ_3	0.8489	69.67	0.00	0.05	0.01	0.001
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6211	1.56	0.06	0.10	0.09	0.082
Time (in seconds)			0.79	12.2	60.2	622.1
	$p=4$					
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.3633	1.69	0.05	0.05	0.01	0.001
ϕ_2	-0.0251	-0.12	0.45	0.85	0.88	0.865
ϕ_3	0.6305	37.68	0.00	0.05	0.01	0.001
ϕ_4	0.0005	0.03	0.49	0.70	0.65	0.623
σ_y	0.3356	20.06	0.00	0.05	0.01	0.001
σ_v	0.6133	0.67	0.25	0.20	0.15	0.185
Time (in seconds)			0.97	20.7	105.0	1237.2

Notes:

1. Except for ϕ_1 and ϕ_2 parameters of SV(3) model and σ_y in all models, we test each coefficient is zero against a right-sided alternative.
2. We cannot test $\phi_1 = 0$ and $\phi_2 = 0$ in SV(3) model since putting each of these restrictions leads to some of the eigenvalues of the latent AR(3) model are outside the unit circle, hence non-stationarity. In these cases, the ARMA-based estimation is infeasible. So we test $\phi_1 = 0.2$ and $\phi_2 = -0.4$ against a right-sided and a left-sided alternative, respectively.
3. We test $\sigma_y = 0.01$ against a right-sided alternative since when $\sigma_y = 0$, SV models are unidentified.

Table 10: Summary statistics of daily returns and RV estimates.

S&P 500, Full Sample of Experiment - 1								
Series	Mean	SD	Kurtosis	Skewness	Range	Minimum	Maximum	LB(10)
y	0.00	0.67	11.39	-0.17	8.83	4.63	-4.20	46.9
y^2	0.45	1.44	99.59	8.58	21.41	21.41	0.00	1117.5
y^*	0.00	2.67	4.80	-0.93	19.56	6.11	-13.45	473.3
RV5	0.00	0.00	108.51	8.05	0.01	0.01	0.00	4275.8
RV5-SS	0.00	0.00	108.51	8.05	0.01	0.01	0.00	4275.8
BV5	0.00	0.00	91.24	7.73	0.01	0.01	0.00	4362.1
BV5-SS	0.00	0.00	91.24	7.73	0.01	0.01	0.00	4362.1
RV10	0.00	0.00	106.17	8.00	0.01	0.01	0.00	4026.4
RV10-SS	0.00	0.00	106.17	8.00	0.01	0.01	0.00	4026.4
RK	0.00	0.00	54.33	6.35	0.01	0.01	0.00	4273.5
RSV5	0.00	0.00	79.42	7.12	0.00	0.00	0.00	3706.0
RSV5-SS	0.00	0.00	79.42	7.12	0.00	0.00	0.00	3706.0
L-RV5	-9.48	1.20	3.18	0.60	7.29	-4.86	-12.15	7569.5
L-RV5-SS	-9.48	1.20	3.18	0.60	7.29	-4.86	-12.15	7569.5
L-BV5	-9.70	1.20	3.23	0.62	7.71	-5.11	-12.82	7831.3
L-BV5-SS	-9.70	1.20	3.23	0.62	7.71	-5.11	-12.82	7831.3
L-RV10	-9.50	1.22	3.18	0.57	7.54	-4.86	-12.40	7144.9
L-RV10-SS	-9.50	1.22	3.18	0.57	7.54	-4.86	-12.40	7144.9
L-RK	-9.67	1.31	3.09	0.40	7.91	-5.28	-13.19	5562.8
L-RSV5	-10.3	1.30	3.02	0.48	8.02	-5.65	-13.66	6017.0
L-RSV5-SS	-10.3	1.30	3.02	0.48	8.02	-5.65	-13.66	6017.0

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1257$.
2. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the squared of residual return and y_t^* is the residual of log squared of residual return.
3. RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel.
4. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
5. L stands for the logarithmic transformation of realized volatility measures.
6. LB(10) is the heteroskedasticity-corrected Ljung-Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 11: Summary statistics of daily returns and RV estimates.

S&P 500, Full Sample of Experiment - 2								
Series	Mean	SD	Kurtosis	Skewness	Range	Minimum	Maximum	LB(10)
y	0.00	0.65	12.78	-0.18	8.83	4.62	-4.20	53.7
y^2	0.42	1.43	102.64	8.78	21.39	21.39	0.00	1181.7
y^*	0.00	2.69	6.45	-1.14	24.16	6.22	-17.94	466.2
RV5	0.00	0.00	111.55	8.21	0.01	0.01	0.00	4483.6
RV5-SS	0.00	0.00	111.55	8.21	0.01	0.01	0.00	4483.6
BV5	0.00	0.00	93.70	7.88	0.01	0.01	0.00	4547.5
BV5-SS	0.00	0.00	93.70	7.88	0.01	0.01	0.00	4547.5
RV10	0.00	0.00	110.33	8.21	0.01	0.01	0.00	4257.2
RV10-SS	0.00	0.00	110.33	8.21	0.01	0.01	0.00	4257.2
RK	0.00	0.00	55.33	6.44	0.01	0.01	0.00	4425.0
RSV5	0.00	0.00	82.90	7.32	0.00	0.00	0.00	3957.9
RSV5-SS	0.00	0.00	82.90	7.32	0.00	0.00	0.00	3957.9
L-RV5	-9.62	1.21	3.34	0.77	7.29	-4.86	-12.15	8036.3
L-RV5-SS	-9.62	1.21	3.34	0.77	7.29	-4.86	-12.15	8036.3
L-BV5	-9.84	1.21	3.40	0.80	7.71	-5.11	-12.82	8215.8
L-BV5-SS	-9.84	1.21	3.40	0.80	7.71	-5.11	-12.82	8215.8
L-RV10	-9.64	1.23	3.33	0.74	7.54	-4.86	-12.40	7568.4
L-RV10-SS	-9.64	1.23	3.33	0.74	7.54	-4.86	-12.40	7568.4
L-RK	-9.79	1.32	3.28	0.53	8.12	-5.28	-13.40	5826.7
L-RSV5	-10.4	1.31	3.17	0.65	8.02	-5.65	-13.66	6520.8
L-RSV5-SS	-10.4	1.31	3.17	0.65	8.02	-5.65	-13.66	6520.8

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1258$.
2. $y_t = r_t - \hat{\mu}_r$ is the residual return, y_t^2 is the squared of residual return and y_t^* is the residual of log squared of residual return.
3. RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel.
4. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
5. L stands for the logarithmic transformation of realized volatility measures.
6. LB(10) is the heteroskedasticity-corrected Ljung-Box statistics with 10 lags. The critical values for LB(10) are: 15.99 (10%), 18.31 (5%), and 23.21 (1%).

Table 12: Relative MSE and associated MCS p-value during moderate volatility regimes.

	$1-day$			$2-day$			$1-week$		
Model	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.196	0.15	0.09	0.749	0.17	0.10	1.114	0.32	0.37
SV(2)	0.194	0.15	0.09	0.742	0.17	0.10	1.124	0.32	0.35
SV(3)	0.180**	1.00	1.00	0.676**	1.00	1.00	0.982*	0.63	0.60
GARCH(1,1)	0.197	0.15	0.09	0.750	0.17	0.10	1.104	0.32	0.37
GARCH(1,2)	0.205	0.15	0.08	0.777	0.17	0.10	1.144	0.32	0.32
GARCH(2,1)	0.197	0.15	0.09	0.750	0.17	0.10	1.104	0.32	0.37
GARCH(2,2)	0.202	0.15	0.09	0.767	0.17	0.10	1.130	0.32	0.34
GJR(1,1)	0.214	0.15	0.06	0.811	0.13	0.10	1.193	0.20	0.25
GJR(2,2)	0.214	0.15	0.05	0.811	0.13	0.10	1.193	0.20	0.21
EGARCH(1,1)	0.215	0.15	0.04	0.816	0.13	0.10	1.201	0.19	0.15
EGARCH(2,2)	0.215	0.15	0.04	0.815	0.13	0.10	1.198	0.19	0.18
HAR-RV5	1.000	0.15	0.02	1.000	0.13	0.10	1.000*	0.63	0.60
HAR-RV5-SS	1.000	0.15	0.02	1.000	0.13	0.10	1.000*	0.63	0.57
HAR-BV5	1.114	0.15	0.03	1.097	0.13	0.10	1.018*	0.63	0.43
HAR-BV5-SS	1.114	0.15	0.03	1.097	0.13	0.10	1.018*	0.63	0.39
HAR-RV10	0.968	0.15	0.02	0.975	0.17	0.10	0.998*	0.63	0.57
HAR-RV10-SS	0.968	0.15	0.02	0.975	0.13	0.10	0.998*	0.63	0.56
HAR-RK	1.076	0.15	0.02	1.067	0.13	0.10	1.022*	0.63	0.39
HAR-RSV5	0.941	0.15	0.02	0.960	0.13	0.10	0.973*	0.63	0.60
HAR-RSV5-SS	0.941	0.15	0.02	0.960	0.13	0.10	0.973**	1.00	1.00

	$2-week$			$3-week$			$1-month$		
Model	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
SV(1)	1.162	0.35	0.26	1.116	0.26	0.23	1.064	0.10	0.24
SV(2)	1.160	0.24	0.23	1.117	0.14	0.19	1.062	0.10	0.26
SV(3)	1.026*	0.57	0.73	0.985*	0.50	0.75	0.967**	1.00	1.00
GARCH(1,1)	1.102	0.48	0.48	1.056	0.26	0.33	1.005*	0.70	0.79
GARCH(1,2)	1.132	0.47	0.31	1.076	0.26	0.26	1.017	0.34	0.46
GARCH(2,1)	1.102	0.47	0.34	1.056	0.26	0.27	1.005	0.34	0.49
GARCH(2,2)	1.122	0.47	0.33	1.068	0.26	0.26	1.012	0.34	0.49
GJR(1,1)	1.173	0.24	0.21	1.111	0.26	0.24	1.043	0.33	0.40
GJR(2,2)	1.173	0.24	0.19	1.111	0.26	0.23	1.044	0.33	0.38
EGARCH(1,1)	1.180	0.19	0.15	1.117	0.14	0.19	1.049	0.32	0.35
EGARCH(2,2)	1.179	0.19	0.17	1.117	0.14	0.20	1.049	0.10	0.31
HAR-RV5	1.000*	0.57	0.73	1.000	0.26	0.33	1.000	0.34	0.49
HAR-RV5-SS	1.000*	0.57	0.73	1.000	0.26	0.39	1.000	0.34	0.49
HAR-BV5	1.024*	0.57	0.61	1.038	0.26	0.27	1.047	0.33	0.37
HAR-BV5-SS	1.024*	0.57	0.54	1.038	0.26	0.26	1.047	0.33	0.37
HAR-RV10	0.996*	0.57	0.73	0.977**	1.00	1.00	0.974*	0.93	0.93
HAR-RV10-SS	0.996*	0.57	0.73	0.977*	0.50	0.75	0.974*	0.71	0.82
HAR-RK	1.058	0.47	0.35	1.038	0.26	0.26	1.026	0.34	0.43
HAR-RSV5	0.996*	0.57	0.73	0.990*	0.50	0.75	1.007	0.34	0.46
HAR-RSV5-SS	0.996**	1.00	1.00	0.990*	0.50	0.75	1.007	0.34	0.46

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1257$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 504$). The in-sample include most volatile part of late-2000s Financial Crisis.
2. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
3. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively.

Table 13: Relative MAE and associated MCS p-value moderate volatility regimes.

	$1-day$			$2-day$			$1-week$		
Model	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.307	0.00	0.00	0.670	0.00	0.00	0.796	0.23	0.16
SV(2)	0.293	0.38	0.38	0.639*	0.58	0.58	0.787	0.25	0.25
SV(3)	0.291**	1.00	1.00	0.634**	1.00	1.00	0.764**	1.00	1.00
GARCH(1,1)	0.372	0.00	0.00	0.816	0.00	0.00	0.989	0.00	0.00
GARCH(1,2)	0.354	0.00	0.00	0.777	0.00	0.00	0.945	0.00	0.00
GARCH(2,1)	0.372	0.00	0.00	0.816	0.00	0.00	0.989	0.00	0.00
GARCH(2,2)	0.351	0.00	0.00	0.772	0.00	0.00	0.940	0.00	0.00
GJR(1,1)	0.324	0.00	0.00	0.712	0.00	0.00	0.866	0.00	0.00
GJR(2,2)	0.324	0.00	0.00	0.713	0.00	0.00	0.867	0.00	0.00
EGARCH(1,1)	0.326	0.00	0.00	0.717	0.00	0.00	0.872	0.00	0.00
EGARCH(2,2)	0.329	0.00	0.00	0.724	0.00	0.00	0.877	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	1.017	0.00	0.00	1.023	0.00	0.00	1.003	0.00	0.00
HAR-BV5-SS	1.017	0.00	0.00	1.023	0.00	0.00	1.003	0.00	0.00
HAR-RV10	0.995	0.00	0.00	0.995	0.00	0.00	1.000	0.00	0.00
HAR-RV10-SS	0.995	0.00	0.00	0.995	0.00	0.00	1.000	0.00	0.00
HAR-RK	1.032	0.00	0.00	1.041	0.00	0.00	1.022	0.00	0.00
HAR-RSV5	0.970	0.00	0.00	0.979	0.00	0.00	0.981	0.00	0.00
HAR-RSV5-SS	0.970	0.00	0.00	0.979	0.00	0.00	0.981	0.00	0.00

	$2-week$			$3-week$			$1-month$		
Model	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.816	0.00	0.02	0.814	0.25	0.40	0.770	0.06	0.17
SV(2)	0.799*	0.59	0.59	0.807*	0.76	0.76	0.762**	1.00	1.00
SV(3)	0.784**	1.00	1.00	0.796**	1.00	1.00	0.779	0.06	0.17
GARCH(1,1)	1.000	0.00	0.00	1.013	0.00	0.00	0.957	0.00	0.00
GARCH(1,2)	0.943	0.00	0.00	0.949	0.00	0.00	0.893	0.00	0.00
GARCH(2,1)	1.000	0.00	0.00	1.013	0.00	0.00	0.957	0.00	0.00
GARCH(2,2)	0.941	0.00	0.00	0.951	0.00	0.00	0.893	0.00	0.00
GJR(1,1)	0.859	0.00	0.00	0.859	0.00	0.01	0.802	0.02	0.03
GJR(2,2)	0.860	0.00	0.00	0.859	0.00	0.00	0.803	0.02	0.02
EGARCH(1,1)	0.866	0.00	0.00	0.866	0.00	0.00	0.810	0.00	0.00
EGARCH(2,2)	0.872	0.00	0.00	0.873	0.00	0.00	0.817	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	1.013	0.00	0.00	1.018	0.00	0.00	1.024	0.00	0.00
HAR-BV5-SS	1.013	0.00	0.00	1.018	0.00	0.00	1.024	0.00	0.00
HAR-RV10	0.993	0.00	0.00	0.985	0.00	0.00	0.984	0.00	0.00
HAR-RV10-SS	0.993	0.00	0.00	0.985	0.00	0.00	0.984	0.00	0.00
HAR-RK	1.026	0.00	0.00	1.024	0.00	0.00	1.016	0.00	0.00
HAR-RSV5	0.990	0.00	0.00	0.994	0.00	0.00	1.009	0.00	0.00
HAR-RSV5-SS	0.990	0.00	0.00	0.994	0.00	0.00	1.009	0.00	0.00

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1257$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 504$). The in-sample include most volatile part of late-2000s Financial Crisis.
2. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
3. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i\in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j\in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively.

Table 14: Relative R2LOG and associated MCS p-value moderate volatility regimes.

	$1-day$			$2-day$			$1-week$		
Model	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
SV(1)	0.558	0.00	0.00	0.755	0.00	0.00	0.823	0.00	0.00
SV(2)	0.489	0.08	0.08	0.692	0.40	0.40	0.756	0.38	0.38
SV(3)	0.477**	1.00	1.00	0.683**	1.00	1.00	0.742**	1.00	1.00
GARCH(1,1)	0.796	0.00	0.00	1.151	0.00	0.00	1.231	0.00	0.00
GARCH(1,2)	0.755	0.00	0.00	1.091	0.00	0.00	1.166	0.00	0.00
GARCH(2,1)	0.796	0.00	0.00	1.151	0.00	0.00	1.231	0.00	0.00
GARCH(2,2)	0.753	0.00	0.00	1.088	0.00	0.00	1.164	0.00	0.00
GJR(1,1)	0.662	0.00	0.00	0.957	0.00	0.00	1.022	0.00	0.00
GJR(2,2)	0.663	0.00	0.00	0.959	0.00	0.00	1.024	0.00	0.00
EGARCH(1,1)	0.673	0.00	0.00	0.974	0.00	0.00	1.041	0.00	0.00
EGARCH(2,2)	0.679	0.00	0.00	0.988	0.00	0.00	1.051	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	0.979	0.00	0.00	0.990	0.00	0.00	0.991	0.00	0.00
HAR-BV5-SS	0.979	0.00	0.00	0.990	0.00	0.00	0.991	0.00	0.00
HAR-RV10	1.011	0.00	0.00	1.006	0.00	0.00	1.001	0.00	0.00
HAR-RV10-SS	1.011	0.00	0.00	1.006	0.00	0.00	1.001	0.00	0.00
HAR-RK	1.023	0.00	0.00	1.019	0.00	0.00	1.011	0.00	0.00
HAR-RSV5	0.958	0.00	0.00	0.976	0.00	0.00	0.994	0.00	0.00
HAR-RSV5-SS	0.958	0.00	0.00	0.976	0.00	0.00	0.994	0.00	0.00

	$2-week$			$3-week$			$1-month$		
Model	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
SV(1)	0.807	0.00	0.00	0.782	0.01	0.06	0.783	0.04	0.15
SV(2)	0.752	0.46	0.46	0.744*	0.82	0.82	0.751**	1.00	1.00
SV(3)	0.734**	1.00	1.00	0.737**	1.00	1.00	0.757*	0.87	0.87
GARCH(1,1)	1.212	0.00	0.00	1.174	0.00	0.00	1.133	0.00	0.00
GARCH(1,2)	1.146	0.00	0.00	1.107	0.00	0.00	1.063	0.00	0.00
GARCH(2,1)	1.212	0.00	0.00	1.174	0.00	0.00	1.133	0.00	0.00
GARCH(2,2)	1.145	0.00	0.00	1.109	0.00	0.00	1.066	0.00	0.00
GJR(1,1)	1.000	0.00	0.00	0.962	0.00	0.00	0.914	0.01	0.01
GJR(2,2)	1.002	0.00	0.00	0.964	0.00	0.00	0.916	0.01	0.00
EGARCH(1,1)	1.020	0.00	0.00	0.981	0.00	0.00	0.934	0.00	0.00
EGARCH(2,2)	1.032	0.00	0.00	0.994	0.00	0.00	0.950	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	0.991	0.00	0.00	1.000	0.00	0.00	1.002	0.00	0.00
HAR-BV5-SS	0.991	0.00	0.00	1.000	0.00	0.00	1.002	0.00	0.00
HAR-RV10	1.001	0.00	0.00	0.995	0.00	0.00	1.000	0.00	0.00
HAR-RV10-SS	1.001	0.00	0.00	0.995	0.00	0.00	1.000	0.00	0.00
HAR-RK	1.008	0.00	0.00	1.004	0.00	0.00	1.007	0.00	0.00
HAR-RSV5	1.001	0.00	0.00	1.000	0.00	0.00	1.005	0.00	0.00
HAR-RSV5-SS	1.001	0.00	0.00	1.000	0.00	0.00	1.005	0.00	0.00

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 and the number of observations is $T = 1257$. The in-sample is from September 01, 2005 to August 31, 2008 ($T = 753$) and the out-of-sample is from September 01, 2008 to August 31, 2010 ($T = 504$). The in-sample include most volatile part of late-2000s Financial Crisis.
2. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
3. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively.

Table 15: Relative MSE and associated MCS p-value during high volatility regimes.

Model	1 − day			2 − day			1 − week		
	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.199	0.11	0.12	0.753	0.16	0.10	1.122	0.40	0.38
SV(2)	0.197	0.11	0.12	0.744	0.16	0.10	1.132	0.31	0.32
SV(3)	0.181**	1.00	1.00	0.677**	1.00	1.00	0.982*	0.66	0.62
GARCH(1,1)	0.197	0.11	0.12	0.743	0.16	0.10	1.097	0.40	0.40
GARCH(1,2)	0.206	0.11	0.09	0.773	0.16	0.10	1.142	0.31	0.30
GARCH(2,1)	0.197	0.11	0.12	0.743	0.16	0.10	1.097	0.40	0.40
GARCH(2,2)	0.203	0.11	0.10	0.763	0.16	0.10	1.128	0.40	0.35
GJR(1,1)	0.215	0.09	0.05	0.811	0.09	0.09	1.196	0.12	0.23
GJR(2,2)	0.216	0.09	0.04	0.811	0.09	0.09	1.197	0.12	0.18
EGARCH(1,1)	0.217	0.09	0.03	0.817	0.09	0.09	1.205	0.11	0.12
EGARCH(2,2)	0.217	0.09	0.03	0.815	0.09	0.09	1.202	0.11	0.14
HAR-RV5	1.000	0.09	0.02	1.000	0.09	0.09	1.000*	0.66	0.62
HAR-RV5-SS	1.000	0.09	0.02	1.000	0.09	0.09	1.000*	0.66	0.62
HAR-BV5	1.113	0.09	0.02	1.097	0.09	0.09	1.017*	0.66	0.48
HAR-BV5-SS	1.113	0.09	0.02	1.097	0.09	0.09	1.0178	0.66	0.40
HAR-RV10	0.964	0.09	0.02	0.974	0.16	0.10	0.998*	0.66	0.62
HAR-RV10-SS	0.964	0.09	0.02	0.974	0.09	0.09	0.998*	0.66	0.62
HAR-RK	1.073	0.09	0.02	1.065	0.09	0.09	1.022*	0.66	0.40
HAR-RSV5	0.947	0.09	0.02	0.963	0.09	0.09	0.974*	0.66	0.62
HAR-RSV5-SS	0.947	0.09	0.02	0.963	0.09	0.09	0.974**	1.00	1.00

Model	2 − week			3 − week			1 − month		
	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
SV(1)	1.169	0.11	0.21	1.123	0.11	0.19	1.071	0.07	0.17
SV(2)	1.165	0.11	0.24	1.123	0.10	0.15	1.068	0.07	0.20
SV(3)	1.027*	0.38	0.73	0.974**	1.00	1.00	0.956**	1.00	1.00
GARCH(1,1)	1.096	0.38	0.49	1.051	0.24	0.34	1.003*	0.71	0.84
GARCH(1,2)	1.129	0.38	0.32	1.075	0.24	0.26	1.018	0.29	0.42
GARCH(2,1)	1.096	0.38	0.38	1.051	0.24	0.34	1.003	0.29	0.47
GARCH(2,2)	1.121	0.38	0.35	1.067	0.24	0.26	1.013	0.29	0.42
GJR(1,1)	1.176	0.11	0.18	1.115	0.11	0.22	1.049	0.12	0.34
GJR(2,2)	1.177	0.11	0.16	1.115	0.11	0.20	1.049	0.12	0.32
EGARCH(1,1)	1.184	0.11	0.12	1.123	0.10	0.15	1.055	0.10	0.25
EGARCH(2,2)	1.182	0.11	0.14	1.121	0.10	0.17	1.054	0.10	0.27
HAR-RV5	1.000*	0.38	0.73	1.000	0.24	0.38	1.000	0.29	0.47
HAR-RV5-SS	1.000*	0.38	0.73	1.000	0.24	0.34	1.000	0.29	0.42
HAR-BV5	1.024	0.38	0.60	1.037	0.24	0.27	1.043	0.29	0.38
HAR-BV5-SS	1.024	0.38	0.53	1.037	0.24	0.26	1.043	0.12	0.34
HAR-RV10	0.996*	0.64	0.79	0.978**	0.94	0.94	0.976*	0.91	0.90
HAR-RV10-SS	0.996*	0.38	0.73	0.978*	0.52	0.70	0.976*	0.91	0.90
HAR-RK	1.057	0.38	0.38	1.036	0.24	0.26	1.024	0.29	0.39
HAR-RSV5	0.996*	0.64	0.79	0.990*	0.52	0.70	1.006	0.29	0.42
HAR-RSV5-SS	0.996**	1.00	1.00	0.990*	0.52	0.70	1.006	0.29	0.42

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1258$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 504$). In this setting, we forecast a highly volatile period.
2. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
3. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively.

Table 16: Relative MAE and associated MCS p-value during high volatility regimes.

	$1-day$			$2-day$			$1-week$		
Model	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.313	0.01	0.01	0.661	0.00	0.00	0.797	0.19	0.14
SV(2)	0.300	0.18	0.18	0.632	0.48	0.48	0.789	0.19	0.18
SV(3)	0.297**	1.00	1.00	0.626**	1.00	1.00	0.763**	1.00	1.00
GARCH(1,1)	0.339	0.00	0.00	0.719	0.00	0.00	0.881	0.00	0.01
GARCH(1,2)	0.330	0.00	0.00	0.701	0.00	0.00	0.862	0.00	0.01
GARCH(2,1)	0.339	0.00	0.00	0.719	0.00	0.00	0.881	0.00	0.01
GARCH(2,2)	0.328	0.00	0.00	0.697	0.00	0.00	0.858	0.00	0.01
GJR(1,1)	0.319	0.01	0.01	0.677	0.00	0.00	0.832	0.03	0.03
GJR(2,2)	0.319	0.01	0.01	0.678	0.00	0.00	0.832	0.03	0.02
EGARCH(1,1)	0.321	0.01	0.01	0.682	0.00	0.00	0.838	0.00	0.01
EGARCH(2,2)	0.323	0.00	0.00	0.685	0.00	0.00	0.841	0.00	0.01
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.01
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.01
HAR-BV5	1.022	0.00	0.00	1.025	0.00	0.00	1.004	0.00	0.01
HAR-BV5-SS	1.022	0.00	0.00	1.025	0.00	0.00	1.004	0.00	0.01
HAR-RV10	0.990	0.00	0.00	0.992	0.00	0.00	0.998	0.00	0.01
HAR-RV10-SS	0.990	0.00	0.00	0.992	0.00	0.00	0.998	0.00	0.01
HAR-RK	1.021	0.00	0.00	1.033	0.00	0.00	1.018	0.00	0.01
HAR-RSV5	0.979	0.00	0.00	0.989	0.00	0.00	0.987	0.00	0.01
HAR-RSV5-SS	0.979	0.00	0.00	0.989	0.00	0.00	0.987	0.00	0.01

	$2-week$			$3-week$			$1-month$		
Model	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
SV(1)	0.825	0.00	0.02	0.831	0.17	0.27	0.801	0.01	0.30
SV(2)	0.807*	0.59	0.59	0.824*	0.58	0.58	0.792**	1.00	1.00
SV(3)	0.792**	1.00	1.00	0.803**	1.00	1.00	0.799*	0.85	0.85
GARCH(1,1)	0.903	0.00	0.01	0.924	0.00	0.02	0.882	0.00	0.02
GARCH(1,2)	0.871	0.00	0.01	0.884	0.00	0.02	0.842	0.00	0.02
GARCH(2,1)	0.903	0.00	0.01	0.924	0.00	0.02	0.882	0.00	0.02
GARCH(2,2)	0.870	0.00	0.01	0.886	0.00	0.02	0.843	0.00	0.02
GJR(1,1)	0.835	0.00	0.02	0.843	0.16	0.16	0.801	0.01	0.30
GJR(2,2)	0.835	0.00	0.02	0.843	0.16	0.15	0.801	0.01	0.30
EGARCH(1,1)	0.841	0.00	0.01	0.850	0.00	0.02	0.807	0.00	0.02
EGARCH(2,2)	0.844	0.00	0.01	0.852	0.00	0.02	0.810	0.00	0.02
HAR-RV5	1.000	0.00	0.01	1.000	0.00	0.02	1.000	0.00	0.02
HAR-RV5-SS	1.000	0.00	0.01	1.000	0.00	0.02	1.000	0.00	0.02
HAR-BV5	1.015	0.00	0.01	1.019	0.00	0.02	1.022	0.00	0.02
HAR-BV5-SS	1.015	0.00	0.01	1.019	0.00	0.02	1.022	0.00	0.02
HAR-RV10	0.992	0.00	0.01	0.984	0.00	0.02	0.983	0.00	0.02
HAR-RV10-SS	0.992	0.00	0.01	0.984	0.00	0.02	0.983	0.00	0.02
HAR-RK	1.023	0.00	0.01	1.019	0.00	0.02	1.011	0.00	0.02
HAR-RSV5	0.994	0.00	0.01	0.997	0.00	0.02	1.010	0.00	0.02
HAR-RSV5-SS	0.994	0.00	0.01	0.997	0.00	0.02	1.010	0.00	0.02

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1258$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 504$). In this setting, we forecast a highly volatile period.
2. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
3. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively.

Table 17: Relative R2LOG and associated MCS p-value during high volatility regimes.

	$1-day$			$2-day$			$1-week$		
Model	RR2LOG	p_{MCS}^{M}	p_{MCS}^{R}	RR2LOG	p_{MCS}^{M}	p_{MCS}^{R}	RR2LOG	p_{MCS}^{M}	p_{MCS}^{R}
SV(1)	0.526	0.00	0.00	0.710	0.00	0.00	0.797	0.00	0.01
SV(2)	0.469	0.20	0.20	0.657*	0.83	0.83	0.738*	0.56	0.56
SV(3)	0.462**	1.00	1.00	0.655**	1.00	1.00	0.728**	1.00	1.00
GARCH(1,1)	0.641	0.00	0.00	0.918	0.00	0.00	1.016	0.00	0.00
GARCH(1,2)	0.615	0.00	0.00	0.882	0.00	0.00	0.975	0.00	0.00
GARCH(2,1)	0.641	0.00	0.00	0.918	0.00	0.00	1.016	0.00	0.00
GARCH(2,2)	0.614	0.00	0.00	0.879	0.00	0.00	0.973	0.00	0.00
GJR(1,1)	0.564	0.00	0.00	0.809	0.00	0.00	0.893	0.00	0.00
GJR(2,2)	0.565	0.00	0.00	0.811	0.00	0.00	0.895	0.00	0.00
EGARCH(1,1)	0.576	0.00	0.00	0.826	0.00	0.00	0.914	0.00	0.00
EGARCH(2,2)	0.582	0.00	0.00	0.833	0.00	0.00	0.921	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	0.988	0.00	0.00	0.995	0.00	0.00	0.997	0.00	0.00
HAR-BV5-SS	0.988	0.00	0.00	0.995	0.00	0.00	0.997	0.00	0.00
HAR-RV10	1.007	0.00	0.00	1.003	0.00	0.00	0.998	0.00	0.00
HAR-RV10-SS	1.007	0.00	0.00	1.003	0.00	0.00	0.998	0.00	0.00
HAR-RK	1.018	0.00	0.00	1.012	0.00	0.00	1.004	0.00	0.00
HAR-RSV5	0.973	0.00	0.00	0.988	0.00	0.00	1.003	0.00	0.00
HAR-RSV5-SS	0.973	0.00	0.00	0.988	0.00	0.00	1.003	0.00	0.00

	$2-week$			$3-week$			$1-month$		
Model	RR2LOG	p_{MCS}^{M}	p_{MCS}^{R}	RR2LOG	p_{MCS}^{M}	p_{MCS}^{R}	RR2LOG	p_{MCS}^{M}	p_{MCS}^{R}
SV(1)	0.828	0.00	0.00	0.809	0.00	0.01	0.844	0.00	0.02
SV(2)	0.760	0.17	0.17	0.754	0.18	0.18	0.782	0.24	0.24
SV(3)	0.720**	1.00	1.00	0.703**	1.00	1.00	0.730**	1.00	1.00
GARCH(1,1)	1.025	0.00	0.00	1.002	0.00	0.00	0.983	0.00	0.00
GARCH(1,2)	0.979	0.00	0.00	0.952	0.00	0.00	0.928	0.00	0.00
GARCH(2,1)	1.025	0.00	0.00	1.002	0.00	0.00	0.983	0.00	0.00
GARCH(2,2)	0.979	0.00	0.00	0.955	0.00	0.00	0.932	0.00	0.00
GJR(1,1)	0.892	0.00	0.00	0.863	0.00	0.00	0.834	0.07	0.07
GJR(2,2)	0.893	0.00	0.00	0.865	0.00	0.00	0.835	0.07	0.06
EGARCH(1,1)	0.912	0.00	0.00	0.885	0.00	0.00	0.855	0.00	0.01
EGARCH(2,2)	0.918	0.00	0.00	0.893	0.00	0.00	0.861	0.00	0.00
HAR-RV5	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-RV5-SS	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.00	0.00
HAR-BV5	0.999	0.00	0.00	1.004	0.00	0.00	1.002	0.00	0.00
HAR-BV5-SS	0.999	0.00	0.00	1.004	0.00	0.00	1.002	0.00	0.00
HAR-RV10	0.996	0.00	0.00	0.994	0.00	0.00	0.994	0.00	0.00
HAR-RV10-SS	0.996	0.00	0.00	0.994	0.00	0.00	0.994	0.00	0.00
HAR-RK	1.005	0.00	0.00	0.999	0.00	0.00	1.000	0.00	0.00
HAR-RSV5	1.006	0.00	0.00	1.005	0.00	0.00	1.008	0.00	0.00
HAR-RSV5-SS	1.006	0.00	0.00	1.005	0.00	0.00	1.008	0.00	0.00

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 and the number of observations is $T = 1258$. The in-sample is from January 01, 2005 to December 31, 2007 ($T = 754$) and the out-of-sample is from January 01, 2008 to December 31, 2009 ($T = 504$). In this setting, we forecast a highly volatile period.
2. HAR stands for Heterogenous Autoregressive model, RV5 is the 5-minute Realized Variance, BV5 is the 5-minute Bipower Variation, RV10 is the 10-minute Realized Variance, RSV5 is the 5-minute Realized Semivariance, and RK is the Realized Kernel. SS denotes the use of 1-minute subsamples in the calculation of realized volatility estimators.
3. These are relative to the reference model HAR-RV5 and values smaller than unity indicate better forecast performance than the HAR-RV5 model.
4. p_{MCS}^{M} and p_{MCS}^{R} are associated with $MCS\_T_{\max,M} = \max_{i\in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j\in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^{*}_{5\%}$ and $\hat{M}^{*}_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^{*}_{5\%}$ and $\hat{M}^{*}_{50\%}$ are identified by two and one asterisks, respectively.

Table 18: Forecasting realized volatility: relative MSE and associated MCS p-value during moderate volatility regimes.

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000**	0.95	0.95	1.000*	0.69	0.69	1.000*	0.81	0.84	1.000	0.24	0.31	1.000	0.21	0.23	1.000	0.15	0.13
SV(1)	**0.991****	1.00	1.00	**0.972****	1.00	1.00	0.986*	0.81	0.84	0.969	0.24	0.31	0.949	0.21	0.23	0.893	0.15	0.13
SV(2)	0.996**	0.95	0.95	0.986*	0.69	0.69	0.981*	0.81	0.84	0.963	0.24	0.31	0.944	0.21	0.23	0.877	0.15	0.13
SV(3)	1.087	0.16	0.17	1.059	0.24	0.32	**0.972****	1.00	1.00	**0.931****	1.00	1.00	**0.881****	1.00	1.00	**0.833****	1.00	1.00
FTSE100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.54	0.54	1.000	0.46	0.44	1.000*	0.91	0.92	1.000*	0.89	0.83	1.000	0.37	0.41	1.000	0.27	0.13
SV(1)	1.004*	0.54	0.54	0.943*	0.89	0.89	1.009	0.41	0.66	**0.975****	1.00	1.00	**0.953****	1.00	1.00	0.950	0.27	0.14
SV(2)	**0.975****	1.00	1.00	**0.941****	1.00	1.00	**0.996****	1.00	1.00	0.981*	0.89	0.83	0.955*	0.85	0.78	0.947	0.27	0.26
SV(3)	1.046	0.27	0.21	1.019	0.46	0.47	1.001*	0.91	0.92	0.982*	0.89	0.83	0.958*	0.85	0.78	**0.943****	1.00	1.00
NASDAQ100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.82	0.82	1.000*	0.93	0.95	1.000*	0.69	0.68	1.000	0.19	0.36	1.000	0.38	0.53	1.000	0.18	0.26
SV(1)	1.018*	0.80	0.79	1.001*	0.93	0.95	0.989*	0.69	0.68	1.020	0.12	0.16	1.021	0.14	0.22	0.984	0.09	0.14
SV(2)	**0.987****	1.00	1.00	**0.996****	1.00	1.00	**0.976****	1.00	1.00	0.985	0.19	0.36	0.976	0.38	0.53	0.943	0.18	0.26
SV(3)	1.087	0.29	0.25	1.070*	0.60	0.64	1.011*	0.69	0.68	**0.970****	1.00	1.00	**0.965****	1.00	1.00	**0.920****	1.00	1.00
N225	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.38	0.32	1.000	0.16	0.15	1.000**	0.99	0.99	1.000	0.30	0.15	1.000*	0.76	0.73	1.000*	0.78	0.70
SV(1)	0.943*	0.67	0.67	0.956	0.16	0.15	1.060	0.26	0.53	1.061	0.30	0.13	1.051	0.23	0.26	1.024	0.21	0.29
SV(2)	**0.937****	1.00	1.00	**0.922****	1.00	1.00	0.998**	0.99	0.99	0.940	0.30	0.15	0.973*	0.94	0.94	0.980*	0.78	0.70
SV(3)	1.203	0.32	0.19	1.084	0.16	0.15	**0.996****	1.00	1.00	**0.907****	1.00	1.00	**0.973****	1.00	1.00	**0.975****	1.00	1.00
SSMI20	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.65	0.60	1.000*	0.80	0.77	**1.000****	1.00	1.00	1.000*	0.66	0.55	1.000	0.37	0.27	1.000	0.14	0.12
SV(1)	0.968*	0.65	0.60	**0.978****	1.00	1.00	1.003**	0.98	0.97	0.984*	0.66	0.55	0.958	0.37	0.27	0.943	0.14	0.12
SV(2)	**0.963****	1.00	1.00	0.990*	0.80	0.77	1.006**	0.98	0.97	0.974*	0.66	0.55	0.948	0.37	0.27	0.924	0.26	0.26
SV(3)	1.128	0.29	0.24	1.127	0.43	0.46	1.011**	0.98	0.97	**0.961****	1.00	1.00	**0.932****	1.00	1.00	**0.912****	1.00	1.00

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 where the in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. The in-sample include the most volatile part of the late-2000s Financial Crisis.
2. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015).
3. These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i\in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j\in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 19: Forecasting realized volatility: relative MAE and associated MCS p-value during moderate volatility regimes.

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}
HAR-RV	1.000	0.01	0.01	1.000	0.10	0.06	1.000	0.18	0.13	1.000	0.15	0.17	1.000	0.05	0.05	1.000	0.05	0.06
SV(1)	0.936*	0.89	0.89	0.950	0.48	0.48	0.916	0.29	0.26	0.849*	0.85	0.79	0.808	0.17	0.18	0.775	0.21	0.23
SV(2)	**0.934****	1.00	1.00	**0.944****	1.00	1.00	0.905	0.36	0.36	0.847*	0.85	0.79	0.807	0.15	0.18	0.771	0.18	0.23
SV(3)	0.968	0.04	0.08	0.983	0.10	0.07	**0.895****	1.00	1.00	**0.843****	1.00	1.00	**0.785****	1.00	1.00	**0.745****	1.00	1.00
FTSE100	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}
HAR-RV	1.000	0.22	0.29	1.000	0.12	0.12	1.000	0.05	0.02	1.000	0.05	0.03	1.000	0.01	0.00	1.000	0.01	0.00
SV(1)	0.986	0.44	0.44	0.987	0.12	0.16	0.960	0.05	0.04	0.862	0.29	0.20	0.826	0.04	0.02	0.789	0.04	0.04
SV(2)	**0.971****	1.00	1.00	**0.950****	1.00	1.00	0.928	0.25	0.25	0.854	0.31	0.31	0.820	0.04	0.03	0.783	0.04	0.04
SV(3)	1.015	0.17	0.23	0.952*	0.92	0.92	**0.902****	1.00	1.00	**0.848****	1.00	1.00	**0.807****	1.00	1.00	**0.771****	1.00	1.00
NASDAQ100	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}
HAR-RV	1.000	0.32	0.39	1.000	0.10	0.11	1.000	0.22	0.12	1.000	0.10	0.04	1.000	0.13	0.08	1.000	0.11	0.05
SV(1)	**0.971****	1.00	1.00	0.966	0.12	0.17	0.944	0.27	0.19	0.883	0.11	0.07	0.866	0.17	0.18	0.859	0.11	0.05
SV(2)	0.976*	0.80	0.80	**0.943****	1.00	1.00	0.932	0.30	0.30	0.868	0.18	0.18	0.848*	0.72	0.72	0.832	0.11	0.10
SV(3)	1.003	0.18	0.24	0.965	0.23	0.23	**0.920****	1.00	1.00	**0.860****	1.00	1.00	**0.847****	1.00	1.00	**0.822****	1.00	1.00
N225	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}
HAR-RV	1.000	0.06	0.03	1.000	0.01	0.01	1.000	0.05	0.09	1.000	0.09	0.06	1.000	0.13	0.05	1.000	0.09	0.05
SV(1)	0.968	0.07	0.07	0.952	0.02	0.02	0.937	0.05	0.09	0.935	0.09	0.08	0.915	0.13	0.07	0.898	0.11	0.18
SV(2)	**0.950****	1.00	1.00	**0.927****	1.00	1.00	**0.915****	1.00	1.00	0.880	0.32	0.32	0.868*	0.50	0.50	0.865**	0.97	0.97
SV(3)	1.060	0.06	0.03	1.003	0.02	0.02	0.944	0.05	0.09	**0.866****	1.00	1.00	**0.860****	1.00	1.00	**0.865****	1.00	1.00
SSMI20	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}	RMAE	p^{M}_{MCS}	p^{R}_{MCS}
HAR-RV	1.000*	0.49	0.54	1.000*	0.56	0.66	1.000	0.17	0.13	1.000	0.01	0.01	1.000	0.04	0.01	1.000	0.04	0.02
SV(1)	0.995*	0.49	0.54	0.990*	0.56	0.66	0.975	0.17	0.13	0.890	0.02	0.02	0.861	0.05	0.05	0.831	0.13	0.10
SV(2)	**0.987****	1.00	1.00	0.984**	0.98	0.98	0.974	0.17	0.13	0.891	0.01	0.01	0.868	0.04	0.03	0.824	0.13	0.10
SV(3)	1.025	0.40	0.37	**0.984****	1.00	1.00	**0.934****	1.00	1.00	**0.862****	1.00	1.00	**0.839****	1.00	1.00	**0.813****	1.00	1.00

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 where the in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. The in-sample include the most volatile part of the late-2000s Financial Crisis.
2. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015).
3. These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model.
4. p^{M}_{MCS} and p^{R}_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^{*}_{5\%}$ and $\hat{M}^{*}_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^{*}_{5\%}$ and $\hat{M}^{*}_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 20: Forecasting realized volatility: relative R2LOG and associated MCS p-value during moderate volatility regimes.

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.01	0.02	1.000	0.07	0.05	1.000	0.07	0.08	1.000	0.27	0.22	1.000	0.16	0.11	1.000	0.07	0.07
SV(1)	0.921	0.46	0.46	0.933	0.09	0.09	0.939	0.07	0.09	0.887	0.51	0.37	0.853	0.16	0.19	0.782	0.24	0.27
SV(2)	**0.906****	1.00	1.00	**0.915****	1.00	1.00	0.909	0.39	0.39	0.878	0.51	0.39	0.852	0.16	0.19	0.783	0.23	0.27
SV(3)	0.990	0.04	0.09	0.973	0.07	0.07	**0.896****	1.00	1.00	**0.861****	1.00	1.00	**0.809****	1.00	1.00	**0.743****	1.00	1.00
FTSE100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.02	0.02	1.000	0.00	0.01	1.000	0.00	0.01	1.000	0.04	0.02	1.000	0.02	0.00	1.000	0.04	0.01
SV(1)	0.959	0.03	0.03	0.977	0.00	0.01	0.940	0.00	0.01	0.873	0.04	0.03	0.834	0.02	0.00	0.786	0.04	0.02
SV(2)	**0.923****	1.00	1.00	**0.927****	1.00	1.00	0.893*	0.68	0.68	0.849	0.11	0.11	0.812	0.02	0.01	0.770	0.04	0.02
SV(3)	1.009	0.02	0.02	0.964	0.14	0.14	**0.885****	1.00	1.00	**0.829****	1.00	1.00	**0.784****	1.00	1.00	**0.744****	1.00	1.00
NASDAQ100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.13	0.13	1.000	0.06	0.10	1.000	0.05	0.09	1.000	0.19	0.16	1.000	0.18	0.10	1.000	0.14	0.07
SV(1)	**0.944****	1.00	1.00	0.959	0.06	0.10	0.968	0.05	0.09	0.925	0.19	0.16	0.912	0.18	0.10	0.857	0.14	0.12
SV(2)	0.952*	0.66	0.66	**0.936****	1.00	1.00	0.934	0.40	0.40	0.896	0.36	0.36	0.874	0.20	0.20	0.816	0.29	0.29
SV(3)	0.999	0.08	0.08	0.964	0.06	0.10	**0.920****	1.00	1.00	**0.884****	1.00	1.00	**0.858****	1.00	1.00	**0.801****	1.00	1.00
N225	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.00	0.00	1.000	0.00	0.00	1.000	0.01	0.03	1.000	0.07	0.03	1.000	0.09	0.05	1.000	0.10	0.07
SV(1)	0.942	0.03	0.03	0.951	0.00	0.00	0.970	0.01	0.03	0.945	0.07	0.05	0.937	0.09	0.06	0.918	0.10	0.08
SV(2)	**0.897****	1.00	1.00	**0.914****	1.00	1.00	**0.902****	1.00	1.00	0.859*	0.89	0.89	**0.843****	1.00	1.00	**0.834****	1.00	1.00
SV(3)	1.128	0.00	0.00	1.102	0.00	0.00	0.967	0.03	0.03	**0.857****	1.00	1.00	0.861	0.35	0.35	0.854	0.31	0.31
SSMI20	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.17	0.10	1.000	0.01	0.10	1.000	0.08	0.06	1.000	0.11	0.08	1.000	0.00	0.01	1.000	0.02	0.00
SV(1)	0.969	0.17	0.10	0.984	0.01	0.10	0.979	0.08	0.06	0.931	0.11	0.08	0.887	0.02	0.02	0.832	0.02	0.01
SV(2)	**0.948****	1.00	1.00	**0.957****	1.00	1.00	0.955	0.09	0.09	0.919	0.11	0.08	0.877	0.00	0.01	0.813	0.02	0.01
SV(3)	1.015	0.17	0.10	1.007	0.01	0.10	**0.915****	1.00	1.00	**0.881****	1.00	1.00	**0.830****	1.00	1.00	**0.772****	1.00	1.00

Notes:

1. The sample period is from September 01, 2005 to August 31, 2010 where the in-sample is from September 01, 2005 to August 31, 2008 and the out-of-sample is from September 01, 2008 to August 31, 2010. The in-sample include the most volatile part of the late-2000s Financial Crisis.
2. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015).
3. These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 21: Forecasting realized volatility: relative MSE and associated MCS p-value during high volatility regimes.

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.94	0.95	1.000*	0.69	0.72	1.000*	0.84	0.89	1.000	0.23	0.31	1.000	0.18	0.23	1.000	0.13	0.13
SV(1)	**0.994****	1.00	1.00	**0.974****	1.00	1.00	0.992*	0.84	0.89	0.981	0.23	0.31	0.967	0.18	0.23	0.914	0.13	0.13
SV(2)	1.002*	0.94	0.95	0.989*	0.69	0.72	0.989*	0.84	0.89	0.974	0.23	0.31	0.960	0.18	0.23	0.898	0.13	0.13
SV(3)	1.082	0.18	0.21	1.061	0.26	0.35	**0.980****	1.00	1.00	**0.940****	1.00	1.00	**0.895****	1.00	1.00	**0.852****	1.00	1.00
FTSE100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.67	0.63	1.000	0.46	0.39	**1.000****	1.00	1.00	1.000*	0.90	0.93	1.000*	0.70	0.69	1.000	0.18	0.17
SV(1)	0.997*	0.67	0.63	0.949*	0.93	0.93	1.015	0.46	0.67	**0.988****	1.00	1.00	**0.970****	1.00	1.00	0.968	0.18	0.17
SV(2)	**0.977****	1.00	1.00	**0.948****	1.00	1.00	1.004*	0.93	0.93	0.993*	0.90	0.93	0.972*	0.88	0.85	0.963	0.20	0.20
SV(3)	1.058	0.18	0.15	1.024	0.46	0.42	1.008*	0.93	0.93	0.991*	0.90	0.93	0.973*	0.88	0.85	**0.960****	1.00	1.00
NASDAQ100	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.78	0.79	1.000**	0.99	0.99	1.000*	0.71	0.77	1.000	0.22	0.40	1.000	0.37	0.56	1.000	0.19	0.31
SV(1)	1.014*	0.78	0.79	**0.994****	1.00	1.00	0.992*	0.71	0.77	1.033	0.12	0.16	1.038	0.13	0.22	1.004	0.07	0.13
SV(2)	**0.976****	1.00	1.00	0.996**	0.99	0.99	**0.981****	1.00	1.00	0.992	0.22	0.40	0.988	0.37	0.56	0.957	0.19	0.31
SV(3)	1.089	0.33	0.25	1.083*	0.54	0.59	1.020*	0.65	0.69	**0.977****	1.00	1.00	**0.976****	1.00	1.00	**0.934****	1.00	1.00
N225	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.41	0.33	1.000	0.17	0.16	**1.000****	1.00	1.00	1.000	0.32	0.23	1.000**	0.97	0.97	1.000*	0.88	0.88
SV(1)	0.948*	0.62	0.62	0.966	0.17	0.16	1.072	0.25	0.50	1.080	0.32	0.18	1.071	0.23	0.29	1.044	0.19	0.28
SV(2)	**0.942****	1.00	1.00	**0.929****	1.00	1.00	1.008**	0.99	0.98	0.960	0.32	0.23	0.991**	0.97	0.97	0.999*	0.77	0.85
SV(3)	1.210	0.30	0.17	1.092	0.17	0.16	1.006**	0.99	0.98	**0.927****	1.00	1.00	**0.990****	1.00	1.00	**0.994****	1.00	1.00
SSMI20	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}	RMSE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.62	0.59	1.000*	0.74	0.74	**1.000****	1.00	1.00	1.000*	0.72	0.70	1.000	0.28	0.31	1.000	0.10	0.11
SV(1)	**0.958****	1.00	1.00	**0.985****	1.00	1.00	1.015*	0.85	0.86	0.997*	0.62	0.61	0.977	0.28	0.31	0.970	0.10	0.11
SV(2)	0.961*	0.83	0.83	0.998*	0.63	0.73	1.022*	0.83	0.82	0.986*	0.72	0.70	0.961	0.28	0.31	0.945	0.27	0.27
SV(3)	1.128	0.23	0.18	1.115	0.39	0.42	1.016*	0.85	0.86	**0.973****	1.00	1.00	**0.948****	1.00	1.00	**0.934****	1.00	1.00

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 where the in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. In this setting, we forecast a highly volatile period.
2. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015).
3. These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 22: Forecasting realized volatility: relative MAE and associated MCS p-value during high volatility regimes.

	$1-day$			$2-day$			$1-week$			$2-week$			$3-week$			$1-month$		
S&P 500	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.05	0.02	1.000	0.07	0.06	1.000	0.29	0.27	1.000	0.25	0.25	1.000	0.09	0.10	1.000	0.11	0.11
SV(1)	0.945*	0.67	0.67	0.952*	0.53	0.53	0.919*	0.62	0.67	0.862*	0.85	0.90	0.831	0.44	0.40	0.812	0.31	0.26
SV(2)	**0.939****	1.00	1.00	**0.947****	1.00	1.00	0.914*	0.86	0.86	**0.859****	1.00	1.00	0.828	0.44	0.40	0.805	0.31	0.26
SV(3)	0.968	0.06	0.11	0.989	0.07	0.06	**0.912****	1.00	1.00	0.860**	0.95	0.95	**0.812****	1.00	1.00	**0.782****	1.00	1.00
FTSE100	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.19	0.23	1.000	0.13	0.19	1.000	0.12	0.06	1.000	0.11	0.04	1.000	0.02	0.01	1.000	0.06	0.02
SV(1)	0.991	0.26	0.26	0.992	0.13	0.19	0.964	0.12	0.09	0.892	0.20	0.13	0.859	0.06	0.03	0.834	0.16	0.11
SV(2)	**0.972****	1.00	1.00	**0.958****	1.00	1.00	0.942	0.36	0.36	0.885	0.22	0.22	0.855	0.06	0.03	0.829	0.16	0.11
SV(3)	1.026	0.12	0.15	0.967*	0.64	0.64	**0.923****	1.00	1.00	**0.876****	1.00	1.00	**0.842****	1.00	1.00	**0.819****	1.00	1.00
NASDAQ100	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.32	0.33	1.000	0.20	0.18	1.000	0.33	0.26	1.000	0.17	0.06	1.000	0.30	0.20	1.000	0.13	0.09
SV(1)	**0.971****	1.00	1.00	0.963	0.32	0.27	0.940*	0.70	0.63	0.904	0.17	0.07	0.886	0.46	0.39	0.902	0.13	0.09
SV(2)	0.971**	0.98	0.98	**0.950****	1.00	1.00	0.937*	0.70	0.63	0.889	0.17	0.12	0.872*	0.68	0.68	0.871	0.15	0.15
SV(3)	0.998	0.28	0.25	0.976	0.32	0.27	**0.929****	1.00	1.00	**0.879****	1.00	1.00	**0.870****	1.00	1.00	**0.862****	1.00	1.00
N225	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.11	0.08	1.000	0.05	0.03	1.000	0.06	0.12	1.000	0.15	0.09	1.000	0.13	0.06	1.000	0.16	0.13
SV(1)	0.970	0.35	0.35	0.967	0.05	0.03	0.953	0.06	0.12	0.954	0.15	0.10	0.941	0.13	0.06	0.935	0.16	0.15
SV(2)	**0.962****	1.00	1.00	**0.944****	1.00	1.00	**0.931****	1.00	1.00	0.904	0.31	0.31	0.895	0.27	0.27	0.902*	0.88	0.88
SV(3)	1.075	0.04	0.03	1.009	0.05	0.03	0.948	0.45	0.45	**0.889****	1.00	1.00	**0.881****	1.00	1.00	**0.900****	1.00	1.00
SSMI20	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}	RMAE	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000*	0.84	0.83	1.000**	0.99	0.99	1.000	0.11	0.10	1.000	0.02	0.01	1.000	0.09	0.03	1.000	0.11	0.05
SV(1)	0.991**	0.95	0.95	0.998**	0.99	0.99	0.986	0.11	0.10	0.911	0.02	0.01	0.879	0.11	0.11	0.863	0.15	0.15
SV(2)	**0.991****	1.00	1.00	1.000**	0.98	0.99	0.991	0.08	0.10	0.912	0.02	0.01	0.886	0.09	0.07	0.849	0.40	0.40
SV(3)	1.036	0.24	0.25	**0.996****	1.00	1.00	**0.945****	1.00	1.00	**0.883****	1.00	1.00	**0.861****	1.00	1.00	**0.844****	1.00	1.00

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 where the in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. In this setting, we forecast a highly volatile period.
2. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015).
3. These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

Table 23: Forecasting realized volatility: relative R2LOG and associated MCS p-value during high volatility regimes.

	1 – *day*			2 – *day*			1 – *week*			2 – *week*			3 – *week*			1 – *month*		
S&P 500	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.00	0.00	1.000	0.01	0.02	1.000	0.43	0.34	1.000	0.39	0.35	1.000	0.18	0.11	1.000	0.18	0.10
SV(1)	0.925	0.25	0.25	0.922	0.40	0.40	0.933	0.43	0.48	0.914	0.54	0.44	0.915	0.18	0.11	0.870	0.18	0.10
SV(2)	**0.899****	1.00	1.00	**0.913****	1.00	1.00	**0.920****	1.00	1.00	0.908	0.54	0.44	0.897	0.18	0.11	0.850	0.18	0.10
SV(3)	0.968	0.00	0.02	0.981	0.01	0.02	0.921**	0.98	0.98	**0.885****	1.00	1.00	**0.844****	1.00	1.00	**0.788****	1.00	1.00
FTSE100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.01	0.01	1.000	0.00	0.02	1.000	0.01	0.14	1.000	0.07	0.10	1.000	0.03	0.04	1.000	0.05	0.06
SV(1)	0.970	0.01	0.01	0.964	0.00	0.02	0.972	0.01	0.14	0.958	0.07	0.10	0.940	0.03	0.04	0.912	0.05	0.06
SV(2)	**0.929****	1.00	1.00	**0.931****	1.00	1.00	**0.942****	1.00	1.00	0.938	0.15	0.15	0.918	0.03	0.04	0.889	0.05	0.06
SV(3)	1.052	0.01	0.01	1.008	0.00	0.02	0.954*	0.62	0.62	**0.915****	1.00	1.00	**0.886****	1.00	1.00	**0.860****	1.00	1.00
NASDAQ100	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.14	0.10	1.000	0.21	0.14	1.000	0.43	0.42	1.000	0.14	0.14	1.000	0.07	0.06	1.000	0.06	0.06
SV(1)	**0.937****	1.00	1.00	0.952	0.21	0.14	0.958	0.43	0.42	0.960	0.14	0.14	0.952	0.07	0.06	0.931	0.06	0.06
SV(2)	0.944*	0.64	0.64	**0.934****	1.00	1.00	0.941**	0.97	0.97	0.916	0.27	0.27	0.893	0.07	0.06	0.851	0.10	0.10
SV(3)	1.015	0.08	0.05	0.980	0.21	0.14	**0.940****	1.00	1.00	**0.897****	1.00	1.00	**0.865****	1.00	1.00	**0.825****	1.00	1.00
N225	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.05	0.03	1.000	0.02	0.05	1.000	0.36	0.36	1.000	0.41	0.38	1.000*	0.67	0.62	1.000*	0.72	0.73
SV(1)	0.970	0.05	0.04	0.984	0.02	0.05	1.012	0.02	0.14	1.009	0.12	0.15	1.042	0.12	0.19	1.060	0.10	0.23
SV(2)	**0.936****	1.00	1.00	**0.953****	1.00	1.00	**0.951****	1.00	1.00	0.916*	0.84	0.84	**0.924****	1.00	1.00	**0.947****	1.00	1.00
SV(3)	1.182	0.00	0.00	1.136	0.00	0.00	1.005	0.26	0.34	**0.913****	1.00	1.00	0.936*	0.67	0.64	0.965*	0.72	0.73
SSMI20	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}	RR2LOG	p^M_{MCS}	p^R_{MCS}
HAR-RV	1.000	0.18	0.18	1.000	0.26	0.30	1.000	0.35	0.35	1.000	0.29	0.28	1.000	0.02	0.04	1.000	0.02	0.03
SV(1)	**0.957****	1.00	1.00	0.984	0.26	0.30	0.996	0.17	0.18	0.990	0.11	0.13	0.959	0.02	0.04	0.933	0.02	0.03
SV(2)	0.958*	0.92	0.92	**0.973****	1.00	1.00	0.982	0.35	0.35	0.963	0.29	0.28	0.918	0.02	0.04	0.867	0.16	0.16
SV(3)	1.061	0.07	0.05	1.048	0.26	0.21	**0.950****	1.00	1.00	**0.929****	1.00	1.00	**0.879****	1.00	1.00	**0.843****	1.00	1.00

Notes:

1. The sample period is from January 01, 2005 to December 31, 2009 where the in-sample is from January 01, 2005 to December 31, 2007 and the out-of-sample is from January 01, 2008 to December 31, 2009. In this setting, we forecast a highly volatile period.
2. HAR stands for the Heterogenous Autoregressive model, and we used 5-minute RV following the results of Liu et al. (2015).
3. These are relative to the reference model HAR-RV and values smaller than unity indicate better forecast performance than HAR-RV model.
4. p^M_{MCS} and p^R_{MCS} are associated with $MCS\_T_{\max,M} = \max_{i \in M} t_{i,\cdot}$ and $MCS\_T_{R,M} = \max_{i,j \in M} \mid t_{i,j} \mid$, respectively.
5. The forecasts in superior model sets $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are defined by the average of $p_{MCS} \geq 0.95$ and the average of $p_{MCS} \geq 0.50$, respectively.
6. The forecasts in $\hat{M}^*_{5\%}$ and $\hat{M}^*_{50\%}$ are identified by two and one asterisks, respectively. Boldface color font highlights the best model.

A.3. Forecasting with SV(p) models

As discussed earlier, SV(p) models can be written as a linear state-space model without losing any information. The state-space representation of SV(p) models is given by

$$\begin{aligned} y_t^* &= w_t + \varepsilon_t, \\ w_t &= \sum_{j=1}^{p} \phi_j w_{t-j} + v_t, \end{aligned} \tag{A.1}$$

where the distribution ε_t is approximated by a normal distribution with mean 0 and variance $\pi^2/2$. Using the standard notations of Hamilton (1994), the model defined in (A.1) can be rewritten as following:

$$\begin{aligned} y_t &= A'x_t + H'\xi_t + w_t, \\ \xi_{t+1} &= F\xi_t + v_{t+1}, \end{aligned} \tag{A.2}$$

with $y_t = y_t^*$, $A' = 0$, $x_t = 1$, $H' = (1, 0, \dots, 0)$ is a $1 \times p$ vector, $w_t = \varepsilon_t$, $R = \mathbb{E}(w_t w_t') = \pi^2/2$,

$$\xi_t = \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_{t-p-1} \end{bmatrix}, \quad F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, v_{t+1} = \begin{bmatrix} v_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad Q = \mathbb{E}(v_t v_t') = \begin{bmatrix} \sigma_v^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

where F and Q are $p \times p$ matrices, and ξ_t are v_{t+1} are $p \times 1$ vectors. Now using (A.2), the Kalman filter can be applied as follows:

- Initialization:

$$\begin{aligned} \hat{\xi}_{1|0} &= \mathbb{E}(\xi_1) = \mathbf{0}_{(p\times 1)}, \\ \mathbf{P}_{1|0} &= \mathbb{E}([(\xi_1 - \mathbb{E}(\xi_1)][(\xi_1 - \mathbb{E}(\xi_1)]') = diag[\sigma_v^2, \dots, \sigma_v^2]_{(p\times p)}, \end{aligned} \tag{A.3}$$

  where $\mathbf{P}_{1|0}$ is the MSE associated with $\hat{\xi}_{1|0}$.

- Sequential updating:

$$\begin{aligned} \hat{\xi}_{t|t} &= \hat{\xi}_{t|t-1} + \mathbf{P}_{t|t-1}H(H'\mathbf{P}_{t|t-1}H + R)^{-1} \times (y_t - H'\hat{\xi}_{t|t-1}) \\ \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}H(H'\mathbf{P}_{t|t-1}H + R)^{-1}\mathbf{P}_{t|t-1}H'. \end{aligned} \tag{A.4}$$

- In-sample prediction:

$$\begin{aligned} \hat{\xi}_{t+1|t} &= F\hat{\xi}_{t|t-1} + F\mathbf{P}_{t|t-1}H(H'\mathbf{P}_{t|t-1}H + R)^{-1} \times (y_t - H'\hat{\xi}_{t|t-1}) \\ \mathbf{P}_{t+1|t} &= F\mathbf{P}_{t|t}F' + Q. \end{aligned} \tag{A.5}$$

  Given (A.5), the forecast of y_{t+1} and the MSE of forecast error are given by

$$\begin{aligned} \hat{y}_{t+1|t} &= H'\hat{\xi}_{t+1|t} \\ \mathbb{E}([y_{t+1} - \hat{y}_{t+1|t}][y_{t+1} - \hat{y}_{t+1|t}]') &= H'\mathbf{P}_{t+1|t}H + R. \end{aligned} \tag{A.6}$$

- Out-of-sample h-step-ahead forecasting:

$$\begin{aligned}\hat{\xi}_{T+h|T} &= F^h \hat{\xi}_{T|T} \\ \hat{y}_{T+h|T} &= H'\hat{\xi}_{T+h|T} = H'F^h\hat{\xi}_{T|T}.\end{aligned} \tag{A.7}$$

The h-step-ahead forecast is computed by (A.7) with the simple estimates plugged in.

A.4. Forecasting with GARCH models

GARCH Model: The generalized autoregressive conditional heteroskedastic (GARCH) model is an extension of the ARCH model by Engle (1982). If a series exhibits volatility clustering, this suggests that past variances might be predictive of the current variance. The GARCH(p,q) model is an autoregressive moving average model for conditional variances, with p GARCH coefficients associated with lagged variances, and q ARCH coefficients associated with lagged squared innovations or lagged squared residual returns. The GARCH(p,q) model of residual return is

$$\begin{aligned}y_t &= \sigma_t z_t, \quad z_t \sim i.i.d \;\; N(0,1), \\ \sigma_t^2 &= \omega + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_p \sigma_{t-p}^2 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_q y_{t-q}^2,\end{aligned}$$

where y_t is the residual return observed at time t and σ_t is the corresponding volatility. For stationarity and positivity, the GARCH model has the following constraints:

- $\omega > 0$,
- $\beta_i \geq 0$, $\alpha_j \geq 0$
- $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} \alpha_j < 1$.

The h-step-ahead forecast of the GARCH$(1,1)$ model is computed according to:

$$\begin{aligned}\hat{\sigma}_{t+h|t}^2 &= \hat{\omega} + \hat{\beta}_1 \hat{\sigma}_{t+h-1|t}^2 + \hat{\alpha}_1 \hat{y}_{t+h-1|t}^2, \\ \hat{y}_{t+h|t}^2 &= \hat{\sigma}_{t+h|t}^2 \quad if \quad h > 0, \\ \hat{y}_{t+h|t}^2 &= y_{t+h}^2 \quad \hat{\sigma}_{t+h|t}^2 = \sigma_{t+h}^2 \quad if \quad h \leq 0.\end{aligned}$$

EGARCH Model: The exponential GARCH (EGARCH) model was developed by Nelson (1991). It is a GARCH variant that models the logarithm of the conditional variance process. In addition to modeling the logarithm, the EGARCH model has additional leverage terms to capture asymmetry in volatility clustering. The EGARCH(p,q) model has p GARCH coefficients associated with lagged log variance terms, q ARCH coefficients associated with the magnitude of lagged standardized innovations, and q leverage coefficients associated with signed, lagged standardized innovations. The form of the EGARCH(p,q) model is

$$\begin{aligned}y_t &= \sigma_t z_t, \quad z_t \sim i.i.d \;\; N(0,1), \\ \log \sigma_t^2 &= \omega + \sum_{i=1}^{p} \beta_i \log \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j \left(|z_{t-j}| - \mathbb{E}\left(|z_{t-j}|\right)\right) + \sum_{j=1}^{q} \gamma_j z_{t-j},\end{aligned}$$

where $z_t := y_t \sigma_t^{-1}$ and to ensure stationarity, all roots of the GARCH coefficient polynomial, $(1 - \beta_1 L - \ldots - \beta_p L^p)$, must lie outside the unit circle. The h-step-ahead forecast of the EGARCH$(1,1)$ model is computed according to:

$$\log \hat{\sigma}^2_{t+h|t} = \hat{\omega} + \hat{\beta}_1 \log \hat{\sigma}^2_{t+h-1|t} + \hat{\alpha}_1 \left(|\hat{z}_{t+h-1|t}| - \mathbb{E}\left(|\hat{z}_{t+h-1|t}| \right) \right) + \gamma_1 \hat{z}_{t+h-1|t}.$$

GJR Model: The GJR-GARCH, or just GJR, model of Glosten et al. (1993) allows the conditional variance to respond differently to the past negative and positive innovations. The GJR(p,q) model has p GARCH coefficients associated with lagged variances, q ARCH coefficients associated with lagged squared innovations, and q leverage coefficients associated with the square of negative lagged innovations. The GJR(p,q) model may be expressed as:

$$y_t = \sigma_t z_t, \quad z_t \sim i.i.d \;\; N(0,1),$$
$$\log \sigma_t^2 = \omega + \sum_{i=1}^{p} \beta_i \sigma^2_{t-i} + \sum_{j=1}^{q} (\alpha_j + \gamma_j I_{[y_{t-j}<0]}) y^2_{t-j},$$

where the indicator function $I[y_{t-j} < 0]$ equals 1 if $y_{t-j} < 0$, and 0 otherwise. Thus, the leverage coefficients are applied to negative innovations, giving negative changes additional weight. For stationarity and positivity, the GJR model has the following constraints:

- $\omega > 0$
- $\beta_i \geq 0$, $\alpha_j \geq 0$
- $\alpha_j + \gamma_j \geq 0$
- $\sum_{i=1}^{p} \beta_i + \sum_{j=1}^{q} (\alpha_j + \frac{1}{2}\gamma_j) < 1$

The GARCH model is nested in the GJR model. If all leverage coefficients are zero, then the GJR model reduces to the GARCH model. The recursive formula for the h-step-ahead forecast of the GJR-GARCH$(1,1)$ model is calculated as:

$$\hat{\sigma}^2_{t+h|t} = \hat{\omega} + \left(\hat{\alpha}_1 + \frac{\hat{\gamma}_1}{2} + \hat{\beta}_1 \right) \hat{\sigma}^2_{t+h-1|t}.$$

A.5. Realized volatility

Let $p_t = \log S_t$ denote the logarithmic of price where S_t is the observed price (at time t) and $r_t = p_t - p_{t-1}$ denote the continuously compounded return from time $t-1$ to t. Assume that the logarithmic price process, p_t, could belong to the class of continuous-time jump diffusion processes,

$$dp_t = \mu_t d_t + \sigma_t dW_t + \kappa_t dq_t, \quad 0 \leq t \leq T \tag{A.1}$$

where μ_t is a continuous and locally bounded variation process and σ_t is a stochastic volatility process; W_t is the standard Brownian motion; dq_t is a counting process such that $dq_t = 1$ represents a jump at time t (and $dq_t = 0$ if no jump) with jump intensity λ_t. If p_{t-} denotes the price immediately prior to the jump at time t, then $\kappa_t = \Delta p_t = p_t - p_{t-}$. The process p_t is consists of a continuous component and a pure jump component. The quadratic variation (QV) of this process is defined by

$$[r,r]_t = \int_0^t \sigma_s^2 dW_s + \sum_{0<s\leq t} \kappa_s^2, \tag{A.2}$$

where the first component, called integrated volatility, comes from the continuous component of (A.1) and the second term is the contribution from discrete jumps. In the absence of jumps, the second term on the right-hand side disappears and the quadratic variation is simply equal to the integrated volatility (IV).

Now, define the intraday return, r_{t_j} , as the difference between two logarithmic prices,

$$r_{t_j} = p_{t_j} - p_{t_{j-1}},$$

where t_j denotes the j-th intraday observation on the t-th day. Let Δ denote the discrete intraday sample period of length, $t_j - t_{j-1}$. The realized volatility (RV) is defined as the sum of squared intraday returns,

$$RV_t = \sum_{j=1}^{n} r_{t_j}^2,$$

where n is the number of Δ-returns during the t-th time horizon (such as a trading day) and is assumed to be an integer. Andersen, Bollerslev, Diebold and Labys (2001) showed that RV is a natural estimator for the QV. Furthermore, The realized volatility satisfies

$$\lim_{\Delta\to 0} RV_t = \int_0^t \sigma_s^2 dW_s + \sum_{0<s\leq t} \kappa_s^2, \tag{A.3}$$

which means that RV_t is a consistent estimator of the QV.

A.6. Heterogenous Autoregressive model of Realized Volatility

Heterogenous Autoregressive model of Realized Volatility (HAR-RV) model proposed by Corsi (2009). In financial markets, either traders are perceived to be heterogeneous in the sense of a different horizon of investments [Müller, Dacorogna, Davé, Olsen, Pictet and von Weizsäcker (1997)] or information arrival is heterogeneous [Andersen and Bollerslev (1998)]. HAR-RV model takes into account the long memory feature, and among the models proposed to forecast volatility, it stands out in terms of performance and simplicity.

A generalized version of HAR-RV model that we used here is as follows:

$$\log RV_{t+1}^{(d)} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + u_{t+1}^d \tag{A.1}$$

where

$$\log RV_t^{(w)} = \frac{1}{5} \sum_{j=0}^{4} \log RV_{t-j}^{(d)},$$

$$\log RV_t^{(m)} = \frac{1}{22} \sum_{j=0}^{21} \log RV_{t-j}^{(d)}.$$

This class of models can be estimated with ordinary least squares. For the details of forecasting in HAR-RV model, see Corsi (2009).

References

Abramowitz, M. and Stegun, I. A. (1970), *Handbook of Mathematical functions*, Dover Publications, New York.

Ahsan, M. N. and Dufour, J.-M. (2015), A simple quasi-efficient estimator for stochastic volatility models, Technical report, McGill University.

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* **18**(2), 351–416.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2011), 'Ultra high frequency volatility estimation with dependent microstructure noise', *Journal of Econometrics* **160**(1), 160–175.

Alizadeh, S., Brandt, M. W. and Diebold, F. X. (2002), 'Range-based estimation of stochastic volatility models', *The Journal of Finance* **57**(3), 1047–1091.

Andersen, T. G. and Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International economic review* pp. 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of financial economics* **61**(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American statistical association* **96**(453), 42–55.

Andersen, T. G., Chung, H.-J. and Sørensen, B. E. (1999), 'Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Econometrics* **91**, 61–87.

Andersen, T. G. and Sørensen, B. E. (1996), 'GMM estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Business and Economic Statistics* **14**(3), 328–352.

Andrews, D. W. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica: Journal of the Econometric Society* pp. 817–858.

Asai, M. (2008), 'Autoregressive stochastic volatility models with heavy-tailed distributions: A comparison with multifactor volatility models', *Journal of Empirical Finance* **15**(2), 332–341.

Bandi, F. M. and Russell, J. R. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics* **79**(3), 655–692.

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**, 294.

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* **76**(6), 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2011), 'Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading', *Journal of Econometrics* **162**(2), 149–169.

Barndorff-Nielsen, O. E., Kinnebrock, S. and Shephard, N. (2010), Measuring downside risk—realised semivariance, *in* T. Bollerslev, J. Russell and M. Watson, eds, 'Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle', New York: Oxford University Press.

Barndorff-Nielsen, O. E., Nicolato, E. and Shephard, N. (2002), 'Some recent developments in stochastic volatility modelling', *Quantitative Finance* **2**(1), 11–23.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), 'Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2), 167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2006), 'Econometrics of testing for jumps in financial economics using bipower variation', *Journal of financial Econometrics* **4**(1), 1–30.

Benati, L. (2008), 'The "great moderation" in the United Kingdom', *Journal of Money, Credit and Banking* **40**(1), 121–147.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (2009), Glossary to ARCH (GARCH), *in* 'Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle', Oxford University Press.

Bollerslev, T. and Zhou, H. (2002), 'Estimating stochastic volatility diffusion using conditional moments of integrated volatility', *Journal of Econometrics* **109**(1), 33–65.

Bougerol, P. and Picard, N. (1992), 'Stationarity of GARCH processes and of some nonnegative time series', *Journal of Econometrics* **52**(1), 115–127.

Box, G. E., Jenkins, G. M. and Reinsel, G. C. (2013), *Time series analysis: forecasting and control*, Wiley. com.

Breidt, F. J. and Carriquiry, A. L. (1996), Improved quasi-maximum likelihood estimation for stochastic volatility models, *in* 'Modelling and Prediction Honoring Seymour Geisser', Springer, pp. 228–247.

Broto, C. and Ruiz, E. (2004), 'Estimation methods for stochastic volatility models: a survey', *Journal of Economic Surveys* **18**(5), 613–649.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica* **60**, 173–180.

Carnero, M. A., Peña, D. and Ruiz, E. (2004), 'Persistence and kurtosis in GARCH and stochastic volatility models', *Journal of Financial Econometrics* **2**(2), 319–342.

Carrasco, M. and Chen, X. (2002), 'Mixing and moment properties of various GARCH and stochastic volatility models', *Econometric Theory* **18**(1), 17–39.

Chan, J. C. and Grant, A. L. (2016), 'Modeling energy price dynamics: GARCH versus stochastic volatility', *Energy Economics* **54**, 182–189.

Chao, J. and Swanson, N. R. (2007), 'Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction', *Journal of Econometrics* **137**(2), 515–555.

Chernov, M., Gallant, A. R., Ghysels, E. and Tauchen, G. (2003), ‘Alternative models for stock price dynamics’, *Journal of Econometrics* **116**(1), 225–257.

Chib, S., Nardari, F. and Shephard, N. (2002), ‘Markov chain Monte Carlo methods for stochastic volatility models’, *Journal of Econometrics* **108**, 281–316.

Cogley, T. and Sargent, T. J. (2005), ‘Drifts and volatilities: monetary policies and outcomes in the post WWII US’, *Review of Economic dynamics* **8**(2), 262–302.

Corsi, F. (2009), ‘A simple approximate long-memory model of realized volatility’, *Journal of Financial Econometrics* **7**(2), 174–196.

Danielsson, J. (1994), ‘Stochastic volatility in asset prices: Estimation with simulated maximum likelihood’, *Journal of Econometrics* **61**, 375–400.

Danielsson, J. and Richard, J.-F. (1993), ‘Accelerated Gaussian importance sampler with application to dynamic latent variable models’, *Journal of Applied Econometrics* **8**, S153–S173.

Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians*, first edn, Oxford University Press, Oxford, U.K.

Davis, R. A. and Mikosch, T. (2009), Probabilistic properties of stochastic volatility models, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, ‘Handbook of Financial Time Series’, Springer-Verlag, Berlin, Heidelberg, pp. 255–267.

Diebold, F. X. and Mariano, R. S. (2002), ‘Comparing predictive accuracy’, *Journal of Business & economic statistics* **20**(1), 134–144.

Dufour, J.-M. (2006), ‘Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics’, *Journal of Econometrics* **133**(2), 443–477.

Dufour, J.-M. and Jouini, T. (2006), ‘Finite-sample simulation-based tests in VAR models with applications to Granger causality testing’, *Journal of Econometrics* **135**(1-2), 229–254.

Dufour, J.-M. and Jouini, T. (2014), ‘Asymptotic distributions for quasi-efficient estimators in echelon VARMA models’, *Computational Statistics and Data Analysis* **73**, 69 – 86.

Dufour, J.-M. and Khalaf, L. (2001), ‘Monte carlo test methods in econometrics’, *Companion to Theoretical Econometrics’, Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, UK* pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002), ‘Simulation based finite and large sample tests in multivariate regressions’, *Journal of Econometrics* **111**(2), 303–322.

Dufour, J.-M. and Pelletier, D. (2005), ‘Practical methods for modelling weak VARMA processes: identification, estimation and specification with a macroeconomic application’, *Preprint* .

Dufour, J.-M. and Taamouti, M. (2003), Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Valéry, P. (2006), ‘On a simple two-stage closed-form estimator for a stochastic volatility in a general linear regression’, *Advances in Econometrics* **20**, 259–288.

Dufour, J.-M. and Valéry, P. (2009), 'Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models', *Journal of Econometrics* **150**, 193–206.

Durbin, J. (1960), 'The fitting of time series models', *Revue de l'Institut International de Statistique* **28**, 233–244.

Durham, G. B. (2006), 'Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models', *Journal of Econometrics* **133**, 273–305.

Durham, G. B. (2007), 'SV mixture models with application to S&P 500 index returns', *Journal of Financial Economics* **85**(3), 822–856.

Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**, 181–187.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**(4), 987–1008.

Fama, E. F. (1965), 'The behavior of stock-market prices', *The Journal of Business* **38**(1), 34–105.

Fan, J. and Wang, Y. (2007), 'Multi-scale jump and volatility analysis for high-frequency financial data', *Journal of the American Statistical Association* **102**(480), 1349–1362.

Flury, T. and Shephard, N. (2011), 'Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models', *Econometric Theory* **27**(05), 933–956.

Francq, C. and Zakoïan, J.-M. (2006), 'Linear-representation based estimation of stochastic volatility models', *Scandinavian journal of statistics* **33**(4), 785–806.

Francq, C. and Zakoïan, J.-M. (2011), *GARCH models: structure, statistical inference and financial applications*, John Wiley & Sons.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, Vol. 230, John Wiley & Sons.

Galbraith, J. W. and Zinde-Walsh, V. (1997), 'On some simple, autoregression-based estimation and identification techniques for ARMA models', *Biometrika* **84**(3), 685–696.

Gallant, A. R. (1987), *Nonlinear Statistical Models*, John Wiley & Sons, New York.

Gallant, A. R., Hsieh, D. and Tauchen, G. (1997), 'Estimation of stochastic volatility models with diagnostics', *Journal of Econometrics* **81**(1), 159–192.

Gallant, A. R., Hsu, C.-T. and Tauchen, G. (1999), 'Using daily range data to calibrate volatility diffusions and extract the forward integrated variance', *Review of Economics and Statistics* **81**(4), 617–631.

Gallant, A. R. and Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657 – 681.

Gatheral, J. and Oomen, R. C. (2010), 'Zero-intelligence realized variance estimation', *Finance and Stochastics* **14**(2), 249–283.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000), 'Stochastic volatility models as hidden Markov models and statistical applications', *Bernoulli* **6**(6), 1051–1079.

Ghysels, E., Harvey, A. and Renault, E. (1996), Stochastic volatility, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 14: Statistical Methods in Finance', Elsevier.

Giraitis, L. and Robinson, P. M. (2001), 'Whittle estimation of ARCH models', *Econometric Theory* **17**(3), 608–631.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The journal of finance* **48**(5), 1779–1801.

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**, 65–99.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K. Translated by Quang Vuong.

Granger, C. W. and Morris, M. J. (1976), 'Time series modelling and interpretation', *Journal of the Royal Statistical Society. Series A (General)* pp. 246–257.

Hafner, C. M. and Linton, O. (2017), 'An almost closed form estimator for the EGARCH model', *Econometric Theory* **33**(4), 1013–1038.

Hamilton, J. D. (1994), *Time series analysis*, Vol. 2, Princeton university press Princeton, NJ.

Hannan, E. J. and Rissanen, J. (1982), 'Recursive estimation of mixed autoregressive-moving average order', *Biometrika* **69**(1), 81–94.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.

Hansen, P. R. (2005), 'A test for superior predictive ability', *Journal of Business & Economic Statistics* **23**(4), 365–380.

Hansen, P. R. and Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business & Economic Statistics* **24**(2), 127–161.

Hansen, P. R., Lunde, A. and Nason, J. M. (2003), 'Choosing the best volatility models: the model confidence set approach', *Oxford Bulletin of Economics and Statistics* **65**, 839–861.

Hansen, P. R., Lunde, A. and Nason, J. M. (2011), 'The model confidence set', *Econometrica* **79**(2), 453–497.

Harvey, A. C. and Shephard, N. (1996), 'Estimation of an asymmetric stochastic volatility model for asset returns', *Journal of Business and Economic Statistics* **14**(4), 429–434.

Harvey, A., Ruiz, E. and Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.

Heber, G., Lunde, A., Shephard, N. and Sheppard, K. (2009), 'Oxford-man institute's realized library, version 0.3'.

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* **119**(7), 2249–2276.

Jacquier, E., Polson, N. and Rossi, P. (1994), 'Bayesian analysis of stochastic volatility models (with discussion)', *Journal of Economics and Business Statistics* **12**, 371–417.

Kalnina, I. and Linton, O. (2008), 'Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error', *Journal of Econometrics* **147**(1), 47–59.

Kambouroudis, D. S. and McMillan, D. G. (2015), 'Is there an ideal in-sample length for forecasting volatility?', *Journal of International Financial Markets, Institutions and Money* **37**, 114–137.

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**, 361–393.

Knight, J. L., Satchell, S. E. and Yu, J. (2002), 'Theory & methods: Estimation of the stochastic volatility model by the empirical characteristic function method', *Australian and New Zealand Journal of Statistics* **44**(3), 319–335.

Koop, G. and Korobilis, D. (2013), 'Large time-varying parameter VARs', *Journal of Econometrics* **177**(2), 185–198.

Koop, G., Leon-Gonzalez, R. and Strachan, R. W. (2009), 'On the evolution of the monetary policy transmission mechanism', *Journal of Economic Dynamics and Control* **33**(4), 997–1017.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005), 'Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* **12**(3), 445–475.

Koreisha, S. and Pukkila, T. (1990), 'A generalized least squares approach for estimation of autoregressive moving average models', *Journal of Time Series Analysis* **11**(2), 139–151.

Kristensen, D. and Linton, O. (2006), 'A closed-form estimator for the GARCH(1,1) model', *Econometric Theory* **22**, 323–337.

Li, Y. and Mykland, P. A. (2007), 'Are volatility estimators robust with respect to modeling assumptions?', *Bernoulli* pp. 601–622.

Lindner, A. M. (2009), Stationarity, mixing, distributional properties and moments of GARCH(p, q)–processes, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, Heidelberg, pp. 43–69.

Liu, L. Y., Patton, A. J. and Sheppard, K. (2015), 'Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes', *Journal of Econometrics* **187**(1), 293–311.

Liu, Y. and Morley, J. (2014), 'Structural evolution of the postwar US economy', *Journal of Economic Dynamics and Control* **42**, 50–68.

Mandelbrot, B. (1963), 'The variation of certain speculative prices', *The Journal of Business* **36**, 394–419.

Meddahi, N. (2001), An eigenfunction approach for volatility modeling, Technical report, CRDE, Université de Montréal, and CIRANO.

Meddahi, N. (2003), 'ARMA representation of integrated and realized variances', *The Econometrics Journal* **6**(2), 335–356.

Melino, A. and Turnbull, S. M. (1990), 'Pricing foreign currency options with stochastic volatility', *Journal of Econometrics* **45**(1), 239–265.

Monfardini, C. (1998), 'Estimating stochastic volatility models through indirect inference', *The Econometrics Journal* **1**(1), C113–C128.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997), 'Volatilities of different time resolutions—analyzing the dynamics of market components', *Journal of Empirical Finance* **4**(2), 213–239.

Nelson, D. (1988), Times Series Behavior of Stock Market Volatility and Returns, PhD thesis, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1,1) model', *Econometric theory* **6**(03), 318–334.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: A new approach', *Econometrica* pp. 347–370.

Newey, W. K. and McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of econometrics* **4**, 2111–2245.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Newey, W. K. and West, K. D. (1994), 'Automatic lag selection in covariance matrix estimation', *The Review of Economic Studies* **61**(4), 631–653.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007), 'Stochastic volatility with leverage: Fast and efficient likelihood inference', *Journal of Econometrics* **140**(2), 425–449.

Pagan, A. R. and Schwert, G. W. (1990), 'Alternative models for conditional stock volatility', *Journal of econometrics* **45**(1-2), 267–290.

Poon, S.-H. and Granger, C. W. (2003), 'Forecasting volatility in financial markets: A review', *Journal of Economic Literature* **41**(2), 478–539.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *The Review of Economic Studies* **72**(3), 821–852.

Richard, J.-F. and Zhang, W. (2007), 'Efficient high-dimensional importance sampling', *Journal of Econometrics* **141**(2), 1385–1411.

Rudin, W. (1976), *Principles of Mathematical Analysis, Third Edition*, McGraw-Hill, New York.

Ruiz, E. (1994), 'Quasi-maximum likelihood estimation of stochastic variance models', *Journal of Econometrics* **63**, 284–306.

Saikkonen, P. (1986), 'Asymptotic properties of some preliminary estimators for autoregressive moving average time series models', *Journal of Time Series Analysis* **7**(2), 133–155.

Sandmann, G. and Koopman, S. J. (1998), 'Estimation of stochastic volatility models via Monte Carlo maximum likelihood', *Journal of Econometrics* **87**(2), 271–301.

Shephard, N. (1994), 'Partial non-Gaussian state space', *Biometrika* **81**(1), 115–131.

Shephard, N. (2005), *Stochastic volatility: selected readings*, Oxford University Press.

Shephard, N. and Andersen, T. G. (2009), Stochastic volatility: Origins and overview, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, Heidelberg, pp. 233–254.

Steel, M. F. (1998), 'Bayesian analysis of stochastic volatility models with flexible tails', *Econometric Reviews* **17**(2), 109–143.

Straumann, D. (2005), *Estimation in conditionally heteroscedastic time series models*, Vol. 181 of *Lecture Notes in Statistics*, Springer, Berlin Heidelberg.

Taylor, S. J. (1982), Financial returns modelled by the product of two stochastic processes – a study of daily sugar prices, *in* O. D. Anderson, ed., 'Time Series Analysis: Theory and Practice', Vol. 1, North-Holland, Amsterdam, pp. 203–226.

Taylor, S. J. (1986), *Modelling Financial Time Series*, John Wiley and Sons, New York.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Mathematical finance* **4**(2), 183–204.

Tsay, R. S. and Tiao, G. C. (1984), 'Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models', *Journal of the American Statistical Association* **79**(385), 84–96.

West, K. D. (1996), 'Asymptotic inference about predictive ability', *Econometrica: Journal of the Econometric Society* pp. 1067–1084.

White, H. and Domowitz, I. (1984), 'Nonlinear regression with dependent observations', *Econometrica: Journal of the Econometric Society* pp. 143–161.

Wishart, J. (1947), 'The cumulants of the z and of the logarithmic χ^2 and t distributions', *Biometrika* **34**(1/2), 170–178.

Xiu, D. (2010), 'Quasi-maximum likelihood estimation of volatility with high frequency data', *Journal of Econometrics* **159**(1), 235–250.

Yu, J. (2002), 'Forecasting volatility in the New Zealand stock market', *Applied Financial Economics* **12**(3), 193–202.

Zhang, L. (2006), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* **12**(6), 1019–1043.

Zhang, L., Mykland, P. A. 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* **100**(472), 1394–1411.