Simple Estimators and Inference for Higher-order Stochastic Volatility Models *

Md. Nazmul Ahsan† McGill University

Jean-Marie Dufour ‡ McGill University

First version: December 2013

Revised:

This version:

Compiled: October 26, 2017, 5:06

*For helpful comments, we thank Manabu Asai, John W. Galbraith, as well as participants at CIREQ PhD Students Conference and the Econometric Society of World Congress. This paper was previously circulating under the title "Closed-Form Estimator for a Higher-order Gaussian Stochastic Volatility Models."

† 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 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

ABSTRACT

We propose several estimators for higher-order stochastic volatility models denoted by SV(p) where the latent volatility process modeled using an AR(p) process. We discuss the motivation and stationarity, ergodicity and mixing properties of SV(p) models. Proposed estimators include two simple estimators and GMM estimators. In literature, several methods have been proposed to estimate SV(1) model, and mostly these methods are costly from the computational viewpoint, inflexible across models, not easy to implement and converge very slowly. Compared to these methods, our simple estimators for SV(p) models are computationally simple and very easy to apply in practice. These estimators do not require choosing a sampling algorithm, initial parameters, and an auxiliary model. We also develop recursive estimation procedures for SV(p) models based on these simple estimators. We derive asymptotic theories for these estimators under standard regulatory assumptions 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 popular Bayesian MCMC estimator. The simple ARMA based estimator, suggested by this study, mostly outperforms other estimators in terms of bias and RMSE. For a larger sample, it uniformly superior to other estimators. Finally, two empirical applications are presented where SV(p) models are estimated by the simple ARMA based estimator. First, SV(p) models fitted with S&P 500 index returns, and we found that a higher-order SV model can better model these returns. We also implemented MC tests to construct more reliable inference and found evidence to support above result. Second, we conducted out-of-sample forecasting experiments to study the accuracy of volatility forecasts between SV(p) models and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) models. The results suggested that SV(p) models performed better than HAR-RV models for forecasting daily volatility. This result is consistent whether high volatility periods (such as Financial Crisis) are in the in-sample or in the out-of-sample. Our findings highlight the importance of higher-order SV models for forecasting volatility.

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

List of Definitions, Assumptions, Propositions and Theorems

1. Introduction

Time-varying volatility is endemic in financial markets. Such features known for a long time, early comments include Mandelbrot (1963) and Fama (1965). Two main classes of parametric models proposed in the literature to model this feature: (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 latter has an additional error term in the variance. 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 important in macroeconomics following the seminal work of Cogley and Sargent (2005) and Primiceri (2005); recent papers include Benati (2008), Koop, Leon-Gonzalez and Strachan (2009), Koop and Korobilis (2013), Liu and Morley (2014).

SV models may be preferable to GARCH-type models for several reasons. *First*, SV models overcome the drawbacks encountered with ARCH models and connected to diffusion processes used in theoretical finance; see Shephard and Andersen (2009). In particular, SV models have simple continuous-time analogues 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 to give the model additional flexibility against misspecification. These modifications substantially improve the performance of the standard GARCH, but these are apparently unnecessary for SV model; 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*, one can easily derive the stationarity, ergodicity and mixing properties of an SV model than a GARCH model which includes explicit feedback of the current volatility with previous volatilities and observations; see Davis and Mikosch (2009). Conditions for the existence of a stationary GARCH process are much more difficult to establish; see Nelson (1990), Bougerol and Picard (1992) and Lindner (2009).

Despite these attractive features, SV models used much 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. *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 standard SV model are available). Also, it seems to be that a single estimation method (such as QML) is applicable for many variants of the GARCH model, whereas there are only a few variants of the SV model but a full series of estimation methods and it is not clear that which one is more efficient compared to others.

Several methods have been proposed to estimate SV(1) models where the latent volatility is modeled as an AR(1) process. 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)], method based on linear representations [Francq and Zakoïan (2006)], moment-based closed-form 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 techniques 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 the computational viewpoint, inflexible across models, not easy to implement in practice, 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 that solely based on the numerical optimization procedures such as QML or GMM require choosing initial parameters. Broto and Ruiz (2004) point 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. The 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 analytically tractable, computationally simple, and easy to implement. However, it tends to be less precise than some other estimators. In Ahsan and Dufour (2015), we propose a simple estimator for SV(1) model. We exploit the feature of an SV model that the log of squared returns can be written as a non-Gaussian ARMA process and derive a simple estimator for the SV(1) model. This estimator not only has the advantages of Dufour and Valéry (2006) but it also solves the efficiency problem. Not that the estimation of SV(1) model is a challenging task and the estimation of higher order stochastic volatility [SV(p)] models are even more challenging. In an SV(p) model the latent volatility process follows an autoregressive process of order p. In this paper, we extend methods of Dufour and Valéry (2006, 2009) and Ahsan and Dufour (2015) and propose simple estimators for SV(p) models. Further, we also suggest GMM-type estimators for SV(p) models.

Financial time series exhibit excess volatility, heavy tails, the absence of autocorrelations in returns and volatility clustering. As noted by Mandelbrot (1963), volatility clustering implies large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. Even though the SV model generates heavier tails than the ARCH-type models [Kim et al. (1998)], sometimes it fails to capture the extreme movements that we observed in the financial return data. This type of extreme movements could be captured by SV(p) models. In financial econometrics literature, SV(p) models are rarely estimated, the exceptions are Gallant, Hsieh and Tauchen (1997), Asai (2008) and Grant and Chen (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) the multifactor stochastic volatility (MFSV) can be interpreted as a linear combination of latent and independent AR(1) processes that aggregate to an ARMA(p,q) process. Since the MFSV follows the ARMA model, the higher-order autoregressive terms in SV models may be reasonable. *Third*, the empirical results of these studies suggest that SV(p) models can describe the extreme values to a certain extent and provide better fit compared to the first order SV model.

Our simple estimators are analytically tractable, computationally simple, and easy to implement in practice. The estimators can readily be implemented without using numerical optimization methods and these does not require choosing arbitrary initial values for the parameters or auxiliary model. Moreover, the fact that many SV models are parametric models involving only a finite number of unknown parameters, our computationally inexpensive estimators can be easily exploited within a 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 originally proposed by Dwass (1957) and Barnard (1963). 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 our proposed estimators under standard regularity assumptions, showing consistency and asymptotic normality when the fourth moment exists. 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 ARMA based estimator has good statistical properties in terms of bias and root mean square error. outperforms the popular Bayesian estimator by a big margin and for larger sample, it uniformly outperformed all other estimators in terms of bias and RMSE. This conclusion is consistent for all the simulation designs and for all individual parameters. Furthermore, the simple estimators are highly time efficient compared to other estimators.

Finally, two empirical applications relating to SV(p) models and simple ARMA based estimator is presented. First, SV(p) models were fitted to a long sample of S & P 500 index returns and we found that these returns can be better modelled as an SV(3) model. We also implemented LMC tests to construct more reliable inference and found evidence in favour of an SV(3) model. Second, we conducted out-of-sample forecasting experiments to study the accuracy of volatility forecasts of S & P 500 index returns based on SV(p) models and Heterogenous Autoregressive model of Realized Volatility (HAR-RV) type models, for forecast horizons 1-day, 5 day and 20-day. The results suggested that higher-order SV models are performed better than HAR-RV type models in the context of forecasting. This result is consistent whether high volatility periods (such as Financial Crisis) are included in the in-sample or in the out-of-sample.

This paper is organized as follows. Section 2 specifies models and assumptions and Section 3 discusses and 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 7 discusses the MC test technique. Section 8 develops asymptotic theory 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 following Taylor (1986), Ghysels et al. (1996). $\mathbb{N}_0$ refers to the non-negative integers.

Assumption 2.1 STOCHASTIC VOLATILITY OF ORDER p. *The process $\{y_t : t \in \mathbb{N}_0\}$ follows an SV model of the type:*

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

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

where y_t is a $T \times 1$ random vector independent of the variables $\{z_\tau, v_\tau, w_\tau : \tau \leq t\}$, the vectors $(z_t, v_t)'$ are i.i.d. according to a $N[0, I_2]$ distribution and $\{\phi_j\}_{j=1}^{p}$, σ_y, σ_v are the fixed parameters of the model. σ_y is a scale parameter, which removes the need for a constant term in the stationary first-order autoregressive process that captures the dynamics of latent volatility process.

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

The above SV model consists of two stochastic processes describing the dynamics of the returns, y_t and the latent volatilities, w_t. The latent process w_t in (2.2) can be interpreted as the random and uneven flow of new information in financial markets, while ϕ_j's are capture the persistence in the volatility. This type of SV models fit more naturally into the theoretical framework within which much of modern finance theory has been developed. Now transforming y_t by taking logarithms of the squares, we can write the measurement equation of the model as

$$\log(y_t^2) = \log(\sigma_y^2) + w_t + \log(z_t^2). \tag{2.3}$$

This transformation entails no information loss since the distribution of z_t is symmetric (see Remark 1 of Francq and Zakoïan (2006)). Under the standard normality assumption, the transformed error, $\log(z_t^2)$ follows a logarithmic chi-squared distribution with 1 degrees of freedom.

Now (2.3) can be re-written as follows

$$\log(y_t^2) = \mu + w_t + \varepsilon_t, \tag{2.4}$$

where $\mu = [\log(\sigma_y^2) + E(\log(z_t^2))]$ and $\varepsilon_t = \log(z_t^2) - E(\log(z_t^2))$ with $E(\log(z_t^2)) = \psi(\frac{1}{2}) - \log(\frac{1}{2})$. Since z_t is a Gaussian noise, the transformed noise, ε_t, has a $\log \chi^2_{(1)}$ distribution with $\sigma_\varepsilon^2 = Var(\log(z_t^2)) = \psi'(\frac{1}{2}) = \pi^2/2$ and $E(\varepsilon_t^4) = \pi^4$ [see Abramowitz and Stegun (1970)].

Note that 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). Furthermore, Several studies approximated the distribution of $\log(z_t^2)$ by a normal distribution characterized by a mean of -1.27, and a variance of $\pi^2/2$ [see Broto and Ruiz (2004)]. Notice that model expressed by (2.4) can be written as

$$y_t^* = w_t + \varepsilon_t, \tag{2.5}$$

where y_t^* is equal to $(\log(y_t^2) - \mu)$, μ is the mean of $\log(y_t^2)$, given by $\mu = [\log(\sigma_y^2) + E(\log(z_t^2))]$. Combining (2.2) and (2.5), the SV model can be written as a linear state space model of the type:

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + v_t, \qquad \text{(State Transition Equation)} \tag{2.6}$$

$$y_t^* = w_t + \varepsilon_t, \qquad \text{(Measurement Equation)} \tag{2.7}$$

where w_t is a logarithm of latent daily volatility, y_t^* is a logarithm of the daily squared return corrected by its mean. The v_t's and ε_t's are $i.i.d.$ $N(0, \sigma_v)$ and $\log \chi^2_{(1)}$ random variables, respectively. Furthermore, all the roots of the characteristic equation of the volatility process $[\Phi(\lambda) = 0]$ are outside the unit circle, i.e. $1 - \phi_1 \lambda - \phi_2 \lambda^2 - \ldots - \phi_p \lambda^p = 0 \Leftrightarrow |\lambda_i| > 1$ for $i = 1, \ldots, p$ ensures stationarity of the volatility process.

Note that several methods have been proposed in the literature that exploits the state space form of SV model by taking logarithms of squares as in model (2.6) and (2.7); 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 estimators for the model defined by (2.6) and (2.7) based on ARMA representation. Furthermore, we 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 in the literature that the volatility process is driven by at least two-factor: one factor capture 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, Ronald Gallant, Ghysels and Tauchen (2003), Durham (2006, 2007). Many of these study also considered more than two factor and try to fit the volatility process that exhibits by the asset returns. These type of factor models are very useful for capturing non-linearities in financial returns and improves the fit of data dramatically. However, these type model need highly complex numerical optimization technique and they are not tractable analytically. It is noted that we can always transform the MFSV models to an SV model that has 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 a MFSV(m) model of the type:*

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

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

where $\Theta_m^{MFSV} \equiv (\sigma_y, \{\phi_{if}\}_{i=1}^m, \{\sigma_{iv}\}_{i=1}^m)$ 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 $E[v_{it} z_t] = 0 \; \forall i$.

Lemma 3.1 SV-ARMA REPRESENTATION OF MFSV MODELS. *The model MFSV(m) defined by assumption 3.1 can be transformed to an SV-ARMA(m,m-1) process with the following representation: The process $\{y_t : t \in \mathbb{N}_0\}$ follows a model of the type :*

$$y_t = exp(\frac{w_t}{2})\sigma_y z_t, \tag{3.8}$$

$$w_t = \sum_{j=1}^{m} \alpha_j w_{t-j} + \sigma_v v_t - \sigma_v \sum_{j=1}^{m-1} \beta_j v_{t-j}, \tag{3.9}$$

where $\beta_{m,m-1}^{SV-ARMA} \equiv (\sigma_y, \{\alpha_j\}_{j=1}^m, \{\beta_j\}_{j=1}^{m-1}, \sigma_v)$ 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.

To understand the lemma **3.1**, we illustrate the following example.

Example 3.1 SV-ARMA(2,1) REPRESENTATION OF MFSV(2) MODEL. Under the assumptions 3.1, we have 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, 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 obtain a new process that is ARMA(2,1).

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 $[-(\phi_{2f}\sigma_{1v}^2 + \phi_{1f}\sigma_{2v}^2)] = -\beta_1\sigma_v^2$.

Although 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) models instead of an SV-ARMA and recuperate the parameters of SV-ARMA from the estimates of SV(p) parameters. This is based on the AR approximation of volatility process. A number of 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 related method 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 has been used by Dufour and Pelletier (2005) and Dufour and Jouini (2014).

Lemma 3.2 SV(K) APPROXIMATION OF SV-ARMA PROCESS. *The model SV-ARMA(p,q) defined by lemma* **3.1** *can be transformed to an SV(k) process with the following representation: The process* $\{y_t : t \in \mathbb{N}_0\}$ *follows a model of the type :*

$$y_t = exp(\frac{w_t}{2})\sigma_y z_t, \tag{3.10}$$

$$w_t = \sum_{j=1}^{k} \phi_j w_{t-j} + \sigma_v v_t, \tag{3.11}$$

where $\Theta_k^{SV} \equiv (\sigma_y, \{\phi_j\}_{j=1}^{k}, \sigma_v)$ *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.*

To understand the lemma **3.2**, 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, using lemma **3.2** we have

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

Solving above equations yields the parameter of SV-ARMA(2,1) 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. MFSV model, where volatility factor structure is driven by two independent AR(1) process, has an SV-ARMA(2,1) representation by aggregation.

2. $SV(\infty)$ representation by using the invertibility of MA part of SV-ARMA.

3. Estimate an SV(3) and recuperate the SV-ARMA(2,1) parameters.

4. From SV-ARMA(2,1) parameters to identify the AR factor polynomials of MFSV.

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

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 is one of the attractive probabilistic features of stochastic volatility models. The problem of finding a necessary and sufficient condition for stationarity of GARCH waited for a solution until Nelson (1990) 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 Zakoian (2011). To establish the large sample property of SV model 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 an SV(p) model.

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 $E(z_t) = 0$ and $E(z_t)^2 = 1$. 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 $[\Phi(\lambda) = 0]$ are outside the unit circle, i.e. $1 - \phi_1 \lambda - \phi_2 \lambda^2 - \ldots - \phi_p \lambda^p = 0 \Leftrightarrow |\lambda_i| > 1$ for $i = 1, \ldots, p$ and there is an integer $s \geq 1$ such that

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

Then

1. *$E[|w_t|]^s < \infty$. 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* $E[|\ln|z_t||^s] < \infty$*, then* $E[|\ln|y_t||^s] < \infty$.

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 model 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* $y_t \equiv exp(\frac{w_t}{2})\sigma_y z_t \Leftrightarrow \ln|y_t| = w_t/2 + \ln|\sigma_y| + \ln|z_t|$. *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 two simple estimators for SV(p) models and recursive algorithms to obtain these estimators. The simple moment based estimator is the extension of Dufour and Valéry (2006, 2009) and the ARMA based estimator is the extension of Ahsan and Dufour (2015).

5.1. Simple moment based estimator

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).

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

The moments and cross-moments of the process $y_t \equiv y_t(\theta)$ where $\theta \equiv (\phi_1, ..., \phi_p, \sigma_y, \sigma_v)'$ are given in the following lemma.

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* $E(U^{2p+1}) = 0, \ \forall p \in \mathbb{N}$ *and* $E(U^{2p}) = \frac{2p!}{2^p p!}, \ \forall p \in \mathbb{N}$*; 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:*

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

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

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

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

Lemma 5.2 CLOSED FORM MOMENT EQUATIONS SOLUTION FOR SV(2). *Using Lemma* **5.1** *and Yule-Walker equations of the volatility process, we have following moment equations solution:*

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

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

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

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

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

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

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

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 $\gamma_j \equiv cov(w_t, w_{t+j})$.

Now it is then 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 higher-order SV process. This estimator is computationally very simple than those that based on sophisticated numerical optimization technique.

5.2. Simple ARMA based estimator

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

Lemma 5.4 ARMA REPRESENTATION OF STOCHASTIC VOLATILITY MODEL. *The model defined by the assumptions 2.1 - 2.2 has following ARMA(p,p) representation:*

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

where $\eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} = v_t + \varepsilon_t - \sum_{j=1}^{p} \phi_j \varepsilon_{t-j}$ *admits an MA(p) process,* $y_t^* = \log(y_t^2)$*, and the error processes* v_t*'s and* ε_t*'s are i.i.d.* $N(0,\sigma_v)$ *and* $\log\chi^2_{(1)}$ *random variables, respectively.*

Lemma 5.5 AUTOCOVARIANCE FUNCTION OF THE OBSERVED PROCESS. *Now under the lemma* **5.4***, the observed process* $\{y_t^*\}$ *has an ARMA(p,p) representation and its autocovariances satisfies the following equations respectively:*

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

The analytical closed-form expressions for the model parameters, $\theta \equiv (\phi_1, ..., \phi_p, \sigma_y, \sigma_v)^{'}$, are given in the following lemma.

Lemma 5.6 CLOSED-FORM PARAMETER SOLUTIONS FOR SV OF ORDER p. *Under the assumptions 2.1 - 2.2, the analytical closed-form expression for the model parameters, $\theta \equiv (\phi_1, ..., \phi_p, \sigma_y, \sigma_v)^{'}$, are given below:*

$$\phi_p = \Gamma_{(k,p)}^{-1} \gamma_{(k,p)}, \tag{5.19}$$

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

$$\sigma_v = [\gamma_{y^*}(0) - \hat{\phi}_p^{'} \hat{\gamma}_{(k,p)} - \psi^{'}(1/2)]^{1/2}, \tag{5.21}$$

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

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

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

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

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

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

This yields the following simple estimators of the SV(p) coefficients:

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

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

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

where $\psi^{'}(1/2)$ is equal to $\pi^2/2$.

In practice, this method may lead to $[\hat{\phi}_p^{'} \hat{\rho}_{(k,p)} + \psi^{'}(1/2)] > 1$ where $\hat{\rho}_{(k,p)} = \hat{\gamma}_{(k,p)} / \hat{\gamma}_{y^*}(0)$ and this makes $\hat{\sigma}_v^2 < 0$. To deal with this problem, Kristensen and Linton (2006) propose that the estimator of ϕ_p can be winsorized (censored) at $1^{'}$ or at $(1-\delta)1^{'}$ for small positive δ.

Note that

$$\phi_p = \sum_{j=1}^{\infty} \omega_j \Gamma_{(k+j,p)}^{-1} \gamma_{(k+j,p)} \tag{5.26}$$

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 can obtain the same efficiency as the Whittle estimator of Giraitis and Robinson (2001). An estimator of ϕ_p based on (5.26) and winsorized at $1^{'}$ or at $(1-\delta)1^{'}$ for small positive δ is as follows:

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

where $1 \leq J \leq T - p$ with $\sum_{j=1}^{J} \omega_j = 1$ and T is the length of financial returns.

5.3. Recursive algorithms for SV(p) models

Previously we show that, it is possible to derive higher order closed form solution for SV models. Here we are using 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. It is actually 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 the convenience, we use a different indexation for the autoregressive parameters of the volatility process in this section. For example, the parameters of an SV(p) model are now $\Theta_p^{SV} \equiv (\{\phi_{p,j}\}_{j=1}^{p}, \sigma_{pv}, \sigma_y)$. Using the following recursive formulae, we can obtain higher order solution of a $SV(p)$ model from a $SV(p-1)$ model.

$$\phi_{p,p} = \frac{\rho_p - \sum_{j=1}^{p-1} \phi_{p-1,j}\, \rho_{p-j}}{1 - \sum_{j=1}^{p-1} a_{p-1,j}\, \rho_j}, \tag{5.28}$$

$$\phi_{p,j} = \phi_{p-1,j} - \phi_{p,p}\phi_{p-1,p-j} \quad \forall j = 1,2,3,\dots,p-1. \tag{5.29}$$

Lemma 5.7 RECURSIVE MOMENT EQUATION SOLUTION. *Under the assumptions 2.1-2.2, the volatility process satisfy the Yule-Walker equations. Thus we can apply Durbin-Levinson algorithm to get the closed form solution recursively. We want to find the closed form estimator for* $\Theta_p^{SV} \equiv (\{\phi_{p,j}\}_{j=1}^{p}, \sigma_{pv}, \sigma_y)$ *and following algorithm is for our recursive solution.*

$$\hat{\sigma}_y = \frac{3^{1/4}\hat{\mu}_2}{(\hat{\mu}_4)^{1/4}}, \tag{5.30}$$

$$\hat{\phi}_{p,p} = \frac{\hat{\rho}_p - \sum_{j=1}^{p-1} \hat{\phi}_{p-1,j}\hat{\rho}_{p-j}}{1 - \sum_{j=1}^{p-1} \hat{\phi}_{p-1,j}\hat{\rho}_j}, \tag{5.31}$$

$$\hat{\phi}_{p,j} = \hat{\phi}_{p-1,j} - \hat{\phi}_{p,p}\hat{\phi}_{p-1,p-j} \quad \forall j = 1,2,3,\dots,p-1, \tag{5.32}$$

$$\hat{\sigma}_v = \left[(1 - \sum_{j=1}^{k} \hat{\phi}_{p-1,j}\hat{\rho}_j) \left[log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}) \right] \right]^{1/2}, \tag{5.33}$$

where

$$\hat{\rho}_j = log(\frac{\hat{\mu}_2(j)}{(\hat{\mu}_2)^2}) / log(\frac{\hat{\mu}_4}{3(\hat{\mu}_2)^2}), \;\; for \;\; j = 1,2,\dots \tag{5.34}$$

The sample auto-correlation of volatility process at lag j is denoted by $\hat{\rho}_j$ *and we can calculate this by using* (5.34)*. After that using these sample auto-correlations, we can estimate the parameters of the volatility process in the second stage with the help of a Durbin-Levinson algorithm. In the proof, we show that how one can obtain moment equations solution of SV(2) recursively from the moment equations solution of SV(1).*

The ARMA based estimator is based on the extended Yule-Walker (EYW) equations. When the MA order is fixed, the system of the EYW equations constitutes a nested Toeplitz system. A *Generalized Durbin-Levinson* algorithm for 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

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

For $i \geq 1$, calculate

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

and

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

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

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

where $\psi^{'}(1/2)$ is equal to $\pi^2/2$.

This algorithm is the same as Tsay and Tiao's algorithm (1984) for calculating the extended sample autocorrelation function under the stationarity assumption.

6. GMM estimation of SV(p) model

In this section, we propose GMM estimator for SV(p) models. 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. In literature, Andersen and Sørensen (1996) proposed GMM estimator for SV(1) but the GMM estimator for SV(p) models remains to be discussed. 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 $Y_T = (y_1, \ldots, y_T)^{'}$, and $\hat{\Omega}_T$ a positive-definite (possibly random) matrix. We compute the corresponding sample averages such that

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

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

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

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 on replacing θ_0 and Ω_* by consistent estimators. The sample analogue 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[1] this infinite sum and using the sample autocovariances, 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. However for Ω_*, we need to use the heteroscedasticity 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 kernal function that can fit into the above equation [2]. 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 using GMM to estimate the parameters of $SV(1)$. One issue of GMM is how many moment conditions to use. You may think that by increasing the number of moment conditions, you are using additional information, which, if weighted appropriately, cannot make the parameter estimates worse. But the weighting matrix, Ω, must itself be estimated, and with q moment conditions, we need to

[1]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)

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

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 in general with GMM is that there is not much guidance on which moment conditions to use. For SV models, moment conditions can be based on infinitely many functions of returns; see Melino and Turnbull (1990).

7. Asymptotic distributional theory

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

7.1. Asymptotic theory for moment based estimator

Now we want to focus on the asymptotic distribution of the moment estimator of SV(p) models. Dufour and Valéry (2006) derived an asymptotic distribution for the SV of order one model where they viewed the SV(1) as a special case of the general class of estimators. Here we want to generalize their proposed framework in the context of SV(p). Our approach for constructing a 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 an SV(1), but without checking the suitable regularity conditions for the SV model. 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 $Y_T \equiv (y_1, \ldots, y_T)^{'}$ and

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

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 neighbourhood 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.

Proposition 7.1 ASYMPTOTIC DISTRIBUTION OF METHOD-OF-MOMENTS ESTIMATOR. *Under the assumptions 7.1-7.3,*

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

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*

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

Here, $V_*(\theta_0)$ 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 on replacing θ_0 and Ω_* by consistent estimators. The sample analogue of Ω_* is given by

$$\hat{\Omega}_* = \hat{\Gamma}_0 + \sum_{i=1}^{\infty} (\hat{\Gamma}_i + \hat{\Gamma}_i^{'}) \tag{7.5}$$

Given this structure, it is natural to estimate $\hat{\Omega}_*$ by truncating[3] this infinite sum and using the sample auto-covariances, where

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

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

$$\hat{\Omega}_{*,HAC} = \hat{\Gamma}_0 + \sum_{i=1}^{T-1} \omega_{i,T} (\hat{\Gamma}_i + \hat{\Gamma}_i^{'}) \tag{7.6}$$

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 kernal function that can fit into the above equation [4]. Thus a consistent estimator of $\hat{V}_*(\theta_0)$ is given by

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

Since we are using 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

[3]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)

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

only need to show that the Assumption 7.1 is hold.

Proposition 7.2 ASYMPTOTIC DISTRIBUTION FOR EMPIRICAL MOMENTS. *Under the assumptions 7.1-7.3 with $p > 1$; we have:*

$$\sqrt{T}[\bar{g}_T(Y_T) - \mu(\theta_0)] \xrightarrow{D} N[0, \Omega_*] \tag{7.7}$$

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

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

7.2. Asymptotic theory for ARMA based estimator

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

A.1 : 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, with $E(z_t) = 0$ and $E(z_t)^2 = 1$. The probability distribution of z_t has a continuous density with respect to Lebesgue measure on real line, and its density is positive on $(-\infty, +\infty)$.

A.2 : The latent process $\{w_t\}$ is strictly stationary with $E[|w_t|]^s < \infty$ that follows from $|\phi| < 1$ and there is an integer $s \geq 1$ such that

$$E\,|v_t|^s < \infty, \qquad |\phi|^s < 1. \tag{7.1}$$

Under the assumptions A.1 and A.2 with $s = 2$, the observed process $\{y_t\}$ is strictly stationarity and geometrically ergodic with exponential β-mixing (see result 4.1 and 4.2). Furthermore, this property is preserved by any continuous transformation of $\{y_t\}$, i.e., $\{g(y_t)\}$ where g is any measurable function $g(\cdot)$. In the following Lemma using Ergodic theorem we prove the consistency of the empirical moments that defined in (5.22).

Lemma 7.1 CONSISTENCY OF EMPIRICAL MOMENTS. *Under the assumptions A.1 and A.2 with $s = 2$, the estimators $\hat{\Gamma}(m) = (\hat{\gamma}_{y^*}(k))_{k=0,\ldots,m}$ and $\hat{\mu}$ in (5.22) satisfy:*

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

The assumptions A.1 and A.2 with $s = 4$ is necessary and sufficient for the SV model to have a strictly stationary solution with a fourth moment. This solution will be β-mixing with geometrically decreasing mixing coefficients. In the following Lemma using we prove the asymptotic distribution of the empirical moments that defined in (5.22).

Lemma 7.2 ASYMPTOTIC DISTRIBUTION OF EMPIRICAL MOMENTS. *Under the assumptions A.1, A.2 with $s = 4$, and lemma **7.1**, the estimators $\hat{\Gamma}(m) = (\hat{\gamma}_{y^*}(k))_{k=0,\ldots,m}$ and $\hat{\mu}$ in (5.22) satisfy:*

$$\sqrt{T}\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\Gamma}(m) - \Gamma(m) \end{bmatrix} \xrightarrow{d} N\left(0, \begin{bmatrix} V_\mu & C_{\mu,\Gamma(m)}^{'} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} \end{bmatrix}\right), \tag{7.2}$$

where

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

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

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

where Λ_t *is an* $(m+1)\times 1$ *vector with* $\Lambda_{t,k} = y_t^* y_{t+k}^* = (\log(y_t^2)-\mu)(\log(y_{t+k}^2)-\mu)$, $k = 0,\ldots,m$, *and* $C_{\mu,\Gamma(m)}$ *is a* $(m+1)\times 1$ *vector given by* $C_{\mu,\Gamma(m)} = (\bar{c},\ \ 0_{[1\times m]})'$, *with* $\bar{c} = 2\sum_{\tau=1}^{\infty} E(\varepsilon_\tau^3)$, *where* $E(\varepsilon_\tau^3) = -14\mathscr{Z}(3)$ *and* $\mathscr{Z}(\cdot)$ *is Riemann's Zeta function with* $\mathscr{Z}(3) = 1.20205$.

Theorem 7.3 ASYMPTOTIC DISTRIBUTION OF SIMPLE ESTIMATOR. *Under the assumptions A.1, A.2 with* $s = 4$, *lemma* **7.1**, *and lemma* **7.2**, *the estimator* $\hat{\theta} \equiv (\hat{\phi}_1,\ldots,\hat{\phi}_p,\hat{\sigma}_y,\hat{\sigma}_v)'$ *given in* (5.23)-(5.25) *is consistent, i.e.,* $\hat{\theta} \stackrel{p}{\longrightarrow} \theta$ *and*

$$\sqrt{T}(\hat{\theta}-\theta) \stackrel{d}{\to} N(0,V), \tag{7.3}$$

where

$$V = \frac{\partial D_p(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\ldots,\gamma_{y^*}(2p))}{\partial(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\ldots,\gamma_{y^*}(2p))} \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(2p)} \\ C_{\mu,\Gamma(2p)} & V_{\Gamma(2p)} \end{bmatrix} \left(\frac{\partial D_p(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\ldots,\gamma_{y^*}(2p))}{\partial(\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\ldots,\gamma_{y^*}(2p))}\right)', \tag{7.4}$$

where the function $D_p = (D_{\phi_p}, D_{\sigma_y}, D_{\sigma_v})'$ *is given by*

$$D_{\phi_p} = {\Gamma_{(k,p)}}^{-1}\gamma_{(k,p)}, \quad D_{\sigma_y} = exp(\mu+1.27)^{1/2}, \quad D_{\sigma_v} = [\gamma_{y^*}(0) - \phi_p'\gamma_{(k,p)} - \psi'(1/2)]^{1/2}, \tag{7.5}$$

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

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

with $k > p$, $\gamma_{y^*}(k) = cov(y_t^*, y_{t-k}^*)$, $y_t^* = (\log y_t^2 - \mu)$, *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 case of SV(1) model, the function* $D_1 = (D_{\phi_1}, D_{\sigma_y}, D_{\sigma_v})'$ *is given by*

$$D_{\phi_1} = \gamma_{y^*}(2)/\gamma_{y^*}(1), \quad D_{\sigma_y} = exp(\mu+1.27)^{1/2}, \quad D_{\sigma_v} = \kappa_1\kappa_2, \tag{7.6}$$

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 $\sigma_y = \sqrt{exp(\mu + 1.27)}$*,* $\kappa_1 = [1 - (\gamma_{y^*}(2)/\gamma_{y^*}(1))^2]$*,* $\kappa_2 = [\gamma_{y^*}(0) - \psi^{'}(1/2)]$*, and* $\psi^{'}(1/2) = \pi^2/2$*.*

In case of SV(2) model,, the function $D_2 = (D_{\phi_1}, D_{\phi_2}, D_{\sigma_y}, D_{\sigma_v})^{'}$ *is given by*

$$
\begin{aligned}
D_{\phi_1} &= \frac{\gamma_{y^*}(2)\gamma_{y^*}(3) - \gamma_{y^*}(1)\gamma_{y^*}(4)}{\gamma^2_{y^*}(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}, \qquad & D_{\phi_2} &= \frac{\gamma_{y^*}(2)\gamma_{y^*}(4) - \gamma^2_{y^*}(3)}{\gamma^2_{y^*}(2) - \gamma_{y^*}(1)\gamma_{y^*}(3)}, \\
D_{\sigma_y} &= exp(\mu + 1.27)^{1/2}, & D_{\sigma_v} &= [\gamma_{y^*}(0) - \psi^{'}(1/2) - \phi_1\gamma_{y^*}(1) - \phi_2\gamma_{y^*}(2)]^{1/2},
\end{aligned} \tag{7.7}
$$

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.,*

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

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

An estimator of the covariance matrix V can be obtained by first estimating V_μ and $V_\Gamma(3)$ using heteroskedasticity and autocorrelation consistent variance estimators and then substituting $\hat{\mu}$, $\hat{\gamma}_{y^*}(0)$, $\hat{\gamma}_{y^*}(1)$, $\hat{\gamma}_{y^*}(2)$ into $\partial D(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2))/\partial(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \gamma_{y^*}(2))$. One can alternatively use the analytic expressions of $\gamma_{y^*}(k)$ to obtain an estimator of V_μ.

8. Monte Carlo tests

We now examine the the usefulness of our simple estimator in the context of simulation based inference, i.e., Monte Carlo test technique. The technique of Monte Carlo tests was originally been 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 $\theta = (\phi, \sigma_y, \sigma_v)$, by its sample analogue 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 $\hat{p}_N[S] \equiv p_N[S_0; S(N)]$, where

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

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

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

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

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 $\hat{p}_N[S|\theta] \equiv p_N[S_0;S(N,\theta)]$, where

$$p_N[x;S(N,\theta)] \equiv \frac{NG_N[x;S(N,\theta)]+1}{N+1}. \tag{8.4}$$

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

$$\sup_{\theta \in \Omega_0} \hat{p}_N[S|\theta] \le \alpha. \tag{8.5}$$

If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:

$$P[\sup_{\theta \in \Omega_0} \{\hat{p}_N[S|\theta]\} \le \alpha] \le \alpha, \tag{8.6}$$

The test defined by $\hat{p}_N[S|\theta] \le \alpha$ 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 $\hat{p}_{TN}[S_T|\theta] \equiv p_{TN}[S_{T0};S_T(N,\theta)]$, where

$$p_{TN}[x;S_T(N,\theta)] \equiv \frac{NG_{TN}[x;S_T(N,\theta)]+1}{N+1}. \tag{8.8}$$

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

$$C_T = \{\theta \in : \left\|\hat{\theta}_T - \theta\right\| < d\} \tag{8.11}$$

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 conduct simulation to investigate the statistical performance of our proposed estimators in terms of bias and root mean square error (RMSE). 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 determine 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 consider sample size of T = 500 and 2000. The choice of samples are adequate in the sense that the sample size of T = 1200 observations are corresponding with roughly five years of daily returns, whereas for high frequency financial data, it is corresponding with roughly fifteen days of five minute intraday returns.

Globally, there is no uniform ranking between the different estimators but the performance of Bayesian-MCMC methods remain superior among the competing methods. Under the above simulation designs, we compare our simple estimators and GMM estimators to popular Bayesian-MCMC estimator. We use the MatLab code of Chan and Grant (2016) for the Bayesian estimation as well as their specified prior.

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

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

where $\mu_1(\theta) \equiv E(|y_t|) = \sigma_y(2/\pi)^{1/2}exp[\gamma_0/8]$, $\mu_2(\theta) \equiv E(y_t^2) = \sigma_y^2 exp[\gamma_0/2]$, $\mu_3(\theta) \equiv E(|y_t|^3) = 2\sigma_y^3(2/\pi)^{1/2}exp[9\gamma_0/8]$, $\mu_4(\theta) \equiv E(y_t^4) = 3\sigma_y^4 exp[2\gamma_0]$, $\mu_{1,1}(j|\theta) \equiv E(|y_t||y_{t-j}|) = \sigma_y^2(2/\pi)exp\left[\gamma_0(1+\rho_j)/4\right]$, $j = 1,...,10$ $\mu_{2,2}(j|\theta) \equiv E(y_t^2 y_{t-j}^2) = \sigma_y^4 exp\left[\gamma_0(1+\rho_j)\right]$, $j = 1,...,10$, 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 asymptotic covariance matrix and HAC matrix with Bartlett Kernel).

Table 1, that reports the estimation results for model $M1$, shows that our ARMA based estimator denoted by AD-ARMA gives superior performance in terms of bias and RMSE whereas GMM, EDV (extension of Dufour and Valéry (2006)) and MCMC estimators are performed poorly. For each parameters, the smallest bias and RMSE is associated with AD-ARMA estimation except for ϕ_2 where the bias of EDV estimation is slightly lower compare to AD-ARMA estimation. The second smallest RMSE for ϕ_1, ϕ_2, σ_y are associated with 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 is very large. Finally, for the larger samples, T=2000, we have almost same identical results for all the estimates as in the case of T = 500. Now AD-ARMA outperforms all other estimators in terms of bias and RMSE and RMSE of AD-ARMA estimates decreases as the sample size increases, proofs the consistency of this simple estimator.

The results of $M2$, $M3$ and $M4$ models are reported in table 3-4 and results are very similar to the ones reported in table 1. However, there are some important points to be noted down. In $M3$, we find that some of EDV and MCMC estimates are explode with very high 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 rank poorly compare to other estimators. However, in some cases (particularly for σ_v) 6-moments GMM estimator performed better compare to 24-moments GMM estimator. This casts 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 for example, Buse (1992), Chao and Swanson (2000), and Dufour and Valéry (2006). In particular, overidentification increases bias 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).

These results are conditional on the convergence of GMM estimator since we observe that GMM posses numerical instability when the sample size is small. We also encounter the non-convergence problem with Bayesian estimation. Furthermore, the EDV estimates of σ_v sometimes produce a negative value. In these cases, we discarded those simulations from the calculation. However, it is noted that in all of the above cases our simple ARMA based estimator denoted by AD-ARMA yield a solution. Several conclusions may be drawn from these simulation results for the AD-ARMA estimator. First, when T=500, our ARMA based estimator provides accurate estimates since it outperforms all other estimators in terms of bias except ϕ_2 in design $M1$. Second, the AD-ARMA is more efficient than other estimators in terms of RMSE for simulation design $M1$, $M3$ and $M4$. In case of $M2$, the efficient GMM estimator with 24 moment (GMM-24M-E) is more efficient for the autoregressive parameters (ϕ_1, ϕ_2). However, in this design, AD-ARMA outperform other estimators in terms of efficiency for (σ_y, σ_v). Third, when T = 2000, AD-ARMA uniformly outperformed all other estimators in terms of bias and RMSE and these include the popular Bayesian estimator. This is consistent for all the simulation designs and for all individual parameters. Furthermore, from table 5, the simple estimators are highly time efficient and the margin of time efficiency is huge compared to other estimators except for EDV and AD-ARMA.

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 compare the volatility forecast performance between SV(p) models with the realized volatility based models that incorporates additional information from the high-frequency data. We used ARMA based estimator in our empirical analysis since in simulations it outperforms other estimators in terms of bias and RMSE.

10.1. Empirical evidence

The SV(p) models are 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 then centred around the sample mean by $y_t = r_t - \mu_r$. 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 Great Depression (1929), second world war (1937-45), OPEC oil price shock(1973), Black Monday (1987), Asian financial crisis (1997), early 2000s recession (Dot-com bubble), late-2000s Financial Crisis (Subprime mortgage crisis / United States housing bubble) and recent Russian financial crisis (2014).

Table 6 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 7 shows the parameter estimates of the SV(p) models obtained by our simple ARMA based estimator where the standard errors are computed using multivariate delta method (see theorem **7.3**). Our result shows that there are 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 $p = 3$ are statistically significant and this gives us the empirical evidence that the volatility process can be treated as an autoregressive process of order more than one. Asai (2008) pointed out that introducing an additional lagged term in the volatility process may capture extreme values to a certain extent. We also estimated an SV(4) model and the coefficient of ϕ_4 is turned out to be insignificant. Again from table 7, we can see that the ARMA based estimator is highly time efficient.

However, the p-values tabulated in table 7 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). However, the asymptotic standard error can be markedly different and may be quite unreliable in finite samples. To construct more reliable inference, we propose parametric Bootstrap or Local Monte Carlo tests since our estimator is convenient for use in the context of computationally costly inference techniques.

To construct more reliable inference, we implemented LMC tests as discussed in section 8 and results are reported in table 8. From the table, we can see that the estimates of ϕ_4 of SV(4) model is not statistically significant in both asymptotic and LMC tests and this entails that an SV(3) model could be more suitable for the volatility dynamics of this sample periods. To summarize, the results presented here indicate that an SV models with additional lag term in the volatility process may be appropriate to model the S&P 500 index.

10.2. Volatility forecast performance

We evaluate the performance of forecasting volatility between SV(p) models and realized volatility based models. 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 of true latent volatility (for details, see appendix A).

However, there are problems in measuring daily realized volatility measures from high frequency return data. First, the *discretization error* in the estimates of volatility due to 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. Thus incorporating noisy information may lead to the parameter inference problem problematic. 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 et al. (2010)], Realized Kernel (RK) [Barndorff-Nielsen, Hansen, Lunde and Shephard (2008, 2011)]. We also consider subsampled versions of all the estimators of RV (except RK, since it using tick-by-tick data, it cannot be subsampled). Subsampling[5], introduced by Zhang et al. (2005), is a simple way to improve efficiency of some sparse-sampled estimators.

In particular we want to conduct some out-of-sample forecasting experiments to study the accuracy of volatility forecasts based on SV(p) models and HAR-RV type models proposed by Corsi (2009). HAR-RV stands for Heterogenous Autoregressive model of Realized Volatility 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:

$$\log \sigma^2_{d,t+1} = c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)}) + u_{d,t+1} \tag{10.1}$$

where

$$\log RV_t^{(w)} = \frac{1}{5}\sum_{j=0}^{4} \log RV_{t-j}^{(d)}, \quad \log RV_t^{(m)} = \frac{1}{22}\sum_{j=0}^{21} \log RV_{t-j}^{(d)},$$

and $\log \sigma^2_{d,t+1}$ is equal to $\log y^2_{t+1}$.

Two measures are used to evaluate the forecast accuracy, namely, the mean square error (MSE) and the mean absolute error (MAE). The MSE and MAE are two of the most popular measures to test the forecasting power of a model for their mathematical simplicity. They are defined by

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \sigma_t^2)^2 \quad \text{and} \quad MAE = \frac{1}{T}\sum_{t=1}^{T}|\hat{\sigma}_t^2 - \sigma_t^2|, \tag{10.2}$$

where σ_t^2 is the square residual return (y_t^2) and $\hat{\sigma}_t^2$ is the forecast of square residual return ($\hat{y}_t^2$).

The sample period of this application is from January 3, 2000 to October 21, 2016, and the number of observations is T = 4200. The high-frequency RV estimates of S&P 500 index are obtained from the OxfordMan Institutes Realized Library and low-frequency daily S&P 500 index is obtained from WRDS. We conduct two out-of-sample forecast experiments. In experiment 1, the in-sample is from January 3, 2000 to December 31, 2009 and the out-of-sample is from January 3, 2010 to October 21, 2016. In experiment 2, the in-sample is from January 3, 2000 to December 31, 2007 and the out-of-sample is from January 3, 2010

[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.

to October 21, 2016. Note that the late-2000s Financial Crisis included in the in-sample for experiment 1 and it included in the out-of-sample for experiment 2.

Table 9-10 reports summary statistics of logarithmic transformation of RV-type estimators. For each forecast experiments, we obtain forecast from three SV models and nine HAR-RV type models and calculated forecast evaluation measures i.e., MSE and MAE. Out-of-sample forecasts are computed using Rolling (moving) window method and computed for forecast horizon 1-day, 5-day and 20-day. Table 11-12 presents the main results of our forecasting experiments.

From table 11, SV models performed better than the HAR-RV type models in experiment 1. It is noted that the MSE statistic indicates that the SV(2) model provides the most accurate forecasts among competing models and this is consistent for forecast horizon 1-day, 5-day and 20-day. According to the MAE statistic, SV(2) also ranks first for forecast horizon 1-day and 5-day. However, the MAE statistic favours the SV(1) model while the SV(2) model is now second best when the forecast horizon is 20-day. The HAR-RV type models did not perform very well in terms of MSE and MAE. These models are ranked after SV models and HAR-RK performed worse when forecast horizon is 1-day, whereas HAR-RSV (5-minute) performed worse for forecast horizon 5-day and 20-day. It is also noted that the subsampled version of the HAR-RSV5 model performed quite well.

From table 12, again SV models performed better than the HAR-RV type models in experiment 2. However, the values of both MSE and MAE are much higher. This could be because of the Financial Crisis is now included in out-of sample. According to MSE statistics, the best model is SV(2) when forecast horizons are 1-day and 20-day while the SV(3) model ranks first for forecast horizon 5-day. However, according to MAE statistics, the best model is SV(3) when forecast horizons are 1-day and 5-day while the SV(1) model ranks first for forecast horizon 20-day. Among the HAR-RV type models, the HAR-RV5 model is the best and it ranked four, according to the values of MSE and MAE and remained in that position across the forecast horizon. This is surprising since RV5 is the most noisy estimator. However, this could be due to the fact that in this experiment the Financial Crisis is included in out-of-sample. The crisis may make the financial market unstable and the noisy RV5 may capture some of the unpredictable features of the market.

In both of our out-of sample experiments, higher-order SV models are performed better than HAR-RV type models indicating that one should consider SV(p) models for forecasting volatility. This result is consistent whether high volatility periods (such as Financial crisis) are included in the in-sample or out-of-sample.

11. Conclusion

In this paper, we proposed several estimators for higher-order SV models and these include two computationally simple estimators and GMM-type estimators. The motivation as well as stationarity, ergodicity and mixing properties of SV(p) models are discussed. Furthermore, the study proposed recursive algorithms for computing simple estimators and derived their asymptotic distribution. The simple estimators, proposed by this study, are especially convenient for use in the context of simulation based inference techniques i.e., Bootstrap or Monte Carlo tests.

In simulation, we compared our proposed estimators to the popular Bayesian-MCMC estimator. We found that GMM estimators and Bayesian estimator have non-convergence problem while the EDV estimates of σ_v^2 sometimes produce a negative value for the variance of the volatility innovation. However, the ARMA based estimator, proposed by this study, mostly outperforms all other estimators and these include the popular Bayesian estimator and for larger sample, it uniformly outperformed all other estimators in terms of bias and RMSE. This conclusion is consistent for all the simulation designs and for all individual parameters. Furthermore, the simple estimators are highly time efficient compared to other estimators.

Finally, two empirical applications relating to SV(p) models and simple ARMA based estimator is presented. First, SV(p) models were fitted to a long sample of S & P 500 index returns and we found that these returns can be better modelled as an SV(3) model. We also implemented LMC tests to construct more reliable inference and found evidence in favour of an SV(3) model. Second, we conducted out-of-sample forecasting experiments to study the accuracy of volatility forecasts of S & P 500 index returns based on SV(p) models and HAR-RV type models, for forecast horizons 1-day, 5 day and 20-day. The results suggested that higher-order SV models are performed better than HAR-RV type models in the context of forecasting. This result is consistent whether high volatility periods (such as Financial Crisis) are included in the in-sample or in the out-of-sample.

Appendix

A. 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 et al. (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.

B. Proofs

PROOF OF LEMMA **3.1**

Under the assumptions 3.1, we have 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 is 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 OrnsteinUhlenbeck models.

∎

PROOF OF LEMMA **3.2**

We consider the SV-ARMA(p,q) model defined by lemma **3.1** with the volatility process driven by a $ARMA(p,q)$ such that

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

where

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

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

$$\sum_{j=0}^{\infty}(-\phi_j)w_{t-j} = \sigma_v v_t, \tag{B.4}$$

where

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

see for example Fuller (1996), Ch. 2, page 74. The process defined by the equation B.4 can be estimated by an truncated AR(k) process and recuperate the ARMA(p,q) parameters from there. The identification requires $k \geq p+q$. Thus we can replace the equation B.4 by the following:

$$\sum_{j=0}^{k}(-\phi_j)w_{t-j} = \sigma_v v_t \tag{B.5}$$

Now substitute equation B.5 into equation 3.9 to get the $SV(k)$ defined by the lemma **3.2**.

∎

PROOF OF LEMMA **5.1**

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

$$\begin{aligned}
\mu_k(\theta) \equiv E(y_t^k) &= \sigma_y^k E(z_t^k) E\left[exp(\frac{kw_t}{2})\right] \\
&= \sigma_y^k \frac{k!}{2^{k/2}(k/2)!} exp\left[\frac{k^2}{8} Var(w_t)\right] \\
&= \sigma_y^k \frac{k!}{2^{k/2}(k/2)!} exp\left[\frac{k^2}{8} \frac{\sigma_v^2}{1-\sum_{j=1}^{p}\phi_j \rho_j}\right], \quad if\ k\ is\ even \\
&= 0, \quad if\ k\ is\ odd
\end{aligned} \tag{B.6}$$

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

$$\begin{aligned}
\mu_{k,l}(m|\theta) \equiv E(y_t^k y_{t+m}^l) &= \sigma_y^{k+l} E(z_t^k) E(z_{t+m}^l) E\left[exp(\frac{kw_t}{2} + \frac{lw_{t+m}}{2})\right] \\
&= \sigma_y^{k+l} \frac{k!}{2^{k/2}(k/2)!} \frac{l!}{2^{l/2}(l/2)!} \left[exp(\frac{k^2}{8} Var(w_t) + \frac{l^2}{8} Var(w_{t+m}) + \frac{2kl}{8} cov(w_t, w_{t+m}))\right] \\
&= \sigma_y^{k+l} \frac{k!}{2^{k/2}(k/2)!} \frac{l!}{2^{l/2}(l/2)!} \left[exp(\frac{k^2}{8}\gamma_0 + \frac{l^2}{8}\gamma_0 + \frac{2kl}{8}\gamma_0 \rho_m)\right] \\
&= \sigma_y^{k+l} \frac{k!}{2^{k/2}(k/2)!} \frac{l!}{2^{l/2}(l/2)!} exp\left[\frac{1}{8}\gamma_0 (k^2 + l^2 + 2kl\rho_m)\right] \\
&= \sigma_y^{k+l} \frac{k!}{2^{k/2}(k/2)!} \frac{l!}{2^{l/2}(l/2)!} exp\left[\frac{1}{8} \frac{\sigma_v^2}{1-\sum_{j=1}^{p}\phi_j \rho_j} (k^2 + l^2 + 2kl\rho_m)\right], \\
&\quad if\ k\ and\ l\ are\ even \\
&= 0, \quad if\ k\ or\ l\ are\ odd
\end{aligned} \tag{B.7}$$

∎

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:

$$\mu_2(\theta) \equiv E(y_t^2) = \sigma_y^2 exp\left[\frac{1}{2}\gamma_0\right] \tag{B.8}$$

$$\mu_4(\theta) \equiv E(y_t^4) = 3\sigma_y^4 exp\left[2\gamma_0\right] \tag{B.9}$$

$$\mu_{2,2}(1|\theta) \equiv E(y_t^2 y_{t-1}^2) = \sigma_y^4 exp\left[\gamma_0(1+\rho_1)\right] \tag{B.10}$$

$$\mu_{2,2}(2|\theta) \equiv E(y_t^2 y_{t-2}^2) = \sigma_y^4 exp\left[\gamma_0(1+\rho_2)\right] \tag{B.11}$$

where $\gamma_0 = \sigma_v^2/(1-\sum_{j=1}^{2}\phi_j \rho_j)$

From (B.8) and (B.9), we get:

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

or,

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

From (B.8)

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

or,

$$\sigma_y = \frac{(E(y_t^2))^{1/2}}{(\frac{E(y_t^4)}{3(E(y_t^2))^2})^{1/4}} = \frac{3^{1/4}E(y_t^2)}{(E(y_t^4))^{1/4}} \tag{B.12}$$

From (B.8) and (B.10), we get:

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

or,

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

Similarly from (B.8) and (B.11), we get:

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

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:

$$\phi_1 = \frac{-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))(log(\frac{3E(y_t^2 y_{t-2}^2)}{E(y_t^4)}))}{(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2 - (log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))^2}, \tag{B.13}$$

$$\phi_2 = \frac{-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))^2 + (log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2}))(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))}{(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2 - (log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))^2} \tag{B.14}$$

$$\sigma_v = [log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) - \phi_1(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})) - \phi_2(log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2}))]^{1/2}. \tag{B.15}$$

∎

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

$$\begin{aligned}\zeta_1(\tau) =& Cov(X_{1,t}, X_{1,t+\tau}) = E([y_t^2 - \mu_2(\theta)][y_{t+\tau}^2 - \mu_2(\theta)]) \\ =& E(y_t^2 y_{t+\tau}^2) - \mu_2^2(\theta) = \sigma_y^4 exp[\gamma_0(1+\rho_\tau)] - \mu_2^2(\theta) = \mu_2^2(\theta)[exp(\gamma_\tau) - 1]\end{aligned} \tag{B.16}$$

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

$$\begin{aligned}\zeta_2(\tau) =& Cov(X_{2,t}, X_{2,t+\tau}) = E[y_t^4 - \mu_4(\theta)][y_{t+\tau}^4 - \mu_4(\theta)] = E(y_t^4 y_{t+\tau}^4) - \mu_4^2(\theta) \\ =& 9\sigma_y^8 exp[4\gamma_0(1+\rho_\tau)] - \mu_4^2(\theta) = \mu_4^2(\theta)[exp(4\gamma_\tau) - 1], \quad \forall \tau \geq 1\end{aligned} \tag{B.17}$$

and,

$$\begin{aligned}\zeta_3(\tau) =& Cov(X_{3,t}, X_{3,t+\tau}) = E([y_t^2 y_{t-1}^2 - \mu_{2,2}(1|\theta)][y_{t+\tau}^2 u_{t+\tau-1}^2 - \mu_{2,2}(1|\theta)]) \\ =& E[y_t^2 y_{t-1}^2 y_{t+\tau}^2 u_{t+\tau-1}^2] - \mu_{2,2}^2(1|\theta) \\ =& \sigma_y^8 E[exp(w_{t-1} + w_{t+\tau} + w_t + w_{t+\tau-1})] - \mu_{2,2}^2(1|\theta) \\ =& \sigma_y^8 exp[2(\gamma_0 + \gamma_1) + \gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}] - \mu_{2,2}^2(1|\theta) \\ =& \sigma_y^8 exp[2(\gamma_0 + \gamma_1)] exp[\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}] - \mu_{2,2}^2(1|\theta) \\ =& \mu_{2,2}^2(1|\theta)[exp(\gamma_{\tau-1} + 2\gamma_\tau + \gamma_{\tau+1}) - 1], \quad \forall \tau \geq 2\end{aligned} \tag{B.18}$$

Finally,

$$\begin{aligned}\zeta_4(\tau) =& Cov(X_{4,t}, X_{4,t+\tau}) = E([y_t^2 y_{t-2}^2 - \mu_{2,2}(2|\theta)][y_{t+\tau}^2 u_{t+\tau-2}^2 - \mu_{2,2}(2|\theta)]) \\ =& E[y_t^2 y_{t-2}^2 y_{t+\tau}^2 u_{t+\tau-2}^2] - \mu_{2,2}^2(2|\theta) \\ =& \sigma_y^8 E[exp(w_{t-2} + w_{t+\tau} + w_t + w_{t+\tau-2})] - \mu_{2,2}^2(2|\theta) \\ =& \sigma_y^8 exp[2(\gamma_0 + \gamma_2) + \gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}] - \mu_{2,2}^2(2|\theta) \\ =& \sigma_y^8 exp[2(\gamma_0 + \gamma_2)] exp[\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}] - \mu_{2,2}^2(2|\theta) \\ =& \mu_{2,2}^2(2|\theta)[exp(\gamma_{\tau-2} + 2\gamma_\tau + \gamma_{\tau+2}) - 1], \quad \forall \tau \geq 3\end{aligned} \tag{B.19}$$

∎

PROOF OF LEMMA **5.4** From (2.7), we get $w_t = y_t^* - \varepsilon_t$ and substitute this into (2.6) and with the results of Granger and Morris (1976), we have completed the proof. ∎

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

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j}, \tag{B.20}$$

or,

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + v_t + \varepsilon_t - \sum_{j=1}^{p} \phi_j \varepsilon_{t-j}. \tag{B.21}$$

Multiply both sides of (B.21) by y_{t-k}^* and taking expectation we get:

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

Setting $k = 0$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + E[v_t y_t^*] + E[\varepsilon_t y_t^*] - \sum_{j=1}^{p} \phi_j E[\varepsilon_{t-j} y_t^*] \\
&= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2 - \sum_{j=1}^{p} \phi_j E[\varepsilon_{t-j}(\phi_j y_{t-j}^* - \phi_j \varepsilon_{t-j})] \\
&= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2 - \sum_{j=1}^{p} \phi_j^2 E[\varepsilon_{t-j} y_{t-j}^* - \varepsilon_{t-j}^2] \\
&= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2 - \sum_{j=1}^{p} \phi_j^2 [\sigma_\varepsilon^2 - \sigma_\varepsilon^2] \\
&= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma_v^2 + \sigma_\varepsilon^2.
\end{aligned} \tag{B.22}$$

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

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + E[v_t y_{t-k}^*] + E[\varepsilon_t y_{t-k}^*] - \sum_{j=1}^{p} \phi_j E[\varepsilon_{t-j} y_{t-k}^*] \\
&= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + 0 + 0 - \phi_k E[\varepsilon_{t-k} y_{t-k}^*] = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) - \phi_k \sigma_\varepsilon^2.
\end{aligned} \tag{B.23}$$

Setting $k > p$, we get

$$\begin{aligned}
\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + E[v_t y_{t-k}^*] + E[\varepsilon_t y_{t-k}^*] - \sum_{j=1}^{p} \phi_j E[\varepsilon_{t-1} y_{t-k}^*] \\
&= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + 0 + 0 - 0 = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j).
\end{aligned} \tag{B.24}$$

Combining (B.22), (B.23), and (B.24), we get the autocovariance structure of the observed process stated at Lemma **5.5**. ∎

PROOF OF LEMMA **5.6**

The estimation of ϕ_p is based on the covariance 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,

$$\Gamma_{(k,p)}.\phi_p = \gamma_{(k,p)}$$

or,

$$\phi_p = \Gamma_{(k,p)}{}^{-1}\gamma_{(k,p)}, \tag{B.25}$$

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

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

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,

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

or, equivalently

$$\sigma_v = [\gamma_{y^*}(0) - \hat{\phi}_p'\hat{\gamma}_{(k,p)} - \psi'(1/2)]^{1/2}, \tag{B.26}$$

where $\sum_{j=1}^{p}\phi_j\gamma_{y^*}(j) = \hat{\phi}_p'\hat{\gamma}_{(k,p)}$, $\sigma_\varepsilon^2 = \psi'(1/2)$ and $\psi'(1/2)$ is equal to $\pi^2/2$.

Now by construction,

$$\mu = E[\log(y_t^2)] = \log(\sigma_y^2) + E[\log(z_t^2)] = \log(\sigma_y^2) - 1.27, \tag{B.27}$$

or, equivalently

$$\sigma_y^2 = exp(\mu + 1.27). \tag{B.28}$$

∎

PROOF OF LEMMA **5.7** Here we are using an alternative method provided by Durbin (1960) that avoids the matrix inversion in the Yule-Walker equations. We derive moment equations solution of SV(2) recursively from the moment equations solution of SV(1) and the results of lemma **5.7** easily identify from there.

At first we solve for the moment equations solution for an SV(1) as follows. We want to find closed form moment equations solution for $\Theta_1^{SV} \equiv (\phi_{11}, \sigma_{1v}, \sigma_y)$. Using Lemma **5.1**, and considering $k = 2$, $k = 4$,

$k = l = 2$ & $m = 1$, we get:

$$\mu_2(\theta) \equiv E(y_t^2) = \sigma_y^2 exp\left[\frac{1}{2}\gamma_0\right] \tag{B.29}$$

$$\mu_4(\theta) \equiv E(y_t^4) = 3\sigma_y^4 exp[2\gamma_0] \tag{B.30}$$

$$\mu_{2,2}(1|\theta) \equiv E(y_t^2 y_{t-1}^2) = \sigma_y^4 exp[\gamma_0(1+\rho_1)] \tag{B.31}$$

$$\text{where } \gamma_0 = \sigma_{1v}^2/(1-\phi_{11}\rho_1)$$

Solving the above equations yield:

$$\gamma_0 = log(\frac{E(y_t^4)}{3(E(y_t^2))^2}), \tag{B.32}$$

$$\sigma_y = \frac{(E(y_t^2))^{1/2}}{(\frac{E(y_t^4)}{3(E(y_t^2))^2})^{1/4}} = \frac{3^{1/4}E(y_t^2)}{(E(y_t^4))^{1/4}}, \tag{B.33}$$

$$\gamma_0\rho_1 = log(\frac{E(y_t^2 u_{t-1}^2)}{(E(y_t^2))^2}), \tag{B.34}$$

or

$$\rho_1 = log(\frac{E(y_t^2 u_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}). \tag{B.35}$$

we know that,

$$\gamma_1 = \phi_{11}\gamma_0 \Leftrightarrow \phi_{11} = \gamma_1/\gamma_0 = \rho_1 \tag{B.36}$$

$$\Leftrightarrow \phi_{11} = log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}), \tag{B.37}$$

and from the Yule-Walker equation we have:

$$\gamma_0 = \phi_{11}\gamma_1 + \sigma_v^2 \Leftrightarrow 1 = \phi_{11}\rho_1 + \sigma_v^2/\gamma_0 = \phi_{11}^2 + \sigma_v^2\gamma_0, \tag{B.38}$$

or,

$$\sigma_v = [(1-\phi_{11}^2)\gamma_0]^{1/2} = [(1-(log(\frac{E(y_t^2 u_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2)(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))]^{1/2} \tag{B.39}$$

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

$$\sigma_y = \frac{(E(y_t^2))^{1/2}}{(\frac{E(y_t^4)}{3(E(y_t^2))^2})^{1/4}} = \frac{3^{1/4}E(y_t^2)}{(E(y_t^4))^{1/4}}, \tag{B.40}$$

$$\phi_{11} = log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}), \tag{B.41}$$

$$\sigma_{1v} = [(1-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2)(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))]^{1/2}. \tag{B.42}$$

SV of order 2:

Here we want to find closed form estimator for $\theta \equiv (\phi_{21}, \phi_{22}, \sigma_{2v}, \sigma_y)$. Using Lemma 1, and considering $k = 2$, $k = 4$, $k = l = 2$ & $m = 1$ and $k = l = 2$ & $m = 2$, we get:

$$\mu_2(\theta) \equiv E(y_t^2) = \sigma_y^2 exp\left[\frac{1}{2}\gamma_0\right] \tag{B.43}$$

$$\mu_4(\theta) \equiv E(y_t^4) = 3\sigma_y^4 exp\left[2\gamma_0\right] \tag{B.44}$$

$$\mu_{2,2}(1|\theta) \equiv E(y_t^2 y_{t-1}^2) = \sigma_y^4 exp\left[\gamma_0(1+\rho_1)\right] \tag{B.45}$$

$$\mu_{2,2}(2|\theta) \equiv E(y_t^2 y_{t-2}^2) = \sigma_y^4 exp\left[\gamma_0(1+\rho_2)\right] \tag{B.46}$$

$$\text{where } \gamma_0 = \sigma_{2v}^2/(1-\textstyle\sum_{j=1}^2 \phi_{2j}\rho_j)$$

Solving the above equations yield:

$$\gamma_0 = log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) \tag{B.47}$$

$$\sigma_y = \frac{(E(y_t^2))^{1/2}}{(\frac{E(y_t^4)}{3(E(y_t^2))^2})^{1/4}} = \frac{3^{1/4}E(y_t^2)}{(E(y_t^4))^{1/4}} \tag{B.48}$$

$$\rho_1 = log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) \tag{B.49}$$

$$\rho_2 = log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) \tag{B.50}$$

One interesting observation is

$$\rho_j = log(\frac{E(y_t^2 y_{t-j}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}), \;\; where \;\; j = 1,2 \tag{B.51}$$

Now using Durbin-Levinson recurrence formula:

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1-\phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\rho_2 - \phi_{11}^2}{1-\phi_{11}^2} \tag{B.52}$$

and

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = \frac{\rho_1 - \rho_1\rho_2}{1-\rho_1^2} \tag{B.53}$$

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

$$\phi_{22} = \frac{log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) - (log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2}{1-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2}, \tag{B.54}$$

or,

$$\phi_{22} = \frac{-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))^2 + (log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2}))(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))}{(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2 - (log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))^2}, \tag{B.55}$$

and

$$\phi_{21} = \frac{log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) - (log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))(log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))}{1-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})/log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2}, \quad \text{(B.56)}$$

or,

$$\phi_{21} = \frac{-(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))(log(\frac{3E(y_t^2 y_{t-2}^2)}{E(y_t^4)}))}{(log(\frac{E(y_t^4)}{3(E(y_t^2))^2}))^2 - (log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2}))^2}, \quad \text{(B.57)}$$

and

$$\sigma_{2v} = [(1-\phi_{21}\rho_1 - \phi_{22}\rho_2)\gamma_0]^{1/2} \quad \text{(B.58)}$$

$$\sigma_{2v} = [log(\frac{E(y_t^4)}{3(E(y_t^2))^2}) - \phi_1(log(\frac{E(y_t^2 y_{t-1}^2)}{(E(y_t^2))^2})) - \phi_2(log(\frac{E(y_t^2 y_{t-2}^2)}{(E(y_t^2))^2}))]^{1/2}. \quad \text{(B.59)}$$

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

∎

PROOF OF LEMMA **7.1** This 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)). Under the assumptions A.1 and A.2 with $s = 2$, the observed process $\{y_t\}$ is strictly stationarity and geometrically ergodic with $E[y_t] < \infty$ and $E[y_t y_{t+k}] < \infty$. The stationary ergodic property is preserved under measurable transformations; that is, if $\{y_t\}$ is stationary and ergodic, so is the sequence $\{g(y_t)\}$ whenever $g : \mathbb{R} \mapsto \mathbb{R}$ is a measurable function (see Davidson (1994), page 291). So $E[g(y_t)] < \infty$ and $E[g(y_t)g(y_{t+k})] < \infty$ under a continuous measure preserving transformation and that completes the proof with $g(y_t) = log(y_t^2) - C$, where C is a constant. ∎

PROOF OF LEMMA **7.2**

To establish the asymptotic normality of empirical moments; we shall use a central limit theorem (C.L.T) for dependent processes (see Davidson (1994), Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this C.L.T holds. Setting

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

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

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

(i) This follows from result 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:

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

(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 $\beta_T = \psi\rho$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\phi})$, with $\phi > 1$ (see Davidson (1994), Definition 16.1, p. 247). To see that

$$\begin{aligned}\lim_{T\to\infty} \frac{\rho^T}{T^{-\phi}} &= \lim_{T\to\infty} T^{\phi} exp(T\log\rho) = \lim_{T\to\infty} exp(\phi\log T)exp(T\log\rho) \\ &= \lim_{T\to\infty} exp(\phi\log T + T\log\rho) = 0\end{aligned}$$

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

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

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

$$\begin{aligned}\zeta_\Psi(\tau) \equiv cov(\Psi_t, \Psi_{t+\tau}) =& E[(\log(y_t^2) - \mu)(\log(y_{t+\tau}^2) - \mu)] \\ =& E[y_t^* y_{t+\tau}^*] \\ =& \gamma_{y^*}(\tau),\end{aligned} \tag{B.60}$$

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

Now we have

$$\limsup_{T\to\infty} Var(T^{-1/2}S_{T1}) = \limsup_{T\to\infty}[\gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau)]$$
$$= \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau) \quad \text{(B.61)}$$
$$= \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau) \leq \sum_{\tau=-\infty}^{\infty}|\gamma_{y^*}(\tau)| < \infty.$$

This convergence is due to the fact that y_t^* follows an ARMA(1,1) process with $|\phi| < 1$. So y_t^* can be viewed as an MA(∞) process with absolute summable coeffcients and this implies absolute summability of autocovariances (see Hamilton 1993, chapter 3, page 52). We deduce that $\limsup_{T\to\infty} T^{-1/2}E|S_{T1}| < \infty$ 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 $\Lambda_{t,k} \equiv y_t^* y_{t+k}^* - \gamma_{y^*}(k)$, where $k = 0, 1, \ldots m$ and compute:

$$\begin{aligned}
\zeta_{\Lambda_k}(\tau) &\equiv cov(\Lambda_{t,k}, \Lambda_{t+\tau,k}) \\
&= E[(y_t^* y_{t+k}^* - \gamma_{y^*}(k))(y_{t+\tau}^* y_{t+\tau+k}^* - \gamma_{y^*}(k))] \\
&= E[y_t^* y_{t+k}^* y_{t+\tau}^* y_{t+\tau+k}^*] - \gamma_{y^*}^2(k) \\
&= E[y_t^* y_{t+k}^*]E[y_{t+\tau}^* y_{t+\tau+k}^*] + cov(y_t^*, y_{t+\tau}^*)cov(y_{t+k}^*, y_{t+\tau+k}^*) \\
&\quad + cov(y_t^*, y_{t+\tau+k}^*)cov(y_{t+k}^*, y_{t+\tau}^*) - \gamma_{y^*}^2(k) \\
&= \gamma_{y^*}^2(k) + \gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) - \gamma_{y^*}^2(k) \\
&= \gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) \quad \forall \tau \geq k,
\end{aligned} \quad \text{(B.62)}$$

$$\begin{aligned}
Var(T^{-1/2}S_{T2,k}) &= \frac{1}{T}\left[\sum_{t=1}^{T} Var(\Lambda_{t,k}) + \sum_{t\neq s} cov(\Lambda_{t,k}, \Lambda_{s,k})\right] = \frac{1}{T}\left[T\zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\Lambda_k}(\tau)\right] \\
&= \zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\Lambda_k}(\tau) \\
&= \gamma_{y^*}^2(0) + \gamma_{y^*}(k)\gamma_{y^*}(-k) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}^2(\tau) + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)].
\end{aligned} \quad \text{(B.63)}$$

Now we have

$$
\begin{aligned}
\limsup_{T\to\infty} Var(T^{-1/2}S_{T2,k}) =& \gamma_{y^*}^2(0)+\gamma_{y^*}(k)\gamma_{y^*}(-k) \\
&+\limsup_{T\to\infty}[2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}^2(\tau)+\gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)]] \\
=& \sum_{\tau=-\infty}^{\infty}[\gamma_{y^*}^2(\tau)+\gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)] \\
=& \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau)+\sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) \\
=& \underbrace{\sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau)}_{<\infty}+\underbrace{\sum_{\tau=-\infty}^{\infty}\gamma_{y^*}^2(\tau+k)}_{<\infty}<\infty.
\end{aligned}
\tag{B.64}
$$

The sums are converges due to the fact that absolute summability implies square-summability. We deduce that $\limsup_{T\to\infty} T^{-1/2}E|S_{T2,k}|<\infty$ 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 Cramer-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

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

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 (B.61) and (B.63) 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

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

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

$$
\begin{aligned}
C_{\mu,\Gamma(m)} =& \sum_t cov(\Psi_t, \Lambda_{t,k}) = 2\sum_{t=1}^{\infty} E[(log(y_t^2)-\mu)(y_t^* y_{t+k}^* - \gamma_{y^*})] = 2\sum_{t=1}^{\infty} E[y_t^*(y_t^* y_{t+k}^* - \gamma_{y^*})] \\
=& 2\sum_{t=1}^{\infty}(E[y_t^{*2} y_{t+k}^*] - \underbrace{E[y_t^*]}_{=0}\gamma_{y^*}) = 2\sum_{t=1}^{\infty} E[y_t^{*2} y_{t+k}^*]; \quad k = 0,1,2,\ldots,m.
\end{aligned} \tag{B.65}
$$

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

$$
\bar{c} \equiv C_{\mu,\Gamma(0)} = 2\sum_{t=1}^{\infty} E[y_t^{*3}] = 2\sum_{t=1}^{\infty}(E[w_t^3] + E[\varepsilon_t^3]) = 2\sum_{t=1}^{\infty} E[\varepsilon_t^3] \tag{B.66}
$$

where $E(\varepsilon_\tau^3) = -14\mathscr{Z}(3)$ and $\mathscr{Z}(\cdot)$ is Riemann's Zeta function with $\mathscr{Z}(3) = 1.20205$. For $k = 1,\ldots,m$, it is easily seen that $C_{\mu,\Gamma(m)} = 0$ from the property of w_t that it has an MA(∞) representation. 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 (B.66). ∎

PROOF OF THEOREM **7.3**

It is easily seen that D is a continuously differentiable mapping of $(\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(2p))$. The convergence result stated in (7.3) follows from the standard result for differentiable transformations of asymptotically normally distributed variables together with the application of multivariate delta method. ∎

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

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

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 Proposition **7.2**. ∎

PROOF OF LEMMA **7.2**

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

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

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

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

(i) This follows from result 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, we note that[6]:

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

(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 6, we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi\rho$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\phi})$, with $\phi > 1$ (see Davidson (1994), Definition 16.1, p. 247). To see that

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

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

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

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

$$\begin{aligned} Var(T^{-1/2} S_{T1}) =& \frac{1}{T}\left[\sum_{t=1}^{T} Var(X_{1,t}) + \sum_{t\neq s} cov(X_{1,s}, X_{1,t})\right] = \frac{1}{T}\left[T\zeta_1(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_1(\tau)\right] \\ =& \zeta_1(0) + 2\sum_{\tau=1}^{T}(1 - \frac{\tau}{T})\zeta_1(\tau) \end{aligned} \tag{B.67}$$

[6] see Davidson (1994), Theorem 14.2, p. 211

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 4. we have

$$\begin{aligned}\zeta_1(\tau) =&\mu_2^2(\theta)[exp(\gamma_\tau)-1]=\mu_2^2(\theta)\left[1+\sum_{k=1}^{\infty}\frac{\gamma_\tau^k}{k!}-1\right]=\mu_2^2(\theta)\left[\gamma_\tau\sum_{k=1}^{\infty}\frac{\gamma_\tau^{k-1}}{k!}\right]\\ =&\mu_2^2(\theta)\left[\gamma_\tau\sum_{k=0}^{\infty}\frac{\gamma_\tau^k}{(k+1)!}\right]\leq\mu_2^2(\theta)\left[\gamma_\tau\sum_{k=0}^{\infty}\frac{\gamma_\tau^k}{(k)!}\right]=\mu_2^2(\theta)\gamma_\tau exp(\gamma_\tau)\end{aligned} \tag{B.68}$$

Therefore, the series

$$\sum_{\tau=1}^{\infty}\zeta_1(\tau)\leq\mu_2^2(\theta)\sum_{\tau=1}^{\infty}\gamma_\tau exp(\gamma_\tau)\leq\mu_2^2(\theta)exp(\gamma_1)\sum_{\tau=1}^{\infty}\gamma_\tau\leq\mu_2^2(\theta)exp(\gamma_1)\underbrace{\sum_{\tau=1}^{\infty}|\gamma_\tau|}_{<\infty}<\infty \tag{B.69}$$

converges and by the Cauchy-Schwarz inequality we deduce that $\limsup_{T\to\infty}T^{-1/2}E\left|S_{T1}\right|<\infty$. 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 4. we have for all $\tau\geq 2$:

$$\begin{aligned}\zeta_3(\tau) &= \mu_{2,2}^2(1\,|\theta\,)[exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})-1]=\mu_{2,2}^2(1\,|\theta\,)\left[1+\sum_{k=1}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^k}{k!}-1\right]\\ &= \mu_{2,2}^2(1\,|\theta\,)\left[(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\sum_{k=1}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^{k-1}}{k!}\right]\\ &= \mu_{2,2}^2(1\,|\theta\,)\left[(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\sum_{k=0}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^{k}}{(k+1)!}\right]\\ &\leq \mu_{2,2}^2(1\,|\theta\,)\left[(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\sum_{k=0}^{\infty}\frac{(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})^{k}}{(k)!}\right]\\ &= \mu_{2,2}^2(1\,|\theta\,)(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\end{aligned} \tag{B.70}$$

Therefore, the series

$$\begin{aligned}\sum_{\tau=1}^{\infty}\zeta_3(\tau) &\leq \zeta_3(1)+\mu_{2,2}^2(1\,|\theta\,)\sum_{\tau=2}^{\infty}(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})exp(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\\ &\leq \zeta_3(1)+\mu_{2,2}^2(1\,|\theta\,)exp(\gamma_1+2\gamma_2+\gamma_3)\sum_{\tau=2}^{\infty}(\gamma_{\tau-1}+2\gamma_\tau+\gamma_{\tau+1})\\ &\leq \zeta_3(1)+\mu_{2,2}^2(1\,|\theta\,)exp(\gamma_1+2\gamma_2+\gamma_3)\left[\underbrace{\sum_{\tau=2}^{\infty}\left|\gamma_{\tau-1}\right|}_{<\infty}+2\underbrace{\sum_{\tau=2}^{\infty}|\gamma_\tau|}_{<\infty}+\underbrace{\sum_{\tau=2}^{\infty}\left|\gamma_{\tau+1}\right|}_{<\infty}\right]<\infty\end{aligned} \tag{B.71}$$

converges and by the Cauchy-Schwarz inequality we deduce that $\limsup_{T\to\infty} T^{-1/2}E|S_{T3}| < \infty$. The proof is very similar for S_{T4}. 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 Cramer-Wold theorem we can establish the limiting result that stated in proposition **7.2** .

∎

C. Tables

Table 1: Comparison of different estimation methods with respect to bias and RMSE for the SV(2) model using simulated data. GMM-6M-E and GMM-24M-E are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_*^{-1}$. GMM-6M-NW and GMM-24M-NW are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_{*,HAC}^{-1}$. Bayesian-MCMC is the Bayesian Markov Chain Monte Carlo methods of Grant and Chen (2016). EDV is the extension of Dufour and Valéry (2006) method and AD-ARMA is the simple ARMA based estimator of this study.

	$T = 500$				$T = 2000$			
	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
AD-ARMA	0.007	-0.023	0.003	0.016	0.003	-0.008	0.001	0.008
	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
AD-ARMA	0.198	0.193	0.016	0.185	0.084	0.081	0.007	0.091

Table 2: Comparison of different estimation methods with respect to bias and RMSE for the SV(2) model using simulated data. GMM-6M-E and GMM-24M-E are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_{*}^{-1}$. GMM-6M-NW and GMM-24M-NW are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_{*,HAC}^{-1}$. Bayesian-MCMC is the Bayesian Markov Chain Monte Carlo methods of Grant and Chen (2016). EDV is the extension of Dufour and Valéry (2006) method and AD-ARMA is the simple ARMA based estimator of this study.

	$T = 500$				$T = 2000$			
	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	-0.476	0.290	2.093	-0.658	-0.656	0.548	1.938	-0.565
AD-ARMA	0.019	-0.035	0.035	0.020	0.026	-0.031	0.006	0.003
	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	4.004	3.990	3.204	1.022	4.428	4.419	2.388	1.136
AD-ARMA	1.169	1.118	0.162	0.336	0.266	0.251	0.074	0.103

Table 3: Comparison of different estimation methods with respect to bias and RMSE for the SV(2) model using simulated data. GMM-6M-E and GMM-24M-E are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_{*}^{-1}$. GMM-6M-NW and GMM-24M-NW are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_{*,HAC}^{-1}$. Bayesian-MCMC is the Bayesian Markov Chain Monte Carlo methods of Grant and Chen (2016). EDV is the extension of Dufour and Valéry (2006) method and AD-ARMA is the simple ARMA based estimator of this study.

	$T = 500$				$T = 2000$			
	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	61.903	-2.248	0.234	0.758	76.059	-2.277
EDV	137.86	172.04	1383.13	27.51	0.01	55.74	1576.95	25.70
AD-ARMA	-0.002	0.001	0.001	-0.004	-0.002	0.000	-0.001	-0.003
	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	109.593	2.277
EDV	1166.50	1166.51	4926.86	55.33	118.80	118.34	3783.13	32.51
AD-ARMA	0.026	0.031	0.037	0.185	0.013	0.014	0.019	0.093

Table 4: Comparison of different estimation methods with respect to bias and RMSE for the SV(2) model using simulated data. GMM-6M-E and GMM-24M-E are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_*^{-1}$. GMM-6M-NW and GMM-24M-NW are 6 moments and 24 moments generalized method of moments with $\Omega = \Omega_{*,HAC}^{-1}$. Bayesian-MCMC is the Bayesian Markov Chain Monte Carlo methods of Grant and Chen (2016). EDV is the extension of Dufour and Valéry (2006) method and AD-ARMA is the simple ARMA based estimator of this study.

	$T = 500$				$T = 2000$			
	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
AD-ARMA	-0.004	-0.012	0.003	0.013	-0.002	-0.004	0.001	0.006
	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
AD-ARMA	0.031	0.033	0.016	0.188	0.014	0.014	0.007	0.093

Table 5: Comparison of different estimation methods with respect to elapsed time (in seconds) for the SV 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
AD-ARMA	0.64	1.41

Table 6: 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

Table 7: ARMA based 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	0.9982	(0.0222)	45.05	0.00	0.9775	(0.0553)	17.68	0.00
ϕ_2					-0.3992	(0.0553)	-7.22	0.00
σ_y	0.3356	(0.0167)	20.06	0.00	0.3356	(0.0167)	20.06	0.00
σ_v	0.9625	(0.0027)	359.16	0.00	1.0350	(0.0382)	27.11	0.00
Time (in seconds)	0.64				0.86			
	$p = 3$				$p = 4$			
	Coefficient	Std. Error	t-stat	p-value	Coefficient	Std. Error	t-stat	p-value
ϕ_1	0.9245	(0.0360)	25.65	0.00	0.8999	(0.0244)	36.86	0.00
ϕ_2	-0.3630	(0.0387)	-9.37	0.00	-0.5636	(0.0247)	-22.80	0.00
ϕ_3	0.4681	(0.0377)	12.40	0.00	0.1773	(0.0238)	7.45	0.00
ϕ_4					0.0203	(0.0238)	0.85	0.20
σ_y	0.3356	(0.0167)	20.06	0.00	0.3356	(0.0167)	20.06	0.00
σ_v	0.9609	(0.0382)	25.17	0.00	1.0434	(0.0169)	61.71	0.00
Time (in seconds)	1.27				1.84			

Table 8: Finite sample inference of SV(p) models using ARMA type 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	0.9982	45.05	0.000	0.05	0.01	0.001
σ_y	0.3355	20.06	0.000	0.05	0.01	0.001
σ_v	0.9625	359.16	0.000	0.05	0.01	0.001
Time (in seconds)			0.6	1.0	2.5	17.9
			$p=2$			
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.9775	17.68	0.000	0.05	0.01	0.001
ϕ_2	-0.3992	-7.22	0.000	0.05	0.01	0.001
σ_y	0.3355	20.06	0.000	0.05	0.01	0.001
σ_v	1.0350	27.11	0.000	0.05	0.01	0.001
Time (in seconds)			0.9	2.0	6.4	55.2
			$p=3$			
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.9245	25.65	0.000	0.05	0.01	0.001
ϕ_2	-0.3630	-9.37	0.000	0.05	0.01	0.001
ϕ_3	0.4681	12.40	0.000	0.05	0.01	0.001
σ_y	0.3356	20.06	0.000	0.05	0.01	0.001
σ_v	0.9609	25.17	0.000	0.05	0.01	0.001
Time (in seconds)			1.3	10.0	44.5	443.8
			$p=4$			
	Coefficient	S_0	Asymptotic tests	Local Monte Carlo tests		
				$N=19$	$N=99$	$N=999$
ϕ_1	0.8999	36.86	0.000	0.05	0.01	0.001
ϕ_2	-0.5636	-22.80	0.000	0.05	0.01	0.001
ϕ_3	0.1773	7.45	0.000	0.05	0.01	0.001
ϕ_4	0.0203	0.85	0.197	0.25	0.28	0.224
σ_y	0.3356	20.06	0.000	0.05	0.01	0.001
σ_v	1.0434	61.71	0.000	0.05	0.01	0.001
Time (in seconds)			1.8	16.6	77.6	771.5

Table 9: Summary Statistics of Logarithmic transformation of RV estimators.

S&P 500 index, 2000 - 2016, number of observations: 4200								
Series	Mean	SD	Kurtosis	Skewness	Range	Max	Min	LB(10)
RV5	-4.22	0.47	3.38	0.46	3.68	-2.11	-5.79	19968
BV5	-4.32	0.46	3.49	0.51	3.75	-2.22	-5.97	22168
RV10	-4.22	0.48	3.34	0.43	3.69	-2.11	-5.80	18605
RSV5	-4.57	0.52	3.23	0.37	4.11	-2.45	-6.57	14442
RK	-4.24	0.46	3.44	0.49	3.60	-2.03	-5.63	21782
RV5-SS	-4.29	0.45	3.53	0.51	3.67	-2.07	-5.74	22708
BV5-SS	-4.35	0.46	3.55	0.53	3.68	-2.10	-5.79	22851
RV10-SS	-4.31	0.46	3.50	0.48	3.69	-2.16	-5.84	21914
RSV5-SS	-4.62	0.49	3.28	0.37	3.95	-2.39	-6.34	17718

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, RK is the Realized Kernel and SS denotes the use of 1-minute subsamples.

Table 10: Summary Statistics of Logarithmic transformation of RV estimators.

S&P 500 index, 2000 - 2010, number of observations: 2738								
Series	Mean	SD	Kurtosis	Skewness	Range	Max	Min	LB(10)
RV5	-4.12	0.45	3.42	0.48	3.23	-2.11	-5.34	14595
BV5	-4.21	0.45	3.47	0.50	3.32	-2.22	-5.54	15672
RV10	-4.12	0.47	3.39	0.46	3.23	-2.11	-5.34	13441
RSV5	-4.47	0.50	3.23	0.37	3.53	-2.45	-5.98	10993
RK	-4.13	0.45	3.50	0.49	3.28	-2.03	-5.32	15183
RV5-SS	-4.18	0.44	3.57	0.52	3.30	-2.07	-5.37	15917
BV5-SS	-4.23	0.45	3.59	0.53	3.42	-2.10	-5.53	15979
RV10-SS	-4.19	0.45	3.52	0.48	3.35	-2.16	-5.51	15258
RSV5-SS	-4.51	0.47	3.32	0.36	3.54	-2.39	-5.94	12726

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, RK is the Realized Kernel and SS denotes the use of 1-minute subsamples.

Table 11: Forecasting performance of competing models.

Series: S&P 500 index, 2000-2016

	h=1				h=5				h=20			
	MSE		MAE		MSE		MAE		MSE		MAE	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
SV(1)	0.1220	3	0.1266	3	0.1227	3	0.1270	2	0.12510	2	0.1282	1
SV(2)	0.1214	1	0.1262	1	0.1225	1	0.1269	1	0.12502	1	0.1283	2
SV(3)	0.1215	2	0.1263	2	0.1226	2	0.1270	3	0.12512	3	0.1284	3
HAR-RV5	0.1978	4	0.1631	6	0.2151	7	0.1686	8	0.22416	9	0.1717	9
HAR-BV5	0.2264	9	0.1723	9	0.2252	9	0.1715	9	0.22751	10	0.1723	10
HAR-RV10	0.2277	10	0.1727	10	0.2286	11	0.1733	11	0.23174	11	0.1746	11
HAR-RSV5	0.2322	11	0.1745	11	0.2331	12	0.1751	12	0.23485	12	0.1757	12
HAR-RK	0.2347	12	0.1752	12	0.2256	10	0.1721	10	0.21909	8	0.1686	8
HAR-RV5-SS	0.2186	8	0.1681	8	0.2159	8	0.1668	7	0.21368	7	0.1645	7
HAR-BV5-SS	0.2140	7	0.1648	7	0.2124	5	0.1639	6	0.21245	5	0.1629	6
HAR-RV10-SS	0.2122	6	0.1623	5	0.2128	6	0.1626	5	0.21245	4	0.1615	5
HAR-RSV5-SS	0.2110	5	0.1609	4	0.2113	4	0.1605	4	0.21251	6	0.1608	4

1. The sample period is from January 3, 2000 to October 21, 2016 and the number of observations is T = 4200. The in-sample is from January 3, 2000 to December 31, 2009 (T=2486) and the out-of-sample is from January 3, 2010 to October 21, 2016 (T=1714). The in-sample include late-2000s Financial Crisis (Subprime mortgage crisis / United States housing bubble).
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, RK is the Realized Kernel and h stands for the forecast horizon.

3. SS denotes the use of 1-minute subsamples.

Table 12: Forecasting performance of competing models.

Series: S&P 500 index, 2000-2010

	h=1				h=5				h=20			
	MSE		MAE		MSE		MAE		MSE		MAE	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
SV(1)	2.082	3	0.471	3	2.110	3	0.4735	3	2.230	3	0.4866	1
SV(2)	2.075	1	0.469	2	2.103	2	0.4735	2	2.226	1	0.4875	2
SV(3)	2.077	2	0.468	1	2.102	1	0.4733	1	2.227	2	0.4883	3
HAR-RV5	3.354	4	0.606	4	3.568	4	0.6166	4	3.897	4	0.6417	4
HAR-BV5	3.926	5	0.647	5	3.996	5	0.6488	5	4.171	5	0.6740	5
HAR-RV10	4.199	6	0.673	6	4.232	6	0.6765	6	4.313	6	0.6785	6
HAR-RSV5	4.311	7	0.675	7	4.341	7	0.6799	7	4.446	7	0.6910	7
HAR-RK	4.465	8	0.693	8	4.497	8	0.6980	8	4.621	9	0.7106	9
HAR-RV5-SS	4.636	9	0.713	9	4.669	9	0.7174	9	4.791	10	0.7311	10
HAR-BV5-SS	4.804	10	0.733	10	4.833	10	0.7354	10	4.946	11	0.7525	11
HAR-RV10-SS	4.959	11	0.752	11	4.993	11	0.7552	11	5.110	12	0.7677	12
HAR-RSV5-SS	5.118	12	0.768	12	5.126	12	0.7627	12	4.580	8	0.6955	8

1. The sample period is from January 3, 2000 to December 31, 2007 and the number of observations is T = 2738. The in-sample is from January 3, 2000 to December 31, 2007 (T=1988) and the out-of-sample is from January 3, 2010 to December 31, 2010 (T=750). The out-of-sample include late-2000s Financial Crisis (Subprime mortgage crisis / United States housing bubble).
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, RK is the Realized Kernel and h stands for the forecast horizon.

3. SS denotes the use of 1-minute subsamples.

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.

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., 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.

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., 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.

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. (2000), Bias and MSE of the IV estimators under weak identification, Technical report, Department of Economics, University of Maryland.

Chernov, M., Ronald Gallant, A., 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* p. nbp001.

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.

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.

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 Zakoian, J.-M. (2011), *GARCH models: structure, statistical inference and financial applications*, Wiley. com.

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.

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', pNH.

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.

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.

Harvey, A. C. and Shephard, N. (1996), 'Estimation of an asymmetric stochastic volatility model for asset returns', *Journal of Business & 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.

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.

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 & 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.

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, 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.

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.

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.

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.

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.', *Economics Group, Nuffield College, University of Oxford, Economics Papers* .

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, Springer.

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 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.

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

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