Estimation and inference for higher-order stochastic volatility models with leverage [*]

Md. Nazmul Ahsan[†]
McGill University

Jean-Marie Dufour [‡]
McGill University

Gabriel Rodriguez-Rondon [§]
McGill University

February 29, 2024

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

[†] Department of Economics, McGill University, Leacock Building, 855 Sherbrooke StreetWest, Montréal, Québec H3A 2T7, Canada; e-mail: md.ahsan@mail.mcgill.ca.

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

[§] 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. e-mail: gabriel.rodriguezrondon@mail.mcgill.ca. Wed page: https://grodriguezrondon.com

ABSTRACT

Statistical inference – estimation and testing – for stochastic volatility models is challenging and computationally expensive. This problem is compounded when leverage effects are allowed. We propose efficient simple estimators for higher-order stochastic volatility models with leverage [SVL(p)], based on a small number of moment equations derived from ARMA representations associated with SVL models, along with the possibility of using "winsorization" to improve stability and efficiency (W-ARMA estimators). The asymptotic distributional theory of the estimators is derived. The computationally simple estimators proposed allow one to easily perform simulation-based (possibly exact) tests, such as Monte Carlo tests (MCT) or bootstrap procedures. In simulation experiments, we show that: (1) the proposed W-ARMA estimators dominate alternative estimators (including Bayesian estimators) in terms of bias, root-mean-square error, and computation time; (2) local and maximized Monte Carlo tests based on W-ARMA estimators yield good control of the size and power of LR-type tests; (3) taking into account leverage improves volatility forecasting. The methods developed are applied to daily returns for three major stock indices (S&P 500, Dow Jones, Nasdaq), confirming the superiority of SVL(p) models over competing conditional volatility models in terms of forecast accuracy.

Key words: Stochastic volatility, financial leverage, asymptotic distribution, higher-order process.
JEL Classification: C15, C22, C53, C58.

Contents

List of Tables

List of Figures

1 Introduction

Stochastic volatility (SV) models, originally proposed by Taylor (1982, 1986), constitute a fundamental model of fluctuating volatility in financial and macroeconomic time series. In this paper, we consider higher-order stochastic volatility models with a leverage effect [SVL(p)], where the underlying volatility process follows an autoregressive process of order p. The leverage effect refers to the inverse relationship between asset returns and volatilities: when asset prices decrease, volatility tends to increase due to the rise in debt-equity ratio, making the asset riskier; see Black (1976) and Christie (1982) for initial discussions. SVL models are empirically significant, due to the ability of capturing asymmetric effects in financial markets. Applications include modeling volatility clustering and persistence [Harvey and Shephard (1996)], improving option pricing [Hull and White (1987), Bates and Watts (1988)], implementing more effective risk mitigation strategies [Chen et al. (2019)], making better-informed asset allocation decisions [Andersen and Bollerslev (1998)], forecasting volatility during periods of market stress and uncertainty [Barndorff-Nielsen et al. (2002)], providing a better empirical understanding of market behavior [Andersen et al. (2003)], etc.

While the SVL model is conceptually simple and theoretically appealing, it poses difficulties for estimation due to the presence of latent variables. Inference on SVL models are thus challenging tasks. The existing methods are mostly limited to the SVL(1) model, computationally expensive, difficult to implement, and typically inefficient. Notable work includes quasi-maximum likelihood (QML) [Harvey and Shephard (1996)] and Bayesian techniques based on Markov Chain Monte Carlo (MCMC) methods [Jacquier et al. (2004), Omori et al. (2007)]. The estimation of SVL(p) models is even more challenging and remains unexplored in the literature, primarily due to their intrinsic complexity.

In this paper, we extend the simple estimator of Ahsan and Dufour (2021) designed for SV(p) models, and we develop a closed-form estimator for SVL(p) models based on the moment structure of the logarithm of squared residual returns. The proposed moment-based closed-form estimator [which we call the CF-ARMA estimator] utilizes the ARMA representation associated with SVL models. It offers analytical tractability and computational efficiency. Specifically, it can be easily implemented without relying on numerical optimization, and eliminates the need for arbitrary initial parameters or auxiliary models.

However, the CF-ARMA estimator may not satisfy stationarity conditions when the sample is not sufficiently large or in the presence of outliers. To overcome this issue, we consider two approaches.

1. A restricted estimation approach, which confines the estimates within an acceptable parameter solution space by adjusting the roots that lie on or outside the unit circle.
2. Winsorized versions of the ARMA-type estimator [which we call W-ARMA estimators]. These modifications significantly increase the likelihood of obtaining acceptable values and improve efficiency [see Hafner and Linton (2017), Ahsan and Dufour (2021)].

In winsorized estimators, the autoregressive parameters of the latent volatility process – which capture volatility clustering – are estimated using an OLS-based weighting. For details on other weighting schemes and simulation experiments regarding precision improvement, see Ahsan and Dufour (2019, 2021). This computationally simple adjustment enhances the stability and accuracy of the estimators. In particular, an OLS-based W-ARMA estimator consistently outperforms all other estimators in terms of bias and RMSE by a significant margin, including the Bayesian estimator proposed in this context. The W-ARMA-OLS estimator proposed in this paper can be interpreted as a *parsimonious moment-based* estimator, where only a few well-chosen moments are utilized. In moment-based (or GMM) inference, employing too many moments can prove to be very costly from the perspective of estimation efficiency and forecasting [see Ahsan and Dufour (2020)]. To be more specific, the other contributions of the paper can be summarized as follows.

First, utilizing these simple estimators, we formulate recursive estimation methods for SVL(p) models by using a Durbin-Levinson-type (DL) algorithm. We discuss the algorithmic framework that extends the recursion outlined in Tsay and Tiao (1984), which is applicable to autoregressive-moving average models and facilitates the recursive computation of parameters for higher-order SV processes. In addition, the asymptotic properties of the proposed simple estimators are established under standard regularity assumptions, demonstrating consistency and asymptotic normality when the fourth moment of the latent volatility process exists.

Second, we emphasize that the proposed computationally inexpensive estimators can be useful in several ways.

- Since SVL models are parametric models involving only a finite number of unknown parameters (which can be easily estimated by our method), one can construct simulation-based tests, even exact tests based on the Monte Carlo test (MCT) technique [see Dufour (2006)], as opposed to procedures based on establishing asymptotic distributions. In particular, exact tests obtained in this way do not depend on stationarity assumptions, so that they can be useful when the latent volatility process has a unit root (or is close to this structure).
- The proposed estimators are helpful for procedures which require repeated estimation based on a rolling window method, *e.g.* backtesting of risk measures (such as Value-at-Risk or Expected Shortfall) in the context of risk management.
- Due to the $\sqrt{T}$-consistency, our simple estimators can be effortlessly applied to very large samples, which are not rare in empirical finance, *e.g.* high-frequency financial returns. In these situations, estimators based on simulation techniques and/or numerical optimization often require substantial computational effort to achieve convergence.

Third, we study by simulation the statistical properties of our estimators and compare them with QML and Bayesian methods. The simulation results confirm that W-ARMA estimators have excellent statistical properties in terms of bias and RMSE. In particular, it outperforms all other estimators studied, including the Bayesian estimator. Furthermore, the proposed ARMA estimators (CF-ARMA and W-ARMA) are extremely efficient in terms of computation time, especially when compared with Bayesian estimators.

Fourth, in a further simulation study, we compare the forecasting performance of SV(p) models with SVL(p) models in the presence of leverage in the true data. We consider different values of the leverage parameter for both $p = 1$ and $p = 2$ models. By employing the MSE loss function with the Diebold-Mariano (DM) test, we find that SVL(p)

models consistently outperform SV(p) models in terms of forecast accuracy. We investigate two data generating processes (DGPs) with $p = 1, 2$, as well as different forecasting horizons. Our findings reveal that, even though the difference in mean squared error (MSE) decreases with longer forecasting horizons, a statistically significant difference persists which favours SVL models.

Fifth, we study the performance of likelihood-ratio-type (LR-type) tests based on W-ARMA estimators for the no leverage hypothesis. Three approaches for controlling the level of the tests are considered: (1) using a standard chi-square asymptotic approximation; (2) local Monte Carlo (LMC) tests (or parametric bootstrapping); (3) maximized Monte Carlo (MMC) tests. Our results show that asymptotic tests are quite unreliable in all scenarios, and exhibit very high degree of over-rejection frequencies. While the LMC method offers improved performance, it still exhibits some degree of over-rejection, particularly evident when the volatility process is highly persistent. In contrast, the MMC procedure consistently maintains control over size, demonstrating robustness even with high persistence. Additionally, when considering level-corrected power (for Asymptotic and LMC tests), we find that LMC performs comparably to infeasible asymptotic tests (infeasible because the true distribution of the test statistic is unknown), while MMC stands out for its effectiveness in power. Moreover, we observe a decline in power with increasing volatility persistence, although larger sample sizes substantially increase power.

Finally, we present applications to daily observations on the returns for three major stock indices – S&P 500, Dow Jones, and NASDAQ – over the period 2000-2023. Using W-ARMA estimation, we find that the returns on these stocks exhibit stochastic volatility with strong persistence. The leverage parameter exhibits significantly high values for all three indices when estimating the SVL(1) model, while low values are observed for SVL(2) and SVL(3) models. Moreover, we implemented MCT techniques to construct more reliable finite-sample inference, particularly since the estimated volatility persistence parameter is close to the unit circle. Tests of on leverage reject the null hypothesis of no leverage when estimating the SVL(1) model. However, for the SVL(2) and SVL(3) models, the asymptotic test still rejects the null hypothesis of no leverage. Conversely, LMC and MMC methods suggest that for SVL(2) and SVL(3) models, we lack sufficient evidence to reject the null hypothesis of no leverage. Our forecasting experiment also shows that the out-of-sample forecasting performance of SVL(p) models is superior to alternative conditional volatility models [GARCH and SV(p)].

This paper is organized as follows. Section 2 specifies the model, assumptions, and motivation. Section 3 describes the simple estimators for SVL(p) models. Section 4 develops the asymptotic distributional theory for these estimators, and Section 5 discusses volatility forecasting with SVL(p) models. Section 6 discusses how finite-sample Monte Carlo tests can be applied using the proposed simple estimator. Section 7 presents the simulation study related to estimation, testing and forecasting. The empirical application is presented in Section 8. We conclude in Section 9. The mathematical proofs, other discussions, and additional results are provided in Appendix.

2 Framework

We consider a discrete-time SVL process of order (p), which can be viewed as an extension of the models considered by Harvey and Shephard (1996) and Omori et al. (2007). Specifically, we say that a variable y_t follows a stationary

SVL(p) process if it satisfies the following assumptions, where $t \in \mathbb{N}_0$, and $\mathbb{N}_0$ represents the non-negative integers.

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

$$y_t = \sigma_y \exp(w_t/2) z_t , \quad z_t \sim i.i.d.\ \mathcal{N}(0,1) , \tag{2.1}$$

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \sigma_v v_t , \quad v_t \sim i.i.d.\ \mathcal{N}(0,1) , \tag{2.2}$$

$$\mathbb{E}[z_{t-1}, v_t] = \mathrm{corr}(z_{t-1}, v_t) = \delta_p , \tag{2.3}$$

where $(\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v, \delta_p)'$ *are fixed parameters,* $\sigma_y > 0$ *and* $\sigma_v > 0$ *control for the level and variance of the latent log volatility process, and* $\delta_p \in (-1, 1)$ *and for any* $i \neq 1$ *and any* t, $\mathrm{corr}(z_{t-i}, v_t) = 0$. δ_p *is the* leverage parameter, *which represents the lag correlation between returns and log volatility shocks.*

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

The latter assumption requires that all the roots of the characteristic equation of the volatility process $[\phi(z) = 0]$ lie outside the unit circle [*i.e.*, $\phi(z) \neq 0$ for $|z| \leq 1$]. This model consists of two stochastic processes, where y_t describes the dynamics of asset returns, and $w_t := \log(\sigma_t^2)$ captures the dynamics of latent log volatilities.[1] The latent process w_t can be interpreted as a random flow of uncertainty shocks or new information in financial markets, the ϕ_j's capture volatility persistence, and δ_p represents the leverage effect. If $\delta_p < 0$, given a negative shock to y_t at time $t-1$, the volatility at time t tends to be larger.

Compared with the SV(p) model in Ahsan and Dufour (2021), a dependence structure between z_{t-1} and v_t is allowed in this SVL(p) model.[2] So the SVL(p) model, defined in Assumptions 2.1 - 2.2, has $p+3$ fixed parameters and has one additional leverage parameters compared to Ahsan and Dufour (2021). On setting $p = 1$, we get the SVL(1) model, considered by Harvey and Shephard (1996), Omori et al. (2007) and Yu (2005).[3] This model is the Euler approximation to the continuous time asymmetric SV model widely used in the option pricing literature; see for example Hull and White (1987), Wiggins (1987), and Chesney and Scott (1989). This SVL(1) model is estimated by QML in Harvey and Shephard (1996) and by MCMC in Omori et al. (2007). As argued in Harvey and Shephard (1996), this model is a martingale difference sequence, consistent with the efficient market hypothesis. This is obvious because for SVL models, we have:

$$\mathbb{E}[y_{t+1}|y_t, w_t] = \sigma_y \exp\big(\phi w_t/2\big)\mathbb{E}[\exp\big(\sigma_v v_{t+1}/2\big)]\mathbb{E}[z_{t+1}|y_t, w_t] = 0 . \tag{2.4}$$

[1] Usually the y_t's are residual returns, such that $y_t := r_t - \mu_r$ and $r_t := 100[\log(p_t) - \log(p_{t-1})]$, where μ_r is the mean of returns (r_t) and p_t is the raw prices of an asset. It is noteworthy to mention that y_t is ordinarily the error term of any time series regression model, see Jurado et al. (2015).

[2] The leverage effect is that stock volatility is correlated to stock returns and it is an empirical fact that stock volatility tends to increase when stock prices drop. There are two common economic explanations for the leverage effect. The first explanation is based on the relationship between volatility and expected returns. When volatility rises, expected returns tend to increase, leading to a drop in the stock price. As a consequence, volatility and stock returns are negatively correlated. The second explanation is based on financial leverage. When stock prices fall, financial leverage increases, leading to an increase in stock return volatility.

[3] Instead of assuming $\mathrm{corr}(z_{t-1}, v_t) = \delta_p$, Jacquier et al. (2004) adopt the specification of $\mathrm{corr}(z_t, v_t) = \delta_p$.

Since the disturbances $(z_{t-1}, v_t')'$ are conditionally Gaussian, we can write

$$v_t = \delta_p z_{t-1} + (1-\delta_p^2)^{1/2}\bar{v}_t\,, \tag{2.5}$$

where $\bar{v}_t \sim N(0,1)$. The state equation (2.2) can then be reformulated as:

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \delta_p \sigma_v z_{t-1} + (1-\delta_p^2)^{1/2}\sigma_v \bar{v}_t\,. \tag{2.6}$$

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

$$\log(y_t^2) = \log(\sigma_y^2) + w_t + \log(z_t^2) = \mu + w_t + \epsilon_t \tag{2.7}$$

where $\mu := \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)]$ and $\epsilon_t := \log(z_t^2) - \mathbb{E}[\log(z_t^2)]$. Under the normality assumption on z_t, the errors ϵ_t are i.i.d. according to the distribution of a centered $\log(\chi_1^2)$ random variable [*i.e.*, ϵ_t has mean zero and variance $\mathbb{E}(\epsilon_t^2)$] and

$$\mathbb{E}[\log(z_t^2)] = \psi(1/2) + \log(2) \simeq -1.2704\,, \quad \sigma_\epsilon^2 := \mathbb{E}(\epsilon_t^2) = \pi^2/2\,, \quad \mathbb{E}(\epsilon_t^3) = \psi^{(2)}(1/2)\,, \quad \mathbb{E}(\epsilon_t^4) = \pi^4 + 3\sigma_\epsilon^2\,, \tag{2.8}$$

where $\psi^{(2)}(z)$ is the *polygamma function* of order 2; see Abramowitz and Stegun (1970, Chapter 6).[4] On setting

$$y_t^* := \log(y_t^2) - \mu \tag{2.9}$$

and combining (2.6) and (2.7), the SVL(p)model can be written in state-space form:

$$\text{State Transition Equation:} \quad w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + \tilde{\delta}_p z_{t-1} + \tilde{\sigma}_v \bar{v}_t\,, \tag{2.10}$$

$$\text{Measurement Equation:} \quad y_t^* = w_t + \epsilon_t\,, \tag{2.11}$$

where $\tilde{\delta}_p := \delta_p \sigma_v$, $\tilde{\sigma}_v := \sqrt{(1-\delta_p^2)}\sigma_v$, and $\bar{v}_t$'s are i.i.d. $\mathcal{N}(0,1)$ and ϵ_t's are i.i.d. $\log(\chi_1^2)$. Note that transformed structural error $\bar{v}_t$ is uncorrelated with z_t and also its transformed error ϵ_t.

3 Simple SV estimation with leverage

In this section, we propose simple estimators for SVL(p) models, including a corresponding recursive procedure. Besides, we suggest alternative methods to improve the performance of these simple estimators.

3.1 ARMA-based estimation

In this subsection, we propose another simple estimator for SVL(p) models, which exploits the autocovariance structure of y_t^*. This estimator is based on considering a small set of moments associated with $y_t^* = \log(y_t^2) - \mu$, and the fact that this process has an ARMA representation.

[4] The $\log(\chi_1^2)$ distribution is often approximated by a normal distribution with mean of -1.2704 and variance of $\pi^2/2$ [see Broto and Ruiz (2004)], or by a mixture distribution [Kim et al. (1998)].

Proposition 3.1. ARMA REPRESENTATION OF SVL(P) MODELS. *Under the Assumptions 2.1 - 2.2, the process y^*_t defined in (2.9) has the following ARMA(p,p) representation*:

$$y^*_t = \sum_{j=1}^{p} \phi_j y^*_{t-j} + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j}\,, \tag{3.1}$$

$$\eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} = v_t + \epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j}, \tag{3.2}$$

where $\{v_t\}$ and $\{\epsilon_t\}$ are mutually independent error processes, the errors v_t are i.i.d. $\mathcal{N}(0,\sigma_v^2)$, and the errors ϵ_t are i.i.d. according to the distribution of a $\log(\chi_1^2)$ random variable.

From the above proposition, we derive simple expressions for the autocovariances and parameters of the SVL(p) model.

Corollary 3.2. AUTOCOVARIANCES OF THE OBSERVED PROCESS. *Under the Assumptions of Proposition 3.1, the autocovariances of the observed process y^*_t defined in (2.9) satisfy the following equations*:

$$\mathrm{cov}(y^*_t, y^*_{t-k}) := \gamma_{y^*}(k) = \begin{cases} \phi_1\gamma_{y^*}(k-1) + \cdots + \phi_p\gamma_{y^*}(k-p) + \sigma_v^2 + \sigma_\epsilon^2 & \text{if } k = 0 \\ \phi_1\gamma_{y^*}(k-1) + \cdots + \phi_p\gamma_{y^*}(k-p) - \phi_k\sigma_\epsilon^2 & \text{if } 1 \le k \le p \\ \phi_1\gamma_{y^*}(k-1) + \cdots + \phi_p\gamma_{y^*}(k-p) & \text{if } k > p\,. \end{cases} \tag{3.3}$$

In order to get the unsquared parameter solution for δ_p, we will exploit the following lemma.

Lemma 3.3. MOMENT EQUATION INVOLVING UNSQUARED LEVERAGE PARAMETER. *Under Assumptions 2.1 - 2.2, the following moment condition holds:*

$$\mathbb{E}\big[|y_t|y_{t-1}\big] = \frac{\delta\sigma_v\sigma_y^2}{\sqrt{2\pi}}\exp\Big(\tilde{\gamma}_p/4\Big) \tag{3.4}$$

where $\tilde{\gamma}_p := \mathrm{var}(w_t) + \mathrm{cov}(w_t, w_{t-1})$.

Corollary 3.4. CLOSED-FORM EXPRESSIONS FOR SVL(p) PARAMETERS. *Under the Assumptions of Proposition 3.1, we have*:

$$\phi_p = \boldsymbol{\Gamma}(p,j)^{-1}\gamma(p,j), \quad j \ge 1\,, \tag{3.5}$$

$$\sigma_y = [\exp(\mu + 1.2704)]^{1/2}\,, \quad \sigma_v = [\gamma_{y^*}(0) - \phi_p^{'}\gamma(1) - \pi^2/2]^{1/2}, \tag{3.6}$$

$$\delta_p = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v\sigma_y^2}\exp\Big(-\frac{1}{4}\tilde{\gamma}_p\Big), \tag{3.7}$$

where $\lambda_y(1) := \mathbb{E}[|y_t|y_{t-1}]$, $\tilde{\gamma}_p := \mathrm{var}(w_t) + \mathrm{cov}(w_t, w_{t-1})$, $\phi_p := (\phi_1, \ldots, \phi_p)^{'}$,

$$\gamma(p,j) := [\gamma_{y^*}(p+j), \ldots, \gamma_{y^*}(2p+j-1)]^{'}, \quad \gamma(p) := [\gamma_{y^*}(1), \ldots, \gamma_{y^*}(p)]^{'} \tag{3.8}$$

are $p \times 1$ vectors, and $\boldsymbol{\Gamma}(p, j)$ is the following $p \times p$ matrix:

$$\boldsymbol{\Gamma}(p, j) := \begin{bmatrix} \gamma_{y^*}(j+p-1) & \gamma_{y^*}(j+p-2) & \cdots & \gamma_{y^*}(j) \\ \gamma_{y^*}(j+p) & \gamma_{y^*}(j+p-1) & \cdots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(j+2p-2) & \gamma_{y^*}(j+2p-3) & \cdots & \gamma_{y^*}(j+p-1) \end{bmatrix}, \tag{3.9}$$

where p is the SV order, $\gamma_{y^*}(k) = \mathrm{cov}(y_t^*, y_{t-k}^*)$, with $y_t^* = [\log(y_t^2) - \mu]$ and $\mu := \mathbb{E}[\log(y_t^2)]$. Expressions of δ_p for the lower order SVL(p) models are:

$$\delta_1 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{\sigma_v^2}{1-\phi_1}\Big), \tag{3.10}$$

$$\delta_2 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{\sigma_v^2}{(1-\phi_1-\phi_2)(1+\phi_2)}\Big), \tag{3.11}$$

$$\delta_3 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{(1-\phi_3)\sigma_v^2}{(1-\phi_1-\phi_2-\phi_3)(1+\phi_1\phi_3+\phi_2-\phi_3^2)}\Big), \tag{3.12}$$

$$\delta_4 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{(1-\phi_3-\phi_1\phi_4-\phi_4^2)(1-\phi_1-\phi_2-\phi_3-\phi_4)^{-1}\sigma_v^2}{(1+\phi_1\phi_3+\phi_2+\phi_4+2\phi_2\phi_4+\phi_1^2\phi_4+\phi_2\phi_4^2-\phi_3^2-\phi_4^2-\phi_4^3-\phi_1\phi_3\phi_4)}\Big). \tag{3.13}$$

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

$$\hat{\gamma}_{y^*}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k} y_t^* y_{t+k}^*, \quad \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\log(y_t^2), \quad \hat{\lambda}_y(1) := \frac{1}{T-1}\sum_{t=1}^{T-1}[|y_t| y_{t-1}], \tag{3.14}$$

where y_t^* is a mean corrected process (by construction). Setting $j = 1$ in (3.5) and replacing theoretical moments by their corresponding empirical moments in (3.5)-(3.7) yield the following *simple CF-ARMA* estimator of the SVL(p) coefficients:

$$\hat{\phi}_p = \hat{\boldsymbol{\Gamma}}(p,1)^{-1}\hat{\gamma}(p,1), \quad \hat{\sigma}_y = [\exp(\hat{\mu}+1.2704)]^{1/2}, \quad \hat{\sigma}_v = [\hat{\gamma}_{y^*}(0) - \hat{\phi}_p^{'}\hat{\gamma}(p) - \pi^2/2]^{1/2}, \tag{3.15}$$

$$\hat{\delta}_p = \frac{\sqrt{2\pi}\hat{\lambda}_y(1)}{\hat{\sigma}_v \hat{\sigma}_y^2}\exp\Big(-\frac{1}{4}\hat{\bar{\gamma}}_p\Big), \tag{3.16}$$

where expressions of $\hat{\delta}_p$ for the lower order SVL(p) models are:

$$\hat{\delta}_1 = \frac{\sqrt{2\pi}\hat{\lambda}_y(1)}{\hat{\sigma}_v \hat{\sigma}_y^2} \exp\Big(-\frac{1}{4}\frac{\hat{\sigma}_v^2}{1-\hat{\phi}_1}\Big), \tag{3.17}$$

$$\hat{\delta}_2 = \frac{\sqrt{2\pi}\hat{\lambda}_y(1)}{\hat{\sigma}_v \hat{\sigma}_y^2} \exp\Big(-\frac{1}{4}\frac{\hat{\sigma}_v^2}{(1-\hat{\phi}_1-\hat{\phi}_2)(1+\hat{\phi}_2)}\Big), \tag{3.18}$$

$$\hat{\delta}_3 = \frac{\sqrt{2\pi}\hat{\lambda}_y(1)}{\hat{\sigma}_v \hat{\sigma}_y^2} \exp\Big(-\frac{1}{4}\frac{(1-\hat{\phi}_3)\hat{\sigma}_v^2}{(1-\hat{\phi}_1-\hat{\phi}_2-\hat{\phi}_3)(1+\hat{\phi}_1\hat{\phi}_3+\hat{\phi}_2-\hat{\phi}_3^2)}\Big), \tag{3.19}$$

$$\hat{\delta}_4 = \frac{\sqrt{2\pi}\hat{\lambda}_y(1)}{\hat{\sigma}_v \hat{\sigma}_y^2} \exp\Big(-\frac{1}{4}\frac{(1-\hat{\phi}_3-\hat{\phi}_1\hat{\phi}_4-\hat{\phi}_4^2)(1-\hat{\phi}_1-\hat{\phi}_2-\hat{\phi}_3-\hat{\phi}_4)^{-1}\hat{\sigma}_v^2}{(1+\hat{\phi}_1\hat{\phi}_3+\hat{\phi}_2+\hat{\phi}_4+2\hat{\phi}_2\hat{\phi}_4+\hat{\phi}_1^2\hat{\phi}_4+\hat{\phi}_2\hat{\phi}_4^2-\hat{\phi}_3^2-\hat{\phi}_4^2-\hat{\phi}_4^3-\hat{\phi}_1\hat{\phi}_3\hat{\phi}_4)}\Big). \tag{3.20}$$

3.2 Restricted estimation

These simple estimators yield values outside the admissible area, *i.e.*, some of the roots of the latent volatility AR(p) model may not correspond to a stationary process. This issue can arise especially in small samples or in the presence of outliers. When this happens, a simple fix consists in altering the eigenvalues that lie on or outside the unit circle. The characteristic equation of the latent AR(p) process is given by $C(\lambda) = \lambda^p - \phi_1 \lambda^{p-1} - \cdots - \phi_p = 0$, and the stationary condition requires all roots lie inside the unit circle, *i.e.*, $|\lambda_i| < 1,\ i = 1, \ldots, p$. If the estimated parameters fail to satisfy this condition, then the restricted estimation can be done in the following two steps:

1. given the estimated unstable parameters, we calculate the roots of the characteristic equation and restrict their absolute values to less than unity;
2. given these restricted roots, we calculate constrained parameters ensuring stationarity.

For an SVL(2) model, the characteristic equation of the latent volatility process is $C(\lambda) = \lambda^2 - \phi_1 \lambda - \phi_2 = 0$. It may have two types of roots: (i) if $\phi_1^2 + 4\phi_2 \geq 0$, then $C(\lambda)$ has two real roots given by $\lambda_{1,2} = [\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}]/2$; (ii) if $\phi_1^2 + 4\phi_2 < 0$ then $C(\lambda)$ has two complex roots given by $\lambda_{1,2} = [\phi_1 \pm i\sqrt{-(\phi_1^2 + 4\phi_2)}]/2$. When the estimated polynomial coefficients produce an unstable solution, then we restrict the absolute value of the roots less than unity, *i.e.* $|\lambda_{1,2}| < 1$ or $|\lambda_{1,2}| = 1 - \Delta$ where Δ is a very small number. Given these restricted roots, we solve for restricted parameters which ensure stationarity. These steps can be done very easily in MATLAB. In MATLAB, the **roots** function calculates the roots given the parameters, and the **poly** function calculates the parameters given the roots. Further, the estimate $\hat{\delta}_p$ may yield a value outside the acceptable region $(-1, 1)$. In this case, we can restrict it within the admissible parameter space.

3.3 ARMA-based winsorized estimation

We can achieve better stability and efficiency of the CF-ARMA estimator by using "winsorization". This procedure substantially increases the probability of getting admissible values. From (3.5), it is easy to see that

$$\phi_p = \sum_{j=1}^{\infty} \omega_j \boldsymbol{B}(p, j) = \sum_{j=1}^{\infty} \omega_j \boldsymbol{\Gamma}(p, j)^{-1} \gamma(p, j) \tag{3.21}$$

for any ω_j sequence with $\sum_{j=1}^{\infty} \omega_j = 1$, where

$$\boldsymbol{B}(p, j) := \boldsymbol{\Gamma}(p, j)^{-1} \gamma(p, j) = [\boldsymbol{B}(p, j)_1, \ldots, \boldsymbol{B}(p, j)_p]' \tag{3.22}$$

is a $p \times 1$ vector. Using (3.21), we can define a more general class of estimators for ϕ_p by taking a weighted average of several sample analogs of the $p \times 1$ vector:

$$\tilde{\phi}_p := \sum_{j=1}^{J} \omega_j \hat{\boldsymbol{B}}(p, j), \quad \hat{\boldsymbol{B}}(p, j) := \hat{\boldsymbol{\Gamma}}(p, j)^{-1} \hat{\gamma}(p, j) = [\hat{\boldsymbol{B}}(p, j)_1, \ldots, \hat{\boldsymbol{B}}(p, j)_p]', \tag{3.23}$$

where $\sum_{j=1}^{J} \omega_j = 1$, $1 \leq J \leq T - p$, and T is the length of time series. We can expect that a sufficiently general class of weights will improve the efficiency of the CF-ARMA estimator.

Using (3.23), we propose alternative estimators of $\phi_{\boldsymbol{p}}$ and we call these estimators *winsorized ARMA* estimators (or *W-ARMA* estimators). For a full list of W-ARMA methods, see Appendix B. These estimation techniques have been examined within the framework of SV(p) models by Ahsan and Dufour (2021) and have also been explored by EGARCH(1, 1) in the context of closed-form estimation for the EGARCH(1, 1) model.

In this paper, we use a W-ARMA estimator based on an OLS regression without intercept, since its performance stands out among the W-ARMA methods [see Section 6 of Ahsan and Dufour (2021)]:

$$\hat{\phi}_p^{\text{ols}} = [A(p, J)' A(p, J)]^{-1} A(p, J)' e(p, J) \tag{3.24}$$

where $e(p, J)$ is a $(pJ) \times 1$ vector and $A(p, J)$ is a $(pJ) \times p$ matrix defined by

$$e(p, J) = [\hat{\gamma}(p, 1)\omega_1^{1/2}, \dots, \hat{\gamma}(p, J)\omega_J^{1/2}]', \quad A(p, J) = [\hat{\boldsymbol{\Gamma}}(p, 1)\omega_1^{1/2}, \dots, \hat{\boldsymbol{\Gamma}}(p, J)\omega_J^{1/2}]'. \tag{3.25}$$

Different OLS-based W-ARMA can be generated by considering different weights $\omega_1, \dots, \omega_J$. In our simulations below as well as in empirical applications, we focus on the case where the weights are equal *i.e.*, $\omega_j = 1/J$ where $j = 1, \dots, J$. For $J = 1$, this estimator is equivalent to the CF-ARMA estimator given in (3.15).

3.4 Recursive estimation for SVL(p) models

In this section, we describe a recursive estimation algorithm tailored for SVL(p) models. For ease of notation, we adopt a different indexing scheme solely for the autoregressive parameters of the volatility process within this section. For instance, the SVL(p) parameters are now represented as $\theta_p^{SVL} := \left(\{\phi_{p,j}\}_{j=1,\dots,p}, \sigma_{pv}, \sigma_y, \delta_p\right)'$.

The recursive estimation of the CF-ARMA estimator capitalizes on the extended Yule-Walker (EYW) equations governing the observed process. When the MA order remains constant, the EYW equations form a nested Toeplitz system. Introducing a *Generalized Durbin-Levinson* algorithm for the CF-ARMA estimator in the SVL(p) model proves beneficial when neither the AR nor the MA order is predetermined. Here, we specifically examine the scenario where $i = p$, meaning the MA order equals p, implying an identical AR order of p.

For $i = 0$, use the Durbin-Levinson algorithm to calculate $\left\{\hat{\phi}_{p,j}^{(0)} \mid p \geq 1, j = 1, \dots, p\right\}$ and for $i \geq 1$, calculate

$$\hat{\phi}_{p,0}^{(i-1)} = -1, \quad \hat{\phi}_{p,j}^{(i)} = \hat{\phi}_{p+1,j}^{(i-1)} - \frac{\hat{\phi}_{p+1,p+1}^{(i-1)}}{\hat{\phi}_{p,p}^{(i-1)}} \hat{\phi}_{p,j-1}^{(i-1)}, \quad p \geq 1, \quad j = 1, \dots, p, \tag{3.26}$$

$$\hat{\sigma}_y = [\exp(\hat{\mu} + 1.2704)]^{1/2}, \quad \hat{\sigma}_{pv} = [\hat{\gamma}_{y^*}(0) - \sum_{j=1}^{p} \hat{\phi}_{p,j} \hat{\gamma}_{y^*}(j) - \pi^2/2]^{1/2}, \tag{3.27}$$

$$\hat{\delta}_p = \frac{\sqrt{2\pi}\hat{\lambda}_y(1)}{\hat{\sigma}_v \hat{\sigma}_y^2} \exp\left(-\frac{1}{4}\hat{\tilde{\gamma}}_p\right), \tag{3.28}$$

where full expressions of $\hat{\delta}_p$ for the lower order SVL(p) models are given in (3.16). This algorithm is the same as the Tsay and Tiao (1984) algorithm [except for the equations involving $\hat{\sigma}_y$ and $\hat{\sigma}_{pv}$] for calculating the extended sample autocorrelation function under the stationarity assumption.

4 Asymptotic distributional theory

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

Assumption 4.1. DISTRIBUTION OF THE ERROR PROCESS. *The error processes z_t and v_t are mutually independent and $\{z_t\}$ is a sequence of i.i.d. real-valued random variables, independent of w_0. The probability distribution of z_t has a continuous density with respect to Lebesgue measure on real line, and its density is positive on $(-\infty, +\infty)$. The transformed error ϵ_t satisfies $\mathbb{E}(|\epsilon_t|)^s < \infty$, where s is an positive integer.*

Assumption 4.2. STATIONARITY OF THE LATENT PROCESS. *The latent process $\{w_t\}$ is strictly stationary with $\mathbb{E}(|w_t|)^s < \infty$ and there is an integer $s \geq 1$ such that $\mathbb{E}(|v_t|)^s < \infty$, $\sum_{j=1}^{\infty} |\bar{\psi}_j|^s < \infty$, where $w_t = \phi^{-1}(B)v_t = \bar{\psi}(B)v_t$ which follows from $\phi(z) \neq 0$ for $|z| \leq 1$, where the characteristic equation of the volatility process $\phi(z) := 1 - \phi_1 z - \cdots - \phi_p z^p = 0$.*

Under Assumptions 4.1 and 4.2 with $s = 2$, the observed process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with exponential β-mixing (see results of Stationarity, Ergodicity and Beta mixing in Ahsan and Dufour (2021)) with finite second moment, *i.e.*, $\mathbb{E}\left[(y_t^*)^2\right] < \infty$. In the following Lemma, using the Ergodic theorem, we prove the consistency of the empirical moments in (3.14).

Lemma 4.1. CONSISTENCY OF EMPIRICAL MOMENTS. *Under the Assumptions 4.1 and 4.2 with $s = 2$, the estimators $\hat{\mu}$, $\hat{\lambda}_y(1)$ and $\hat{\Gamma}(m) := [\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \dots, \hat{\gamma}_{y^*}(m)]'$ defined by (3.14) satisfy:*

$$\hat{\mu} \xrightarrow{p} \mu, \qquad \hat{\lambda}_y(1) \xrightarrow{p} \lambda_y(1) \quad \text{and} \quad \hat{\Gamma}(m) \xrightarrow{p} \Gamma(m) := [\gamma_{y^*}(0), \gamma_{y^*}(1), \dots, \gamma_{y^*}(m)]'. \tag{4.1}$$

Assumptions 4.1 and 4.2 with $s = 4$ are sufficient for the SVL model to have a strictly stationary solution with a finite fourth moment of y_t^*, *i.e.*, $\mathbb{E}\left[(y_t^*)^4\right] < \infty$. Note that the fourth moment of y_t^* translates into the eighth moment of y_t. This solution will be β-mixing with geometrically decreasing mixing coefficients. In the following Lemma, using a Central Limit theorem for the stationary and ergodic process (Lindeberg-Levy theorem for the dependent process), we present the asymptotic distribution of the empirical moments in (3.14).

Lemma 4.2. ASYMPTOTIC DISTRIBUTION OF EMPIRICAL MOMENTS. *Under the Assumptions 4.1 and 4.2 with $s = 4$, the estimators $\hat{\mu}$, $\hat{\lambda}_y(1)$ and $\hat{\Gamma}(m) := [\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1), \dots, \hat{\gamma}_{y^*}(m)]'$ defined by defined by (3.14) satisfy:*

$$\sqrt{T}\begin{bmatrix} \hat{\mu} - \mu \\ \hat{\Gamma}(m) - \Gamma(m) \\ \hat{\lambda}_y(1) - \lambda_y(1) \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} & C_{\mu,\lambda_y(1)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} & C_{\Gamma(m),\lambda_y(1)} \\ C_{\mu,\lambda_y(1)} & C'_{\Gamma(m),\lambda_y(1)} & V_{\lambda_y(1)} \end{bmatrix}\right), \tag{4.2}$$

where

$$V_\mu = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty} \gamma_{y^*}(\tau), \quad V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t) + 2\sum_{\tau=1}^{\infty} \mathrm{cov}(\Lambda_t, \Lambda_{t+\tau}), \quad V_{\lambda_y(1)} = \lambda_y^2(1), \tag{4.3}$$

$$C_{\mu,\Gamma(m)} = (\bar{c},\ 0_{[1\times m]})', \quad \bar{c} := C_{\mu,\Gamma(0)} = 2\sum_{t=1}^{\infty} \mathbb{E}[y_t^{*3}] = 2\sum_{t=1}^{\infty} \mathbb{E}[\epsilon_t^3], \quad \Lambda_t := [\Lambda_{t,0}, \Lambda_{t,1}, \dots, \Lambda_{t,m}]', \tag{4.4}$$

$$\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k) = [\log(y_t^2) - \mu][\log(y_{t+k}^2) - \mu] - \gamma_{y^*}(k), \quad k = 0, \dots, m, \tag{4.5}$$

$$C_{\mu,\lambda_y(1)} = 2\sum_{t=1}^{\infty} \frac{\delta_p \sigma_v \sigma_y^2}{\sqrt{2\pi}} \exp\left(\frac{\gamma_w(0)}{4}\right)\left[8 - \left(\gamma^{\epsilon} + \psi^{(0)}(0.5)\right)\exp\left(\frac{\gamma_w(1)}{4}\right)\right], \tag{4.6}$$

$$C_{\Gamma(m),\lambda_y(1)} = [C_{\lambda 0}, C_{\lambda 1}, \dots, C_{\lambda m}]', \tag{4.7}$$

$$\begin{aligned} C_{\lambda k} \quad &= \quad 2\sum_{t=1}^{\infty}\left(\frac{\delta_p \sigma_v \sigma_y^2}{\sqrt{2\pi}}\exp\left(\frac{\gamma_w(0)}{4}\right)\left(\gamma_w(0) + \bar{\psi}_k \gamma_w(1) + 0.5\bar{\psi}_k \gamma_w(k) + \gamma_w(k+1) - \frac{\gamma^{\epsilon} + \psi^{(0)}(0.5)}{2}\right)\right), \\ & \quad\;\; k = 0, 1, 2, \dots, m. \end{aligned} \tag{4.8}$$

where $\bar{\psi}_k$ is k-th parameter of the MA(∞) representation of the latent log volatility process, $\gamma_w(k) = \mathrm{cov}(w_t, w_{t+k})$, $\psi^{(0)}(z)$ is the digamma function *and γ^{ϵ} is the Euler-Mascheroni constant.*

This in turn yields the asymptotic distribution of the simple ARMA-type estimator $(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v, \delta_p)'$.

Theorem 4.3. ASYMPTOTIC DISTRIBUTION OF SIMPLE ARMA-SV ESTIMATOR. *Under the Assumptions 4.1 and 4.2 with $s = 4$, the estimator $\hat{\theta} := (\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\sigma}_y, \hat{\sigma}_v, \hat{\delta}_p)'$ given in (3.15) is consistent,* i.e., *$\hat{\theta} \xrightarrow{p} \theta$, and*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, V), \tag{4.9}$$

where $\theta := (\phi_1, \dots, \phi_p, \sigma_y, \sigma_v, \delta_p)'$ and

$$V = G(\beta)\begin{bmatrix} V_{\mu} & C'_{\mu,\Gamma(2p)} & C_{\mu,\lambda_y(1)} \\ C_{\mu,\Gamma(2p)} & V_{\Gamma(2p)} & C_{\Gamma(2p),\lambda_y(1)} \\ C_{\mu,\lambda_y(1)} & C'_{\Gamma(2p),\lambda_y(1)} & V_{\lambda_y(1)} \end{bmatrix} G(\beta)', \tag{4.10}$$

$$G(\beta) := \frac{\partial D_p}{\partial \beta'}, \quad D_p := D_p(\beta) = (D_{\phi_{\boldsymbol{p}}}, D_{\sigma_y}, D_{\sigma_v}, D_{\delta_p})', \quad \beta := [\mu, \gamma_{y^*}(0), \gamma_{y^*}(1), \dots, \gamma_{y^*}(2p), \lambda_y(1)]', \tag{4.11}$$

$$D_{\phi_{\boldsymbol{p}}} := \boldsymbol{\Gamma}(p,1)^{-1}\gamma(p,1), \quad D_{\sigma_y} := \exp(\mu + 1.2704)^{1/2}, \quad D_{\sigma_v} := [\gamma_{y^*}(0) - \phi'_{\boldsymbol{p}}\gamma(1) - \pi^2/2]^{1/2}, \tag{4.12}$$

$$D_{\delta_{\boldsymbol{p}}} = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2}\exp\left(-\frac{1}{4}\tilde{\gamma}_p\right), \quad \lambda_y(1) := \mathbb{E}[|y_t| y_{t-1}], \tag{4.13}$$

$$\phi_p = (\phi_1, \dots, \phi_p)', \quad \gamma(p,1) = [\gamma_{y^*}(p+1), \dots, \gamma_{y^*}(2p)]', \gamma(p) = [\gamma_{y^*}(1), \dots, \gamma_{y^*}(p)]', \tag{4.14}$$

$$\boldsymbol{\Gamma}(p,1)_{[p\times p]} = \begin{bmatrix} \gamma_{y^*}(p) & \gamma_{y^*}(p-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(p+1) & \gamma_{y^*}(p) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p-1) & \gamma_{y^*}(2p-2) & \cdots & \gamma_{y^*}(p) \end{bmatrix}, \tag{4.15}$$

$$\gamma_{y^*}(k) = \mathrm{cov}(y_t^*, y_{t-k}^*), \quad y_t^* = (\log y_t^2 - \mu), \quad \mu := \mathbb{E}[\log(y_t^2)], \quad \tilde{\gamma}_p := \mathrm{var}(w_t) + \mathrm{cov}(w_t, w_{t-1}). \tag{4.16}$$

The explicit form of the analytical moment derivative, $G(\beta)$, is given in the proof. An estimator of the covariance matrix V can be obtained by using heteroskedasticity and autocorrelation consistent (HAC) covariance estimators [see Den Haan and Levin (1997) and Robinson and Velasco (1997)] and then substituting $\hat{\beta} = [\hat{\mu},\ \hat{\gamma}_{y^*}(0),\ \hat{\gamma}_{y^*}(1),\ \hat{\gamma}_{y^*}(2), \dots,\ \hat{\gamma}_{y^*}(2p),\ \hat{\lambda}_y(1)]'$ into $G(\beta)$. In our empirical applications, we use a Bartlett kernel estimator with the bandwidth varying with the sample size; see Newey and West (1994). One can alternatively use the analytic expressions of $\gamma_{y^*}(k)$ to obtain an estimator of V_μ. The ARMA-type estimator can be viewed as a GMM-type estimator, so one can also use GMM standard errors.

Theorem 4.3 covers the simplest CF-ARMA estimator. It is easy to see that the asymptotic distribution of more general winsorized estimators can be derived in the same way upon using Lemmas 4.1 - 4.2.

5 Forecasting with SVL(p) models

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

$w_{t+1} = \sum_{j=1}^{p} \phi_j w_{t-j+1} + \tilde{\delta}_p z_t + \tilde{\sigma}_v \bar{v}_{t+1}$,

$$y_t^* = w_t + \epsilon_t, \quad w_{t+1} = \sum_{j=1}^{p} \phi_j w_{t-j+1} + \tilde{\delta}_p z_t + \tilde{\sigma}_v \bar{v}_{t+1}, \tag{5.1}$$

where $\tilde{\delta} = \delta\sigma_v$, $\tilde{\sigma}_v = \sqrt{(1-\delta^2)}\sigma_v$, $\bar{v}_t$'s are i.i.d. $\mathcal{N}(0,1)$, and ϵ_t's are approximated by a normal distribution with mean 0 and variance $\pi^2/2$. Using similar notation as that found in Hamilton (1994), this model can also be rewritten as follows:

$$y_t^* = H'\xi_t + \epsilon_t, \quad \xi_{t+1} = F\xi_t + \sigma_v u_{t+1}, \tag{5.2}$$

$$u_{t+1} = \delta\eta_t + (1-\delta^2)^{1/2}\zeta_{t+1}, E[u_t u_t'] = E[z_t z_t'] = E[\tilde{v}_t \tilde{v}_t'] = \Omega_1, \tag{5.3}$$

where $H' = [1, 0, \ldots, 0]$ is a $1 \times p$ vector, u and

$$\xi_t = \begin{bmatrix} w_t \\ w_{t-1} \\ w_{t-2} \\ \vdots \\ w_{t-p+1} \end{bmatrix}, F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \eta_t = \begin{bmatrix} z_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \zeta_t = \begin{bmatrix} \tilde{v}_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, u_{t+1} = \begin{bmatrix} v_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Omega_1 = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix},$$

where F and Ω_1 are $p \times p$ matrices, and ξ_t, u_t, η_t, ζ_t are $p \times 1$ vectors. sing (5.2), the Kalman filter can be applied as follows.

- Initialization:

$$\begin{aligned} \hat{\xi}_{1|0} &= E[\xi_1] = \mathbf{0}_{(p\times 1)}, \\ \boldsymbol{P}_{1|0} &= E[(\xi_1 - E[\xi_1])(\xi_1 - E[\xi_1])'] = \mathrm{diag}[\sigma_v^2, \ldots, \sigma_v^2]_{(p\times p)}, \end{aligned} \tag{5.4}$$

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

- Sequential updating:

$$\begin{aligned} K_t = \boldsymbol{P}_{t|t-1} H \left(H' \boldsymbol{P}_{t|t-1} H + \frac{\pi^2}{2} \right)^{-1}, \quad \hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + K_t \left(y_t^* - H'\hat{\xi}_{t|t-1} \right), \\ \boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - K_t H' P_{t|t-1}. \end{aligned} \tag{5.5}$$

- In-sample prediction:

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t} + \tilde{\delta}\eta_{t|t}, \quad P_{t+1|t} = FP_{t|t}F_v'^2\Omega_1, \quad \hat{y}_{t+1|t} = H'\hat{\xi}_{t+1|t},$$
$$E[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})'] = H'\boldsymbol{P}_{t+1|t}H + \frac{\pi^2}{2}, \tag{5.6}$$

  where the last two lines follow from (5.2).

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

$$\hat{\xi}_{T+h|T} = F^h\hat{\xi}_{T|T} + F^{(h-1)}\tilde{\delta}\eta_{T|T}, \quad \hat{y}^*_{T+h|T} = H'\hat{\xi}_{T+h|T}. \tag{5.7}$$

Hence, the h-step ahead forecast is computed using (5.7) and the simple estimates. Note that here, the leverage parameter δ appears in prediction equation given in the first line of (5.6) and (5.7). When there is no leverage, this value will simply be zero and as a result, in sample predictions will also differ. It can also be seen from the first line in (5.7) that as h becomes larger, the effect from leverage will decay. Hence, we may expect that forecasts from a model with leverage and that from a model without leverage may not differ too much when forecasting very far out of sample.

6 Hypothesis testing

In this section, we discuss how to test a hypothesis on an SVL(p) model. First, we discuss asymptotic tests based on LR-type test statistics. Second, we show how to construct finite-sample tests using the Monte Carlo test technique. Finally, we show how we can derive implicit standard errors (ISE) for the parameters of our model from simulation-based confidence intervals.

6.1 Asymptotic tests

The SVL(p) model has three parameters in addition to the p autoregressive parameters (*i.e.*, a total of $p+3$ parameters). These are given by $\theta = (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v, \delta_p)'$. To test the values of individual parameters, we can consider t-type statistics of the form:

$$T(\theta_i) = (\hat{\theta}_i - \theta_1)/SE(\hat{\theta}_i) \tag{6.1}$$

where the standard error $SE(\hat{\theta}_i)$ is calculated from the asymptotic covariance matrix given in (4.9). Alternatively, we can also consider GMM-based LR-type statistics, based on the following moment-based objective function:

$$M_T(\theta) := g_T(\theta)' A_T\, g_T(\theta) \tag{6.2}$$

where $\theta := (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v, \delta_p)'$, $g_T(\theta)$ is $(p+3) \times 1$ vector of moment functions, defined as

$$g_T(\theta) = \begin{bmatrix} \hat{\mu} + 1.2704 - \log(\sigma_y^2) \\ \hat{\gamma}_{y^*}(0) + \hat{\gamma}_{y^*}(1) - (\pi^2/2) - (1-\phi_1)^{-1}\left(\sum_{j=2}^{p}\left[\phi_j(\hat{\gamma}_{y^*}(j-1) + \hat{\gamma}_{y^*}(j)\right] - \sigma_v^2\right) \\ \hat{\gamma}_{y^*}(p+1) - \phi_1\hat{\gamma}_{y^*}(p) + \cdots + \phi_p\hat{\gamma}_{y^*}(1) \\ \vdots \\ \hat{\gamma}_{y^*}(2p) - \phi_1\hat{\gamma}_{y^*}(2p-1) + \cdots + \phi_p\hat{\gamma}_{y^*}(p) \\ \delta - \frac{\sqrt{2\pi}\hat{\lambda}_{y(1)}}{\sigma_v\sigma_y^2}\exp\left(-\frac{1}{4}\tilde{\gamma}_p\right) \end{bmatrix} \tag{6.3}$$

and A_T is an appropriate weighting matrix. $M_T(\theta)$ is (up to an asymptotically negligible term) a GMM objective function.

The first moment function follows from first part of (3.6) after taking logarithm and the second moment condition follows from the Corollary of Autocovariances of SV Process on adding the equations for $k = 0$ and $k = 1$ [in (3.3)]: this yields

$$\hat{\gamma}_{y^*}(0) + \hat{\gamma}_{y^*}(1) - (\pi^2/2) - (1-\phi_1)^{-1}\left(\sum_{j=2}^{p}\left[\phi_j(\hat{\gamma}_{y^*}(j-1) + \hat{\gamma}_{y^*}(j)\right] - \sigma_v^2\right) = 0. \tag{6.4}$$

The preceding p moment conditions prior to the final one are corresponds to (3.3) with $k = p+1, \ldots, 2p$ and finally the fourth equation comes from (3.7) where $\tilde{\gamma}_p$ is a function of $(\phi_1, \ldots, \phi_p, \sigma_v)$.

Since the number of moment functions in (6.3) is equal to the number of parameters, we take $A_T = I_{(p+3)}$ and consider the GMM-type objective function

$$M_T^*(\theta) = g_T(\theta)' g_T(\theta). \tag{6.5}$$

To test hypotheses on θ, the LR-type statistic is the difference between the restricted and unrestricted optimal values of the objective function:

$$LR_T = T[M_T^*(\hat{\theta}_0) - M_T^*(\hat{\theta})] \tag{6.6}$$

where $\hat{\theta}$ is the unrestricted estimator and $\hat{\theta}_0$ is the constrained estimator under the null hypothesis. Under standard regularity conditions, the asymptotic distribution of LR_T is χ_r^2 where r is the number of constraints; see Newey and West (1987), Newey and McFadden (1994), Dufour et al. (2017). Note however that usual regularity conditions may not be satisfied when some parameters are not identified or the null hypothesis involves the frontier of the parameter space.

6.2 Simulation-based finite-sample tests

In this section, we discuss simulation-based inference procedures for SVL(p) models with leverage. The simulation-based methods are more attainable in the context of this study for two reasons. First, the SVL(p) model is a parametric model with a finite number of parameters, and we can effortlessly simulate this model. Second, we can simulate the test statistic of SVL(p) parameters, which is based on a computationally inexpensive estimator. So, using our proposed computationally simple estimators, one can easily construct more reliable finite-sample inference.

It should be noted that the simulation-based procedure may not be attainable when the estimator is computationally expensive, so that we cannot simulate the test statistic easily.

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

Let us first consider the case of pivotal statistics, *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 for testing a null hypothesis H_0.

1. Compute the observed test statistic S_0 from the available data.

2. Generate by Monte Carlo methods a vector $S(N) = (S_1, \dots, S_N)$ of N i.i.d. replications of S under H_0.

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

$$p_N[x, S(N)] := \frac{NG_N[x; S(N)] + 1}{N+1}, \tag{6.7}$$

$$G_N[x; S(N)] := \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty)}(S_i - x), \quad I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \in [0,\infty), \\ 0 & \text{if } x \notin [0,\infty). \end{cases} \tag{6.8}$$

   In other words,

$$p_N[S_0; S(N)] = \frac{NG_N[S_0; S(N)] + 1}{N+1} \tag{6.9}$$

   where $NG_N[S_0; S(N)]$ is the number of simulated values greater than or equal to S_0. When $S_0, S_1, \dots, S_N$ are all distinct [an event with probability one when the vector $(S_0, S_1, \dots, S_N)'$ 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 $S_0, S_1, \dots, S_N$.

4. The MC critical region for a test of level α $(0 < \alpha < 1)$ is

$$\hat{p}_N[S] \le \alpha. \tag{6.10}$$

   If $\alpha(N+1)$ is an integer and the distribution of S is continuous under the null hypothesis, then under H_0,

$$P[\hat{p}_N[S] \le \alpha] = \alpha; \tag{6.11}$$

   see Dufour (2006).

Consider now 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
$$H_0 : \bar{\theta} \in \Omega_0 \tag{6.12}$$
where $\emptyset \neq \Omega_0 \subset \Omega$, compute the observed test statistic S_0 from the available data.

2. For each $\theta \in \Omega_0$, we can generate by Monte Carlo methods a vector $S(N,\theta) = [(S_1(\theta), \ldots, S_N(\theta)]$ of N i.i.d. replications of S.

3. The simulated test statistics define the MC p-value function $\hat{p}_N[S|\theta] := p_N[S_0; S(N,\theta)]$ where
$$p_N[x; S(N,\theta)] := \frac{NG_N[x; S(N,\theta)] + 1}{N+1}. \tag{6.13}$$

4. The p-value function $\hat{p}_N[S\,|\,\theta]$ as a function of θ is maximized over the parameter values compatible with Ω_0, *i.e.*, under the null hypothesis, and H_0 is rejected if N
$$\sup\{\hat{p}_N[S|\theta] : \theta \in \Omega_0\} \leq \alpha. \tag{6.14}$$
If the number N of simulated statistics is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:
$$P_{\bar{\theta}}[\sup\{\hat{p}_N[S|\theta] : \theta \in \Omega_0\} \leq \alpha] \leq \alpha. \tag{6.15}$$
Consequently the critical region in (6.14) has *level* α for testing H_0; for a proof, see Dufour (2006).

Because of the maximization in the critical region (6.14), the above test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as 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 algorithms and quasi-Newton methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, *e.g.* simulated annealing or Particle Swarm Optimization.

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. Compute S_{T0} the observed test statistic (based on data). By assumption, the distribution of S involves nuisance parameters under the null hypothesis H_0 in (6.12).

2. We suppose we have a consistent set estimator C_T of $\bar{\theta}$ (under H_0), *i.e.* C_T satisfies
$$\lim_{T\to\infty} P_{\bar{\theta}}[\bar{\theta} \in C_T] = 1 \text{ under } H_0. \tag{6.16}$$

3. For each $\theta \in \Omega_0$, we can generate by Monte Carlo methods a vector $S(N,\theta) = [(S_1(\theta), \ldots, S_N(\theta)]$ of N i.i.d. replications of S.

4. he simulated test statistics define the MC p-value function $\hat{p}_{TN}[S_T \mid \theta] := p_{TN}[S_{T0}; S_T(N,\theta)]$, where

$$p_{TN}[x; S_T(N,\theta)] := \frac{NG_{TN}[x; S_T(N,\theta)] + 1}{N+1}. \tag{6.17}$$

5. The p-value function $\hat{p}_{TN}[S_T \mid \theta]$ as a function of θ is maximized with respect to θ in C_T, and H_0 is rejected if

$$\sup\{\hat{p}_{TN}[S_T \mid \theta] : \theta \in C_T\} \leq \alpha. \tag{6.18}$$

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

$$\lim_{T\to\infty} P_{\bar{\theta}}[\sup\{\hat{p}_{TN}[S_T \mid \theta] : \theta \in C_T\} \leq \alpha] \leq \alpha\,. \tag{6.19}$$

   The critical region in (6.18) has level α asymptotically.

In practice, it is easy to find a consistent set estimate of $\bar{\theta}$, whenever a *consistent* point estimate $\hat{\theta}_T$ of $\bar{\theta}$ is available (*e.g.*, a GMM estimator). For instance, any set of the form

$$C_T = \{\theta \in \Omega_0 : \left\|\hat{\theta}_T - \theta\right\| < \varepsilon\} \tag{6.20}$$

with $\varepsilon > 0$ a fixed positive constant independent of T, satisfies (6.16). 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 which 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).

The test based on simulations using a point nuisance parameter estimate is called a *local Monte Carlo* (LMC) test. The term local reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. If the set C_T in (6.18) is reduced to a single point estimate $\hat{\theta}_T$, *i.e.* $C_T = \{\hat{\theta}_T\}$, we get a LMC test

$$\hat{p}_{TN}[S_T \mid \hat{\theta}_T] \leq \alpha \tag{6.21}$$

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, a fundamental difference between the latter procedure and the parametric bootstrap method.

Even if $\hat{\theta}_T$ is a consistent estimate of $\bar{\theta}$ (under the null hypothesis), the set $C_T = \{\hat{\theta}_T\}$ does not generally satisfy condition (6.16). 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 MMC test is clearly not significant at level α.

Since Ahsan and Dufour (2021) consider testing individual parameters of a higher order SV(p) model using t-type statistics briefly described above, we will instead focus on testing for no leverage using the LR-type statistic discussed previously. Consequently, in the description of the simulation-based finite-sample tests above, we set $S_0 = LR_0$ and $S(N,\theta) = LR(N,\theta)$, where $LR(N,\theta) = [LR_1(\theta),\ldots,LR_N(\theta)]$. This testing procedure requires estimating the model under both the null hypothesis and the alternative hypothesis using a computationally inexpensive estimator; see Section 3 for computationally simple methods for SVL(p) models. Furthermore, the constrained model utilizes the consistent point estimate $\hat{\theta}_T$. We discuss in more detail the test for no leverage in the next section, which also includes other simulation-based results.

6.3 Implicit standard error

In this subsection, we show that *implicit standard errors* (ISE) for the components of a parameter vector $\theta = (\theta_1,\ldots,\theta_m)'$ can be derived from simulation-based confidence intervals. The asymptotic standard error proposed in Section 4 can be markedly different and may be quite unreliable in finite samples. To construct a more reliable standard error, we derive the ISE in the following way.

1. Calculate the (typically restricted) estimate $\hat{\theta}_0 = (\hat{\theta}_{10},\ldots,\hat{\theta}_{m0})'$ from observed data ($\mathbb{Y}_0$).
2. Using $\hat{\theta}_0$ as parameter value, generate N i.i.d. replications $\mathbb{Y}(N) = (\mathbb{Y}_1,\ldots,\mathbb{Y}_N)$ of $\mathbb{Y}$, by Monte Carlo methods.
3. From $\mathbb{Y}(N) = (\mathbb{Y}_1,\ldots,\mathbb{Y}_N)$, compute the corresponding parameter estimates.
4. For each component θ_i of θ, the confidence interval $[C_i(\alpha_L), C_i(\alpha_H)]$, with coverage $\alpha = \alpha_L - \alpha_H$, is constructed using the empirical α_{iL} quantile and the empirical α_{iH} quantile of $\hat{\theta}_i(N) = (\hat{\theta}_{i0},\hat{\theta}_{i1},\ldots,\hat{\theta}_{iN})$.
5. By analogy with usual Gaussian-based confidence intervals, we set $C_i(\alpha_L) = \hat{\theta}_{i0} - z(\alpha/2)\,\hat{\sigma}_{iL}$ and $C_i(\alpha_H) = \hat{\theta}_{i0} + z(\alpha/2)\,\hat{\sigma}_{iH}$, where $z(\alpha/2)$ satisfies $P[Z \geq z(\alpha/2)] = \alpha/2$ and $Z \sim \mathrm{N}(0,1)$. This suggests that two numbers could play the role of "standard errors" here:
$$\hat{\sigma}_{iL} = \frac{\hat{\theta}_{i0} - C_i(\alpha_L)}{z(\alpha/2)} := ISE_{iL}\,, \quad \hat{\sigma}_{iU} = \frac{C_i(\alpha_H) - \hat{\theta}_{i0}}{z(\alpha/2)} := ISE_{iH}\,. \tag{6.22}$$
6. Finally, a conservative ISE for θ_i is given by $\min\{ISE_{iL},\ ISE_{iH}\}$ and a liberal ISE is given by the average or $\max\{ISE_{iL},\ ISE_{iH}\}$.

Table 1. Comparison with competing estimators: Bias and RMSE

				Bias				RMSE			
				ϕ	σ_y	σ_v	δ	ϕ	σ_y	σ	δ
				True Value							
T	Estimators	RCT	NIV	0.95	0.15	1	-0.95	0.95	0.15	1	-0.95
500	Bayes	31929.7	0	-0.0568	0.0830	0.1343	0.6639	0.0617	0.1108	0.1530	0.6653
	QML	1028.8	0	-0.0767	0.0176	-0.8469	0.3481	0.2208	0.0810	1.2603	0.4151
	CF-ARMA	1.0	61	-0.0142	0.0160	-0.0146	0.1542	0.0396	0.0777	0.2822	0.3861
	W-ARMA ($J=10$)	1.7	0	-0.0137	**0.0155**	**0.0155**	0.1609	**0.0268**	**0.0768**	**0.1231**	0.3880
	W-ARMA ($J=100$)	2.6	0	**-0.0064**	**0.0155**	-0.0466	**0.1521**	0.0236	**0.0768**	0.1412	**0.3818**
2000	Bayes	218068.0	0	-0.0155	0.0346	-0.0151	0.1588	0.1070	0.0573	0.0621	0.1663
	QML	1025.5	0	-0.0482	**0.0022**	-0.4264	0.3273	0.1513	0.0363	0.8612	0.3855
	CF-ARMA	1.0	3	-0.0031	0.0030	-0.0238	-0.0313	0.0188	0.0356	0.1678	0.1173
	W-ARMA ($J=10$)	1.4	0	-0.0038	0.0031	**-0.0039**	**-0.0315**	**0.0105**	**0.0356**	**0.0617**	0.1172
	W-ARMA ($J=100$)	1.8	0	**-0.0017**	0.0031	-0.0233	-0.0322	0.0106	**0.0356**	0.0781	**0.1158**

Notes: We simulate 1000 samples from each model. W-ARMA ($J=10,100$) is the winsorized ARMA estimator based on OLS and J is the winsorizing parameter. These methods are proposed in Section 3. QML is the quasi-maximum likelihood estimator of Harvey and Shephard (1996). We used R package *stochvol* of Kastner (2016) for the Bayesian estimation based on Markov Chain Monte Carlo methods, where the posteriors are based on 50000 draws of the sampler, after discarding 50000 draws. RCT stands for the relative computational time w.r.t. the CF-ARMA estimator. The number of inadmissible values (NIV) of ϕ is also reported, these are out of 1000. Boldface font highlights the smallest bias and RMSE with no NIV. Boldface font also highlights the estimator, which has the best overall performance.

7 Simulation study

In this section, we evaluate the properties of our proposed estimators through simulation, focusing specifically on bias and root mean square error (RMSE). Additionally, we present simulation findings regarding the finite-sample properties of the LR-type test discussed in Section 6. Moreover, through simulations, we illustrate the significance of leverage in the context of volatility forecasting.

7.1 Estimation

Now, we assess the statistical performance of our proposed estimators. These estimators include the closed-form ARMA (CF-ARMA) estimator, as well as two winsorized ARMA (W-ARMA) estimators with no intercept regression, where $J=10$ and $J=100$, respectively, for SVL(p) models. Overall, there is no consistent ranking among the different estimators. However, the Bayesian estimator consistently outperforms other methods in the context of an SV(1) model; refer to Jacquier et al. (1994) for details. Consequently, we compare our proposed estimators with both QML [see Harvey and Shephard (1996)] and Bayesian [see Jacquier et al. (1994), Kim et al. (1998)] estimators.

We consider an SVL(1) model with parameter values set at $(\phi,\sigma_y,\sigma_v,\delta)=(0.95,0.15,1,-0.95)$. These parameter choices are typical in empirical studies focusing on hourly or daily returns. We conduct simulations with 1000 replications and present results for two distinct sample sizes ($T=500,2000$).

Bayesian estimates based on Markov Chain Monte Carlo methods are computed using the R package *stochvol* [see Kastner (2016)], where the SVL(1) model is reparametrized as $y_t=\exp(w_t/2)z_t$ and $w_t=\mu_w+\phi(w_{t-1}-\mu_w)+\sigma_v v_t$, where $\sigma_y=\exp(\mu_w/2)$. Regarding this reparametrization, see also Kim et al. (1998). In Kastner (2016), independent priors are considered for μ_w, ϕ, σ_v and δ, where $\mu_w\sim\mathcal{N}\left(\mu_{w,0},\sigma^2_{w,0}\right)$, $\sigma_v\in\mathbb{R}^+$ with $\pm\sqrt{\sigma_v^2}\sim\mathcal{N}\left(0,B_{\sigma_v}\right)$, $\phi\in(-1,1)$

Table 2. Empirical size of tests for no leverage in SVL(1) model (moderate persistence)

H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$						
	$\phi = 0.90$, $\sigma_y = 0.10$, $\sigma_v = 0.75$			$\phi = 0.75$, $\sigma_y = 0.10$, $\sigma_v = 1.00$		
T	Asy	LMC	MMC	Asy	LMC	MMC
500	79.3	12.6	3.3	71.1	9.9	1.5
1000	82.8	12.8	2.0	72.8	10.1	1.4
2000	85.0	11.4	1.8	76.7	11.4	1.0
5000	89.0	10.2	1.0	77.9	10.8	1.4

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N = 99$ simulations.

with $(\phi+1)/2 \sim \mathcal{B}\left(\alpha_{(\phi,0)}, \beta_{(\phi,0)}\right)$, and $\delta \in (-1,1)$ with $(\delta+1)/2 \sim \mathcal{B}\left(\alpha_{(\delta,0)}, \beta_{(\delta,0)}\right)$. Here, $\mathcal{B}$ denotes the beta distribution, and α and β represent positive hyperparameters. The latter are set to: $\mu_{w,0} = 0$, $\sigma_{w,0} = 1$, $B_{\sigma_v} = 1$, $\alpha_{(\phi,0)} = 5$, $\beta_{(\phi,0)} = 15.5$, $\alpha_{(\delta,0)} = 75$, and $\beta_{(\delta,0)} = 1$. We apply this Bayesian method on simulated datasets and draw 100000 observations from the posterior distributions. Bayesian estimators are obtained by averaging the last $50,000$ draws.

In our simulations, we find the CF-ARMA estimator of ϕ occasionally produced values outside the stationary region. These samples were discarded from calculation. Therefore, the bias and RMSE obtained are thus conditional on the non-occurrence of inadmissible values. Such problems are practically non-existent with the OLS-based W-ARMA estimator. Table 1 presents the estimation outcomes, revealing that the W-ARMA estimator consistently exhibits superior performance in bias and RMSE across all parameter values and sample sizes. Specifically, the winsorized ARMA estimator with $J = 10$ consistently outperforms all other estimators, including QML and the Bayesian estimator, demonstrating significantly lower bias and RMSE. Moreover, winsorized estimators are notably more time-efficient, as evidenced by the considerable margin of time savings compared to alternative methods. These findings offer valuable insights into the relative performance of various estimation techniques for SVL(1) models.

7.2 Testing

Here, we are interested in testing for no leverage and hence consider the following hypothesis

$$H_0 : \delta = 0 \quad vs \quad H_1 : \delta \neq 0 \tag{7.1}$$

To test the hypothesis of no leverage, we utilize the vector of moment conditions given in equation (6.3) to derive $g_T(\theta_1)$, which remains consistent with the unconstrained model. We then simply set $\delta = 0$ in the final moment condition to derive $g_T(\theta_0)$. It is worth noting that all moment conditions, except the last one, are independent of δ and remain identical under both the null and alternative hypotheses.

The unrestricted and restricted objective functions are denoted by $M^T(\theta_1) = g_T(\theta_1)' g_T(\theta_1)$ and $M^T(\theta_0) = g_T(\theta_0)' g_T(\theta_0)$ respectively. Here, the complete parameter vector under H_1 is represented by $\theta = (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v, \delta)'$, while under H_0, the vector of nuisance parameters reduces to $\tilde{\theta} = (\phi_1, \ldots, \phi_p, \sigma_y, \sigma_v)'$. Hence, within the framework of LMC tests, the values $\hat{\tilde{\theta}}_0 = (\hat{\phi}_1^0, \ldots, \hat{\phi}_p^0, \hat{\sigma}_y^0, \hat{\sigma}_v^0)'$, obtained from estimating the restricted model using observed data, are used to simulate the null distribution of the test statistic.

Table 3. Empirical power of tests for no leverage in SVL(1) model (moderate persistence)

	H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$											
	$\delta = -0.90$			$\delta = -0.70$			$\delta = 0.70$			$\delta = 0.90$		
T	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC
	$\phi = 0.90$, $\sigma_y = 0.10$, $\sigma_v = 0.75$											
500	61.9	65.5	20.9	42.3	45.0	12.5	42.5	43.9	11.4	63.8	66.6	19.5
1000	84.9	81.9	36.8	59.8	57.1	19.7	61.9	58.2	18.7	85.6	82.7	35.4
2000	99.1	99.3	58.7	85.6	87.3	31.7	85.6	87.1	31.9	98.7	99.0	60.3
5000	100.0	100.0	94.9	99.8	99.9	70.6	99.9	99.9	68.4	100.0	100.0	93.9
	$\phi = 0.75$, $\sigma_y = 0.10$, $\sigma_v = 1.00$											
500	96.2	95.9	62.3	80.8	80.2	35.8	80.8	80.3	35.2	96.5	96.4	63.6
1000	100.0	100.0	93.3	99.0	98.7	70.5	98.8	98.5	72.8	100.0	100.0	92.9
2000	100.0	100.0	99.7	100.0	100.0	94.5	100.0	100.0	94.4	100.0	100.0	99.9
5000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N = 99$ simulations. Asymmetric test results are level-corrected using $N = 10000$ and true parameters (*i.e.*, infeasible test in practice as it requires knowing the true DGP). LMC test results are level-corrected using $N = 1{,}000$ and true parameters.

Table 4. Empirical size of tests for no leverage in SVL(1) model (high persistence)

	H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$					
	$\phi = 0.99$, $\sigma_y = 0.10$, $\sigma_v = 0.25$			$\phi = 0.95$, $\sigma_y = 0.10$, $\sigma_v = 0.50$		
T	Asy	LMC	MMC	Asy	LMC	MMC
500	81.8	19.5	5.3	79.7	13.6	2.5
1000	88.7	17.1	5.2	83.3	13.4	2.3
2000	92.1	16.8	5.4	87.9	11.4	1.6
5000	93.2	13.3	4.8	91.2	10.9	1.4

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N = 99$ simulations.

Table 5. Empirical power of tests for no leverage in SVL(1) model (high persistence)

	H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$											
	$\delta = -0.90$			$\delta = -0.70$			$\delta = 0.70$			$\delta = 0.90$		
T	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC
	$\phi = 0.99$, $\sigma_y = 0.10$, $\sigma_v = 0.25$											
500	23.4	24.1	8.4	19.4	19.8	7.9	22.8	24.1	6.2	25.6	27.0	8.1
1000	34.1	33.4	12.4	26.6	26.4	10.4	23.2	23.1	8.6	29.7	29.6	9.2
2000	43.7	42.7	16.5	30.2	28.7	12.0	30.7	29.7	12.4	44.3	42.4	17.9
5000	68.8	67.2	31.7	46.9	46.1	20.0	44.5	43.8	18.1	66.0	65.0	27.5
	$\phi = 0.95$, $\sigma_y = 0.10$, $\sigma_v = 0.50$											
500	42.9	40.6	12.8	29.6	27.9	8.0	30.9	28.6	8.1	44.9	41.5	13.2
1000	67.9	67.1	26.2	46.3	45.2	14.6	44.4	43.4	13.0	67.1	65.8	24.1
2000	87.6	87.3	40.1	62.7	62.3	20.3	65.7	65.6	22.7	89.8	89.4	43.9
5000	100.0	100.0	82.3	96.4	96.4	48.0	96.0	96.0	43.7	100.0	100.0	79.9

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N = 99$ simulations. Asymmetric test results are level-corrected using $N = 10000$ and true parameters (*i.e.*, infeasible test in practice as it requires knowing the true DGP). LMC test results are level-corrected using $N = 1{,}000$ and true parameters.

Table 6. Empirical size of tests for no leverage in SVL(2) model

H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$						
	$\phi_1 = 0.05$, $\phi_2 = 0.85$, $\sigma_y = 1.00$, $\sigma_v = 1.00$			$\phi_1 = 0.05$, $\phi_2 = 0.70$, $\sigma_y = 1.00$, $\sigma_v = 1.00$		
T	Asy	LMC	MMC	Asy	LMC	MMC
500	57.9	13.8	6.7	53.1	8.0	2.7
1000	62.8	10.7	3.9	54.3	6.3	1.7
2000	68.0	8.8	2.4	55.8	7.1	1.5
5000	77.0	8.4	1.9	57.1	6.8	2.5

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N = 99$ simulations.

Table 7. Empirical power of tests for no leverage in SVL(2) model

H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$												
	$\delta = -0.90$			$\delta = -0.70$			$\delta = 0.70$			$\delta = 0.90$		
T	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC
$\phi_1 = 0.05$, $\phi_2 = 0.85$, $\sigma_y = 1.00$, $\sigma_v = 1.00$												
500	62.4	67.5	47.0	49.4	56.9	34.1	50.0	55.2	34.3	62.1	65.8	45.2
1000	74.3	75.3	52.8	60.1	63.0	39.5	61.5	65.8	40.9	75.5	76.2	53.8
2000	90.3	88.1	67.4	80.9	80.3	52.5	80.9	80.2	49.9	91.0	87.4	66.3
5000	99.8	98.7	88.6	99.1	96.2	77.4	99.4	95.9	78.8	100.0	98.8	87.8
$\phi_1 = 0.05$, $\phi_2 = 0.70$, $\sigma_y = 1.00$, $\sigma_v = 1.00$												
500	97.7	96.7	87.7	90.5	92.8	74.6	90.9	92.3	72.0	97.6	96.5	88.8
1000	100.0	99.2	96.6	99.8	98.7	92.7	99.7	98.7	92.5	100.0	99.4	96.7
2000	100.0	99.9	99.4	100.0	99.9	98.4	100.0	99.9	97.8	100.0	99.9	98.9
5000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo test uses $N = 99$ simulations. Asymmetric test results are level-corrected using $N = 10000$ and true parameters (*i.e.*, infeasible test in practice as it requires knowing the true DGP). LMC test results are level-corrected using $N = 1,000$ and true parameters.

When considering MMC tests, we consider the consistent set define by

$$C_T^{(1)}(\delta^0) = \left\{\tilde{\theta}_0 \in \tilde{\Omega}_0 : \left\|\phi_k - \hat{\phi}_k^0\right\| \leq 0.01, |\phi_k| \leq 0.999, \left\|\sigma_y - \hat{\sigma}_y^0\right\| \leq 0.05, |\sigma_y| \geq 0.01, \left\|\sigma_v - \hat{\sigma}_v^0\right\| \leq 0.05, |\sigma_v| \geq 0.01\right\} \quad (7.2)$$

for SVL(1) models. For SVL(2) models, we will consider a larger set which includes more values of the variance parameters. This set is given by

$$C_T^{(2)}(\delta^0) = \left\{\tilde{\theta}_0 \in \tilde{\Omega}_0 : \left\|\phi_k - \hat{\phi}_k^0\right\| \leq 0.01, |\phi_k| \leq 0.999, \left\|\sigma_y - \hat{\sigma}_y^0\right\| \leq 0.10, |\sigma_y| \geq 0.01, \left\|\sigma_v - \hat{\sigma}_v^0\right\| \leq 0.10, |\sigma_v| \geq 0.01\right\}. \quad (7.3)$$

We report the rejection frequency of the LMC and MMC test procedures when testing for no leverage. For comparison, we also provide results for the asymptotic test where the critical value is obtained from a $\chi^2_{(1)}$ distribution rather than from the simulated null distribution. As we show, the asymptotic test performs very poorly in this setting as size is not controlled. All rejection frequencies are obtained using $1,000$ replications. Rejection frequencies when the alternative hypothesis is true (*i.e.*, power study) are obtained using critical values that are locally level-corrected. Note that we use the term "locally level-corrected" instead of "size-corrected" because a true size correction would require one to ensure that the probability of rejecting the null hypothesis under all distributions compatible with null hypothesis (*i.e.*, for all values of the nuisance parameters) be less than or equal to the level α. However, finding the appropriate size-corrected critical values requires a numerical search that was not performed in the experiments.

The level-corrected critical value is obtained by simulating the test statistic under the null hypothesis with $10,000$ replications for asymptotic tests and $1,000$ replications for LMC tests. It is worth noting that, as a result, the power results for the asymptotic test shown below are infeasible in practice as they require knowing the true DGP. These are only provided for comparison. Although the LMC test power results are also level corrected, as is shown, the degree of over-rejection is not as severe as is the case for the asymptotic test. In addition, we provide results in the appendix where the power of the LMC test is not level-corrected and there is little difference for the examples considered here. On the other hand, the MMC test does not involve the use of level-corrected critical values as it already involves a search over values consistent with the null hypothesis. Further, this test does not display any over-rejection under the null hypothesis. The maximization of MMC tests was done using a particle swarm algorithm and calculations were performed with the R system [see MaxMC package Dufour and Neves (2018); Bendtsen. (2022)]. For all scenarios, we examine the performance of each test procedure across various sample sizes: $T = 500$, $1,000$, $2,000$, $5,000$. Our findings can be summarized as follows.

First, we consider testing for leverage in an SVL(1) model. Table 2 reports the empirical size of the asymptotic test, LMC test, and MMC test for no leverage across four different sample sizes and two distinct DGPs. These DGPs possess moderate persistence in the volatility process, characterized by $\theta = (0.90, 0.10, 0.75, 0.00)$ and $\theta = (0.75, 0.10, 1.00, 0.00)$, respectively, consistent with the null hypothesis. As observed, the asymptotic test exhibits a significant degree of over-rejection, failing to control the test size ($\alpha = 5\%$). Conversely, the simulation-based Monte Carlo procedures outlined here perform well in comparison, achieving a rejection frequency much closer to the nominal test level. However, we note that the LMC procedure still exhibits some over-rejection, particularly evident in the case where $\phi = 0.90$, although showing a decreasing trend as the sample size increases. These findings also indicate that as persistence increases from $\phi = 0.75$ to $\phi = 0.90$, both the asymptotic test and the LMC test demonstrate higher levels of over-rejection. On the contrary, this analysis highlights the effectiveness of the MMC procedure, which successfully controls the test size by maintaining a rejection frequency less than or equal to the nominal level, as prescribed by theory.

Second, to showcase the power of these test procedures we considering DGPs with the same values for ϕ, σ_y, and σ_v, but consider different values of δ that are consistent with the alternative hypothesis. Table 3 contains the rejections frequencies for the two DGPs (moderate persistence) and four different sample sizes when $\delta = -0.90, -0.70, 0.70, 0.90$. Although it is commonly the case when considering financial time series that the leverage parameter is negative, the last six columns include test results when leverage is positive to show that the power of these tests appear to be symmetrical, in that they are very comparable to the rejection frequencies when leverage is negative values of the same magnitude. We also consider more values of δ and summarize all these results graphically in Appendix Figures A2 and A3. When considering these results, it is important to remember that the rejection frequencies for the asymptotic test and the LMC test are obtained using critical values that are level-corrected as described above. These are used instead of the conventional χ^2_1 critical values so that they may be more comparable since, as seen from Table 2, using the conventional critical values would likely lead to suggest there is higher degree of power that is actually due to the over-rejection of these test procedures. Hence, the power results for the asymptotic test shown here are infeasible in practice as we do not typically know the true DGP parameter values under the null hypothesis. However, Appendix Table A1 shows the same results except that the LMC test rejection frequencies are no longer level-corrected and here we see that there is not a large difference in terms of power, suggesting that the LMC test still performs quite well in terms of power. The Appendix also provides similar figures as described above when the LMC test is not level corrected. Those results are summarized graphically in Appendix Figures A4 and A5. As expected, the power of all test procedures increase as the sample size increases. These results also show that when persistence increases from $\phi = 0.75$ to $\phi = 0.90$, the power of all test procedures decreases slightly. This is especially the case for smaller sample sizes but less prevalent for larger samples sizes.

Third, we consider two more DGPs where we increase the degree of persistence. Specifically, we consider a $\phi = 0.95$ and $\phi = 0.99$. When considering these DGPs under the null hypothesis, such that $\delta = 0$, we find that there is a higher degree of over-rejection from both the asymptotic test and the LMC test. However, for the LMC test, the over-rejection decreases as the sample size increases. This can be seen from Table 4. Here, we see once again the benefit of the MMC test procedure. The MMC procedure is able to control the size of the test even for the more difficult cases where persistence is very high (e.g., $\phi = 0.99$) and even does so when considering small sample sizes.

Fourth, Table 5 shows that when persistence is much higher the power of the test is also affected. While the power remains relatively high when $\phi = 0.95$, it decreases notably when $\phi = 0.99$, such that the power of the LMC test, for instance, is not as substantial. Nevertheless, the power of the LMC test is comparable to the infeasible asymptotic test, which suggest it still performs quite well. As can be seen, the power still increases with the sample size when persistence is high as expected and so we should expect that for cases where persistence is very high, a larger sample size can achieve even better power results. As before, we also provide a table where the LMC test is not locally level-corrected; see Appendix Table A2. Similarly, Appendix Figures A6-A9 summarize results for more values of δ for both DGPs considered here and for the case where the LMC test is locally level-corrected and when it is not.

Fifth, we test the hypothesis of no leverage when the DGP includes an additional lag (*i.e.*, when $p = 2$). Specifically, we consider DGPs that are similar to the ones considered for the SVL(1) case in the sense that the persistence is similar *i.e.*, the roots are close to the unit root. Table 6 summarizes the results of the test procedures under the null hypothesis. The results for the asymptotic and LMC tests are similar to when we considered an SVL(1) model in that the the asymptotic test displays a very high degree of over-rejection, while the LMC test displays a small degree of over-rejection that appears to decrease for larger sample sizes and is especially smaller for the second DGP where $\phi_1 = 0.05$ and $\phi_2 = 0.70$. As previously mentioned, here we expand the consistent set used for the MMC test slightly by considering a wider space for the variance parameters. As before, we find that the MMC test performs remarkably well in controlling the size of the test, highlighting once more the benefit of this test procedure, which is feasible for the computationally simple moment-based estimation procedure proposed here.

Finally, Table 7 show results when the test procedures are applied to an SVL(2) model that is consistent with the alternative hypothesis. Specifically, we again report results when $\delta = -0.90, -0.70, 0.70, 0.90$ but results for more values of δ are provided in Appendix Figures A10-A13. The results shown here are quite similar to the SVL(1) case, though the MMC procedure appears to have even higher power for smaller sample sizes. Overall, we find that the performance of the MC simulation-based tests are favourable under various settings.

7.3 Forecasting

When forecasting, we work with the log of the squared returns, which according to (2.9) and (5.7) is given by $\log(\hat{y}^2_{T+h|T}) = \mu + H'\hat{\xi}_{T+h|T}$. Table 8 reports the ratio of the MSE_{SV} and MSE_{SVL} where the subscript SV and SVL

Table 8. MSE ratios and DM test for forecasting volatility in SV models with and without leverage.

	SV(1) vs. SVL(1)						SV(2) vs. SVL(2)					
	$h=1$		$h=5$		$h=10$		$h=1$		$h=5$		$h=10$	
δ	MSE ratio	DM Stat.	MSE ratio	DM Stat.	MSE ratio	DM Stat.	MSE ratio	DM Stat.	MSE ratio	DM Stat.	MSE ratio	DM Stat.
	$\phi=0.90$, $\sigma_y=0.10$, $\sigma_v=0.75$						$\phi_1=0.05$, $\phi_2=0.85$, $\sigma_y=1.00$, $\sigma_v=1.00$					
-0.9	1.14	7.86***	1.08	8.69***	1.05	8.75***	1.14	7.87***	1.03	7.12***	1.01	7.07***
-0.7	1.06	3.77***	1.04	5.71***	1.03	5.70***	1.08	6.59***	1.02	5.43***	1.01	5.50***
-0.5	1.02	1.75*	1.02	3.64***	1.01	3.57***	1.04	4.86***	1.01	3.94***	1.00	4.00***
0.5	1.02	2.67***	1.01	3.96***	1.01	4.54***	1.04	5.08***	1.01	4.11***	1.00	4.37***
0.7	1.07	5.43***	1.04	6.91***	1.03	7.27***	1.08	6.59***	1.01	5.96***	1.01	6.21***
0.9	1.15	7.94***	1.09	9.13***	1.05	9.11***	1.14	7.97***	1.03	7.86***	1.01	7.95***
	$\phi=0.75$, $\sigma_y=0.10$, $\sigma_v=1.00$						$\phi_1=0.05$, $\phi_2=0.70$, $\sigma_y=1.00$, $\sigma_v=1.00$					
-0.9	1.19	8.13***	1.06	8.56***	1.03	8.57***	1.17	7.68***	1.03	7.78***	1.01	7.78***
-0.7	1.10	5.91***	1.04	7.32***	1.02	7.35***	1.09	5.94***	1.02	6.10***	1.01	6.09***
-0.5	1.04	3.42***	1.02	5.51***	1.01	5.54***	1.04	3.73***	1.01	3.98***	1.00	3.97***
0.5	1.03	1.92*	1.01	1.82*	1.00	1.99**	1.04	4.37***	1.01	4.39***	1.00	4.44***
0.7	1.08	4.65***	1.03	5.62***	1.02	5.75***	1.09	6.43***	1.02	6.44***	1.01	6.45***
0.9	1.19	7.92***	1.06	8.69***	1.03	8.70***	1.18	7.98***	1.03	8.12***	1.01	8.12***

Notes: We simulate a process with $T=2,010$ observations, estimate the model using the first $T_{est}=2,000$, and forecast the last 10 observations. This process is repeated $B=3,000$ times. The MSE ratio is computed by using the MSE of the SV(p) model in the numerator and the MSE of the SVL(p) model in the denominator, so that a value greater than 1 suggests that the model with leverage provides a better forecast. The reported values represent the Mean Squared Error (MSE) ratio for each model at different horizons: $h=1$ (one day), $h=5$ (one week), and $h=10$ (two weeks). The DM statistic is computed as in equation (7.6). The significance level of the DM statistic is indicated using $(*)$ for 10% significance level, $(**)$ for 5% significance level, and $(***)$ for 1% significance level.

denotes the MSE from the model with no leverage (*i.e.*, $\delta=0$) and the MSE from the model with leverage (*i.e.*, $\delta \neq 0$) respectively. Specifically, the MSE of each model is computed as,

$$MSE_m = \frac{1}{B}\sum_{i=1}^{B}\left(\sum_{j=1}^{h}\left[\log(y_{t+j}^2)-\log(\hat{y}_{m,t+j|t}^2)\right]^2\right), \tag{7.4}$$

where $m=\{SV, SVL\}$ and B is the number of simulations. In practice, we would set B to be the number of out-of-sample observations after t as in the empirical section below. The ratio is then computed as MSE ratio $= MSE_{SV}/MSE_{SVL}$ and hence, a value greater than 1 suggests the model with leverage performs better. This ratio is provided for values of $\delta=-0.9,-0.7,-0.5,0.5,0.7,0.9$ along different rows and horizons $h=1,5,10$ along different columns. Specifically, we consider some of the same DGPs as in the hypothesis test for no leverage study presented above. Additionally, we use the DM test of Diebold and Mariano (2002) as a means to determine the statistical significance of the difference in the forecast performance of a model with leverage and a model with no leverage. That is, we compute

$$d_i = \sum_{j=1}^{h}\left[\log(y_{t+j}^2)-\log(\hat{y}_{SV,t+j|t}^2)\right]^2 - \sum_{j=1}^{h}\left[\log(y_{t+j}^2)-\log(\hat{y}_{SVL,t+j|t}^2)\right]^2 \tag{7.5}$$

where as before, i denotes the number of the simulations. From here, the DM statistic is computed as

$$DM = \frac{\bar{d}}{\sqrt{(\gamma_d(0)+2\sum_{k=1}^{B^{1/3}}\gamma_d(k))/B}} \tag{7.6}$$

where $\bar{d}=\frac{1}{B}\sum_{i=1}^{B} d_i$ and $\gamma_d(k)$ is the sample auto-covariance of d_i at lag k. As described in Diebold and Mariano (2002) this test statistic has a standard normal distribution under the null hypothesis of no statistical difference is

Table 9. Empirical W-ARMA estimates of SVL(p) models

	SVL(1)				SVL(2)					SVL(3)					
	$\hat{\phi}$	$\hat{\sigma}_y$	$\hat{\sigma}_v$	$\hat{\delta}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_y$	$\hat{\sigma}_v$	$\hat{\delta}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\sigma}_y$	$\hat{\sigma}_v$	$\hat{\delta}$
								S&P 500							
est.	0.972	0.850	0.294	-0.774	0.427	0.549	0.850	0.719	-0.024	0.141	0.355	0.477	0.850	0.639	-0.151
SE	(0.001)	(0.055)	(0.011)	(0.115)	(0.033)	(0.031)	(0.131)	(0.056)	(0.498)	(0.006)	(0.086)	(0.009)	(0.109)	(0.075)	(0.171)
								DOWJ							
est.	0.970	0.811	0.294	-0.749	0.513	0.459	0.811	0.689	-0.035	0.300	0.235	0.434	0.811	0.628	-0.154
SE	(0.001)	(0.049)	(0.011)	(0.128)	(0.074)	(0.072)	(0.111)	(0.061)	(0.492)	(0.076)	(0.084)	(0.027)	(0.093)	(0.072)	(0.193)
								NASDQ							
est.	0.985	1.135	0.209	-0.998	0.424	0.563	1.135	0.621	-0.008	0.234	0.448	0.302	1.135	0.598	-0.046
SE	(0.001)	(0.099)	(0.005)	(0.001)	(0.031)	(0.031)	(0.25)	(0.070)	(0.506)	(0.044)	(0.047)	(0.080)	(0.217)	(0.076)	(0.487)

Notes: Sample for each index is from 2000-Jan-04 to 2023-May-31 ($T = 5,889$). Estimates are obtained using W-ARMA estimator given in (3.24) with $J = 250$. Standard errors of parameters are obtained using the implicit standard errors procedure described in Section 6.3.

Table 10. Asymptotic and finite-sample tests for the presence of leverage in SVL(p) models

	H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$								
	SVL(1)			SVL(2)			SVL(3)		
	Asymptotic	LMC	MMC	Asymptotic	LMC	MMC	Asymptotic	LMC	MMC
S&P 500	0.00	0.00	0.01	0.07	0.83	0.95	0.00	0.32	0.70
Dow Jones	0.00	0.00	0.01	0.01	0.84	0.94	0.00	0.37	0.46
NASDAQ	0.00	0.00	0.01	0.56	0.87	0.92	0.00	0.70	0.89

Notes: Sample for each index is from 2000-Jan-04 to 2023-May-31 ($T = 5,889$). The reported values are p-values for test procedure when testing for leverage (*i.e.* $H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$). When estimating the constrained and unconstrained models we used W-ARMA estimator given in (3.24) with $J = 250$. We use $N = 999$ Monte Carlo simulations to simulate the null distribution for LMC and MMC tests. Table A5 extends this table by considering $N = 99$ and $N = 299$ for both the LMC and MMC test procedures.

the forecast error. That is, we test the null hypothesis $H_0 : \mathrm{DM} = 0$ against an alternative hypothesis $H_1 : \mathrm{DM} \neq 0$.

From Table 8, we find that when leverage is present, the forecasts of the model with leverage outperforms the forecasts of the model without leverage. This is seen as the MSE ratio is above 1. Specifically, as we should expect, we find that when δ is higher, the ratio becomes larger indicating the performance difference is greater. This is also reflected in the DM statistic, which is larger for larger values of δ. As previously discussed, we see from the first line of equation (5.7) that δ shows up in the forecast but is multiplied by $F^{(h-1)}$ and so for stable models, as h increases, the effect of δ will decay. This feature can be seen in Table 8 since we can see the MSE ratio decrease as h increases. However, we still find that the difference in forecasting performance is still statistically significant and in favour of the model with leverage even at horizon $h = 10$ (*e.g.*, two weeks ahead if working with daily data).

8 Application to stock price indices

In this section, we demonstrate the performance of the SVL(p) model proposed in this paper when applied to three financial indices. Namely, we consider the Standard and Poor's 500 Composite Price Index (S&P500), the Dow Jones Industrial Average Price Index (DOWJ), and the Nasdaq Composite Price Index (NASDQ). SVL(p) models are fitted to daily demeaned returns of these indices. Specifically, the raw series p_t are converted to returns by the transformation $r_t = 100[\log(p_t) - \log(p_{t-1})]$ and then the returns are demeaned by $y_t = r_t - \hat{\mu}_r$, where $\hat{\mu}_r$ is the sample average of

Table 11. Empirical volatility forecasting performance with competing conditional volatility models

	S&P 500			Dow Jones			NASDAQ		
Models	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
ARCH(1)	13.629	65.351	128.294	11.982	60.489	121.549	14.48	68.309	132.324
ARCH(2)	9.514	50.653	105.372	9.482	50.428	104.794	8.981	47.185	97.015
ARCH(3)	8.936	47.558	100.909	8.882	47.319	100.274	8.690	45.376	94.246
GARCH(1,1)	8.297	42.552	87.398	8.279	42.450	87.042	7.918	40.114	**81.540**
GARCH(2,2)	8.259	42.378	87.181	8.260	42.300	86.859	7.884	39.988	**81.390**
GARCH(3,3)	8.251	42.326	87.128	8.260	42.288	86.883	7.889	39.989	**81.404**
EGARCH(1,1)	7.971	40.683	82.891	**8.006**	**40.792**	83.014	7.674	38.927	**78.895**
EGARCH(2,2)	7.878	40.272	82.137	**7.941**	**40.486**	82.466	8.109	40.711	**82.047**
EGARCH(3,3)	7.792	39.916	81.499	**11.809**	**59.887**	**110.934**	9.855	53.878	**161.959**
GJR(1,1)	8.059	41.551	85.553	8.051	41.430	85.102	7.723	39.313	**79.968**
GJR(2,2)	8.034	41.432	85.409	8.030	**41.267**	84.842	7.702	39.206	**79.821**
GJR(3,3)	8.024	41.372	85.335	8.036	**41.280**	84.912	7.690	39.189	**79.825**
SV(1)	6.261	**31.721**	**64.657**	**6.191**	**31.379**	**63.867**	6.088	**30.669**	**62.200**
SV(2)	6.257	32.863	67.359	**6.183**	**32.433**	**66.462**	6.027	31.370	**63.962**
SV(3)	**6.193**	33.077	67.616	**6.092**	**32.704**	**66.798**	5.987	31.607	**64.239**
SVL(1)	**6.148**	**31.526**	**64.508**	**6.102**	**31.128**	**63.579**	**5.949**	**30.323**	**61.786**
SVL(2)	**6.174**	32.690	67.183	**6.118**	**32.262**	**66.283**	**5.927**	31.186	**63.775**
SVL(3)	**6.097**	32.915	67.451	**6.007**	**32.533**	**66.623**	**5.875**	31.432	**64.063**

Notes: The reported values represent the Mean Squared Error (MSE) for each model at different horizons: $h=1$ (one day), $h=5$ (one week), and $h=10$ (two weeks). The values in bold indicate that the model is part of the Model Confidence Set (MCS) for that horizon and that specific index. The MCS is determined using a 5% significance level. The values in bold red indicate the models in the MCS with the lowest MSE. Sample for each index is from 2000-Jan-04 to 2023-May-31 ($T = 5{,}889$). Estimates are obtained using W-ARMA estimator given in (3.24) with $J = 250$. Out-of-sample forecasting is performed using a rolling window scheme of size $T_{est} = 1{,}000$, where the first estimation window ends at observation $1{,}000$, and the last estimation window starts at observation $4{,}880$ but ends at observation $5{,}879$.

returns. We consider a sample from January 4, 2000 to May 31, 2023 (*i.e.*, $T = 5{,}889$) so that all indices are available for the same time period and so that we capture various volatile periods.

The Appendix Figure A14 shows plots of y_t for each index. Additionally, Appendix Table A4 reports summary statistics of the daily residual returns (y_t as well as summery statistics for y_t^2 and y_t^*, where $y_t^* = \log(y_t^2) - \mu$ as in equation (2.9), for each financial index. The skewness and kurtosis of y_t and y_t^2 for each index show some evidence of non-normal distribution, while the distribution of log-transformed residual returns is closer to normal. These findings are consistent with most empirical studies.

Table 9 shows the parameter estimates of the SVL(p) models (where p = 1, 2, 3) using our W-ARMA estimator. For all estimates in this section we use (3.24) with $J = 250$ with equal weights to estimate ϕ_p and the other expressions in equations (3.15) and (3.16) to obtain estimates of the other parameters. This table also reports implicit standard errors of each parameter, which are obtained using the procedure described in section 6.3. Our results show some persistence in the volatility process for each index during the sample considered here. We also find that leverage (*i.e.*, δ) is very high when using an SVL(1) model, but very small and in many cases close to 0 when considering an SVL(2) or SVL(3) model.

As in the simulation section above, we examine a test for no leverage in SVL(p) models, specifically testing $H_0 : \delta = 0$ against $H_1 : \delta \neq 0$, using the moment-based LR test described in Section 6.1. This entails utilizing the moments provided in (6.3) for both the constrained and unconstrained models to calculate the LR statistic.

The LR test results are reported in Table 10, where we consider the asymptotic test, the LMC test, and the MMC test procedures. The asymptotic test always appears to reject the null hypothesis of no leverage at a significance level of 10%, even for the SVL(2) and SVL(3) case where δ was shown to be very small in Table 9, with the only exception being when the SVL(2) model is applied to NASDAQ. This is likely due to the over-rejection we found in the simulation study of section 7.2. On the other hand, the LMC test results appear to be more in line with what we should expect given the values of δ being closer to 0 when using an SVL(2) or SVL(3) model. That is, we reject the null hypothesis of no leverage when considering an SVL(1) model but fail to reject the null hypothesis when considering an SVL(2) or SVL(3) model.

For the MMC test, we use the same confidence set as in (7.2) when searching the nuisance parameter space for a maximum p-value. The reason for choosing this set and not the wider set given in (7.3) is because we see that with the LMC test, only the SVL(1) model appears to reject the null hypothesis of no leverage. As a result we find it suitable to use the same confidence set as in the SVL(1) case shown in the simulation section where we saw this set was large enough to allow the MMC test procedure to control size. Whereas with the SVL(2) and SVL(3) models, the LMC procedure already fails to reject the null hypothesis and so it is not necessary to search a very wide space.

Table 11 provides results of a forecast study. Specifically, we consider various different GARCH-type, SV-type and SVL-type models along different rows and report their MSE when forecasting the log squared residual returns of each index at horizon $h = 1$, 5, and 10. In order to compute the MSE, we use a fixed estimation window of $T_{est} = 1{,}000$ and then roll the estimation window forward, while always forecasting the next $h = 1, \ldots, 10$ periods. Since we have a total sample size of $T = 5{,}889$, this results in $B = 4{,}880$ forecast errors for each horizon. The reason for keeping a short estimation window is based on the premise that very distant observations may contain little useful information for use in forecasting in more recent periods as discussed in Kambouroudis and McMillan (2015). The squared forecast errors are also used to perform the Model Confidence Set (MCS) test of Hansen et al. (2011) using a significance level of 5% in order to determine the set of models that outperform the rest in terms of forecasting performance. When using the MCS test we are testing the following hypothesis

$$H_0 : \mu_i = 0 \quad \text{for all} \quad i \in \mathcal{M} \quad \text{and} \quad H_1 : \mu_i \neq 0 \quad \text{for some} \quad i \in \mathcal{M} \tag{8.1}$$

where $\mu_i = E[d_i]$, which is computed as $\bar{d}_i = m^{-1} \sum_{j \in \mathcal{M}} \bar{d}_{i,j}$ with i, j indexing two different models and $\mathcal{M}$ is the set of all models being considered here. In this case, the test statistic of interest is $MCS_{T_{max,\mathcal{M}}} = \max_{i \in \mathcal{M}} t_i$ where $t_i = \frac{\bar{d}_i}{\sqrt{\hat{\text{Var}}(\bar{d}_i)}}$ and where $\hat{\text{Var}}(\bar{d}_i)$ is a bootstrap estimate of $\text{Var}(\bar{d}_i)$. That is, we consider the T_{max} version of the test discussed in Hansen et al. (2011). This test is applied using R package 'MCS' (see Bernardi (2017)). Our results show that the SVL(1) model is the only model that is always included in the MCS at all horizons and for all indices. From Table 11 we can also see that the SV(1), SV(3) and SVL(2) and SVL(3) model appear in the MCS for specific horizons for all three indices. When considering Dow Jones and NASDAQ data, the SV(2) model also appears in the MCS for specific horizons. Further, for these latter two indices, some EGARCH models also appear in the MCS. The MCS even includes some GARCH models when considering NASDAQ data.

We also consider a larger estimation window $T_{est} = 4{,}880$ and for which the results are shown in Appendix Table A6. As before, we still find that the SVL(1) model is always part of the MCS but in this case only the SV(1), SV(3) and SVL(3) also appear in the MCS.

9 Conclusion

Despite their conceptual appeal, SVL models pose challenges for estimation and inference due to the presence of latent variables, especially when extending beyond the SVL(1) model. To address this, we propose a novel moment-based simple CF-ARMA estimator for SVL(p) models. This estimator offers analytical tractability and computational efficiency, thereby overcoming limitations associated with existing methods. Moreover, we introduce enhancements such as restricted and winsorized versions of the CF-ARMA estimator, which further improve the stability and accuracy of the estimator, particularly in the presence of outliers or small samples. We demonstrate that these simple estimators are especially convenient for use in simulation-based inference techniques, such as Bootstrap or Monte Carlo tests.

Our study encompasses several simulation experiments and empirical applications. Utilizing the W-ARMA-OLS estimator, we showcase the advantages of LMC and MMC methods in controlling the size and power of LR-type tests. Additionally, we explore the statistical properties and forecasting performance of the proposed estimators, highlighting the significance of leverage in volatility forecasting. Importantly, our empirical findings emphasize the superiority of SVL(p) models over competing conditional volatility models, such as GARCH and SV(p), in terms of forecast accuracy.

In summary, our research contributes to the advancement of estimation and inference techniques for SVL(p) models, providing valuable insights into understanding and modeling volatility dynamics in financial markets. Through empirical applications and simulation studies, we offer practitioners and researchers practical tools and methodologies that are beneficial for risk management, option pricing, asset allocation, and volatility forecasting.

References

Abramowitz, M. and Stegun, I. A. (1970), *Handbook of Mathematical Functions*, Vol. 55, Dover Publications Inc., New York.

Ahsan, M. N. and Dufour, J.-M. (2019), A simple efficient moment-based estimator for the stochastic volatility model, *in* I. Jeliazkov and J. Tobias, eds, 'Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling', Vol. 40 of *Advances in Econometrics*, Emerald, Bingley, U.K., pp. 157–201.

Ahsan, M. N. and Dufour, J.-M. (2020), Volatility forecasting and option pricing with higher-order stochastic volatility models, Technical report, McGill University.

Ahsan, M. N. and Dufour, J.-M. (2021), 'Simple estimators and inference for higher-order stochastic volatility models', *Journal of Econometrics* **224**(1), 181–197. Annals Issue: PI Day.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**(2), 579–625.

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., Nicolato, E. and Shephard, N. (2002), ‘Some recent developments in stochastic volatility modelling’, *Quantitative Finance* **2**(1), 11–23.

Bates, D. M. and Watts, D. G. (1988), *Nonlinear Regression Analysis and its Applications*, John Wiley & Sons, New York.

Bendtsen., C. (2022), *pso: Particle Swarm Optimization.* R package version 1.0.4.
URL: *https://CRAN.R-project.org/package=pso*

Bernardi, L. C. . M. (2017), *MCS: Model Confidence Set Procedure.* R package version 0.1.3.
URL: *https://CRAN.R-project.org/package=MCS*

Black, F. (1976), ‘Studies of stock market volatility changes’, *Proceedings of the American Statistical Association, Business & Economic Statistics Section, 1976* .

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

Chen, L., Zerilli, P. and Baum, C. F. (2019), ‘Leverage effects and stochastic volatility in spot oil returns: A Bayesian approach with VaR and CVaR applications’, *Energy Economics* **79**, 111–129.

Chesney, M. and Scott, L. (1989), ‘Pricing European currency options: a comparison of the modified Black- Scholes model and a random variance model’, *The Journal of Financial and Quantitative Analysis* **24**(3), 267–284.

Christie, A. (1982), ‘The stochastic behavior of common stock variances Value, leverage and interest rate effects’, *Journal of Financial Economics* **10**(4), 407–432.

Den Haan, W. J. and Levin, A. (1997), A practitioner’s guide to robust covariance matrix estimation, *in* G. S. Maddala and C. R. Rao, eds, ‘Handbook of Statistics: Robust Inference’, Vol. 15, Elsevier, pp. 299–342.

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

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

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

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'A Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Neves, J. (2018), Maximized Monte Carlo: an R package, Technical report, McGill University.

Dufour, J.-M., Trognon, A. and Tuvaandorj, P. (2017), 'Invariant tests based on M-estimators, estimating functions, and the generalized method of moments', *Econometric Reviews* **36**(1-3), 182–204.

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

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

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

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

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

Hull, J. and White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *The Journal of Finance* **42**(2), 281–300.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2004), 'Bayesian analysis of stochastic volatility models with fat-tails and correlated errors', *Journal of Econometrics* **122**(1), 185–212.

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.

Jurado, K., Ludvigson, S. C. and Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

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

Kastner, G. (2016), 'Dealing with stochastic volatility in time series using the R package stochvol', *Journal of Statistical Software* **69**(5), 1–30.

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

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), 'Hypothesis testing with efficient method of moments estimators', *International Economic Review* **28**, 777–787.

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.

Robinson, P. M. and Velasco, C. (1997), Autocorrelation-robust inference, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics: Robust Inference', Vol. 15, Elsevier, pp. 267–298.

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

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

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

Wiggins, J. B. (1987), 'Option values under stochastic volatility: Theory and empirical estimates', *Journal of Financial Economics* **19**(2), 351–372.

Yu, J. (2005), 'On leverage in a stochastic volatility model', *Journal of Econometrics* **127**(2), 165–178.

Estimation and inference for higher-order stochastic volatility models with leverage

Supplementary Appendix

Md. Nazmul Ahsan[†], Jean-Marie Dufour[‡] and Gabriel Rodriguez-Rondon[§]

February 29, 2024

This online Supplementary Appendix contains additional details and complementary results relevant to the paper.

[†]Department of Economics, McGill University, Leacock Building, 855 Sherbrooke StreetWest, Montréal, Québec H3A 2T7, Canada; e-mail: md.ahsan@mail.mcgill.ca.

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

[§]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. e-mail: gabriel.rodriguezrondon@mail.mcgill.ca. Wed page: https://grodriguezrondon.com

A Proofs

PROOF OF PROPOSITION 3.1 From (2.10) - (2.11), we have

$$\phi(B)w_t = v_t\,, \quad y_t^* = w_t + \epsilon_t \tag{A.1}$$

where $\phi(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$ and $v_t = \delta_p z_{t-1} + (1-\delta_p)^{1/2}\bar{v}_t$. Furthermore, v_t's and ϵ_t's are uncorrelated. This result follows from

$$\mathrm{Cov}(v_t, \epsilon_t) = \mathbb{E}\big[v_t(h(z_t) - \mathbb{E}[\epsilon_t])\big] = \mathbb{E}\big[v_t(h(z_t)\big] = \mathbb{E}\big[\mathbb{E}(v_t|z_t)h(z_t)\big]\,, \tag{A.2}$$

where $h(z_t) = \log(z_t^2)$. Since $\mathbb{E}\big[\mathbb{E}(v_t|z_t)h(z_t)\big]$ is an odd function of z_t, and the density of z_t $[f(z_t)]$ is symmetric, its expected value is zero. Thus, $\mathrm{Cov}(v_t, \epsilon_t) = 0$.

Now applying $\phi(B)$ to both sides of (2.11) yields

$$\phi(B)y_t^* = \phi(B)w_t + \phi(B)\epsilon_t = v_t + \phi(B)\epsilon_t. \tag{A.3}$$

The right hand side of (A.3) is clearly a covariance stationary process. By the Wold decomposition theorem it must have a moving average representation. Since the autocovariance function cuts off for lags $k > p$ it must be an $MA(p)$ process, say $\theta(B)\eta_t = (1-\theta_1 B - \cdots - \theta_p B^p)\eta_t$. Hence, y_t^* must be an ARMA(p, p) process [see equation (2.1) of Granger and Morris (1976)].

The moving average parameters $\theta_1, \theta_2, \ldots, \theta_p$ and the white noise variance σ_η^2 of this ARMA(p, p) process can be found by equating the autocovariance function of the right hand side of (A.3) with that of $\theta(B)\eta_t$ for lags $k = 0, 1, \ldots, p$ and solving the $p+1$ resulting nonlinear equations

$$\begin{aligned}
(1+\theta_1^2+\cdots+\theta_p^2)\sigma_\eta^2 &= \sigma_v^2 + (1+\phi_1^2+\cdots+\phi_p^2)\sigma_\epsilon^2\,,\\
(-\theta_1+\theta_1\theta_2+\cdots+\theta_{p-1}\theta_p)\sigma_\eta^2 &= (-\phi_1+\phi_1\phi_2+\cdots+\phi_{p-1}\phi_p)\sigma_\epsilon^2\,,\\
&\ \ \vdots\\
(-\theta_{p-1}+\theta_1\theta_p)\sigma_\eta^2 &= (-\phi_{p-1}+\phi_1\phi_p)\sigma_\epsilon^2\,,\\
-\theta_p\sigma_\eta^2 &= -\phi_p\sigma_\epsilon^2.
\end{aligned} \tag{A.4}$$

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

PROOF OF COROLLARY 3.2 From Proposition 3.1, the observed process y_t^* satisfies the following equation:

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j} \tag{A.5}$$

or

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + v_t + \epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j}. \tag{A.6}$$

Multiply both sides of (A.6) by y^*_{t-k}, and taking expectation, we get:

$$\gamma_{y^*}(k) = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \mathbb{E}[v_t y^*_{t-k}] + \mathbb{E}[\epsilon_t y^*_{t-k}] - \sum_{j=1}^{p} \phi_j \mathbb{E}[\epsilon_{t-j} y^*_{t-k}]. \tag{A.7}$$

For $k=0$, we get

$$\begin{aligned}\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \mathbb{E}[v_t y^*_t] + \mathbb{E}[\epsilon_t y^*_t] - \sum_{j=1}^{p} \phi_j \mathbb{E}[\epsilon_{t-j} y^*_t] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma^2_v + \sigma^2_\epsilon - \sum_{j=1}^{p} \phi_j \mathbb{E}[\epsilon_{t-j}(\phi_j y^*_{t-j} - \phi_j \epsilon_{t-j})] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma^2_v + \sigma^2_\epsilon - \sum_{j=1}^{p} \phi_j^2 \mathbb{E}[\epsilon_{t-j} y^*_{t-j} - \epsilon^2_{t-j}] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma^2_v + \sigma^2_\epsilon - \sum_{j=1}^{p} \phi_j^2 [\sigma^2_\epsilon - \sigma^2_\epsilon] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \sigma^2_v + \sigma^2_\epsilon.\end{aligned} \tag{A.8}$$

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

$$\begin{aligned}\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \mathbb{E}[v_t y^*_{t-k}] + \mathbb{E}[\epsilon_t y^*_{t-k}] - \sum_{j=1}^{p} \phi_j \mathbb{E}[\epsilon_{t-j} y^*_{t-k}] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + 0 + 0 - \phi_k \mathbb{E}[\epsilon_{t-k} y^*_{t-k}] = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) - \phi_k \sigma^2_\epsilon.\end{aligned} \tag{A.9}$$

Setting $k > p$, we get

$$\begin{aligned}\gamma_{y^*}(k) &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \mathbb{E}[v_t y^*_{t-k}] + \mathbb{E}[\epsilon_t y^*_{t-k}] - \sum_{j=1}^{p} \phi_j \mathbb{E}[\epsilon_{t-1} y^*_{t-k}] \\ &= \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + 0 + 0 - 0 = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j).\end{aligned} \tag{A.10}$$

Combining (A.8), (A.9), and (A.10), we get the autocovariance structure of the observed process that stated in the Corollary. □

PROOF OF LEMMA 3.3 From Assumptions 2.1 - 2.2, we have

$$\mathbb{E}\big[|y_t| y_{t-1}\big] = \sigma^2_y \mathbb{E}\Big[\exp\Big(\frac{w_t}{2} + \frac{w_{t-1}}{2}\Big) z_{t-1}\Big] \mathbb{E}\Big[|z_t|\Big]. \tag{A.11}$$

For a standard normal random variable, z_t , it can be shown that

$$\mathbb{E}\Big[|z^i_t|\Big] = \frac{2^{i/2}}{\sqrt{\pi}} \Gamma\left(\frac{i+1}{2}\right), \tag{A.12}$$

thus for $i=1$, we have

$$\mathbb{E}\Big[|z_t|\Big] = \sqrt{2/\pi}. \tag{A.13}$$

Now define $A := \frac{w_t}{2} + \frac{w_{t-1}}{2}$ and $B := z_{t-1}$, we have:

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_A^2 & \frac{\delta\sigma_v}{2} \\ \frac{\delta\sigma_v}{2} & 1 \end{pmatrix}\right) \tag{A.14}$$

where

$$\sigma_A^2 = \mathrm{var}\left(\frac{w_t}{2} + \frac{w_{t-1}}{2}\right) = \frac{1}{4}\mathrm{var}(w_t + w_{t-1}) = \frac{1}{2}\left[\tilde{\gamma}_p\right], \tag{A.15}$$

$$\tilde{\gamma}_p := \gamma_p(0) + \gamma_p(1), \quad \gamma_p(0) := \mathrm{var}(w_t), \quad \gamma_p(1) := \mathrm{cov}(w_t, w_{t-1}). \tag{A.16}$$

The moment generating function for a bivariate mean zero normal random variable is given by

$$M(t_1, t_2) \equiv \mathbb{E}\left[\exp\left(t_1 A + t_2 B\right)\right] = \exp\left(\frac{\sigma_A^2 t_1^2 + \sigma_B^2 t_2^2 + 2\rho_{A,B}\sigma_A\sigma_B t_1 t_2}{2}\right). \tag{A.17}$$

Using (A.17) and (A.14), we get:

$$\begin{aligned} \left.\frac{\partial M}{\partial t_2}\right|_{t_1=1, t_2=0} &= \mathbb{E}[\exp(A)B] = \rho_{A,B}\sigma_A\sigma_B \exp\left(\frac{\sigma_A^2}{2}\right) \\ &= \frac{\delta\sigma_v}{2} \exp\left(\tilde{\gamma}_p/4\right) \end{aligned} \tag{A.18}$$

and, on combining (A.11), (A.13) and (A.18),

$$\mathbb{E}\left[|y_t| y_{t-1}\right] = \frac{\delta\sigma_v\sigma_y^2}{\sqrt{2\pi}} \exp\left(\tilde{\gamma}_p/4\right) \tag{A.19}$$

where $\tilde{\gamma}_p$ depends on the order of latent volatility process (p). For $p = 1, \ldots, 5$, this yields the following expressions for $\tilde{\gamma}_p$:

$$\tilde{\gamma}_1 = \frac{\sigma_v^2}{1-\phi_1}, \tag{A.20}$$

$$\tilde{\gamma}_2 = \frac{\sigma_v^2}{(1-\phi_1-\phi_2)(1+\phi_2)}, \tag{A.21}$$

$$\tilde{\gamma}_3 = \frac{(1-\phi_3)\sigma_v^2}{(1-\phi_1-\phi_2-\phi_3)(1+\phi_1\phi_3+\phi_2-\phi_3^2)}, \tag{A.22}$$

$$\tilde{\gamma}_4 = \frac{(1-\phi_3-\phi_1\phi_4-\phi_4^2)\sigma_v^2}{(1-\phi_1-\phi_2-\phi_3-\phi_4)(1+\phi_1\phi_3+\phi_2+\phi_4+2\phi_2\phi_4+{\phi_1}^2\phi_4+\phi_2{\phi_4}^2-{\phi_3}^2-{\phi_4}^2-{\phi_4}^3-\phi_1\phi_3\phi_4)}, \tag{A.23}$$

$$\tilde{\gamma}_5 = \frac{(1-\phi_3-\phi_1\phi_4-\phi_4^2-\phi_5-\phi_2\phi_5+\phi_3\phi_5-\phi_1^2\phi_5+\phi_2\phi_5^2-\phi_5^2+\phi_5^3-\phi_1\phi_4\phi_5)\sigma_v^2}{(1-\phi_1-\phi_2-\phi_3-\phi_4-\phi_5)\tilde{\kappa}}, \tag{A.24}$$

where

$$\begin{aligned} \tilde{\kappa} \ : \ &= 1+\phi_1\phi_3+\phi_2+\phi_4+2\phi_2\phi_4+{\phi_1}^2\phi_4+\phi_2{\phi_4}^2-{\phi_3}^2-{\phi_4}^2-{\phi_4}^3-\phi_1\phi_3\phi_4+\phi_1^3\phi_5 \\ &-\phi_1^2\phi_3\phi_5-\phi_1^2\phi_5^2+\phi_1\phi_2\phi_4\phi_5+3\phi_1\phi_2\phi_5+2\phi_1\phi_3\phi_5^2-\phi_1\phi_4^2\phi_5-2\phi_1\phi_4\phi_5-\phi_1\phi_5^3+\phi_1\phi_5 \\ &-\phi_2^2\phi_5^2-2\phi_2\phi_3\phi_5+\phi_2\phi_4\phi_5^2-\phi_2\phi_5^2+3\phi_3\phi_4\phi_5-\phi_3\phi_5^3+\phi_3\phi_5-\phi_4\phi_5^2+\phi_5^4-2\phi_5^2. \end{aligned} \tag{A.25}$$

□

PROOF OF COROLLARY 3.4 The estimator of $\phi_{\boldsymbol{p}}$ is based on the autocovariance structure of the process y_t^*. This is the solution of p-system of equations from (3.3) with $k = p+1, \dots, 2p$. Thus

$$\begin{bmatrix} \gamma_{y^*}(p) & \gamma_{y^*}(p-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(p+1) & \gamma_{y^*}(p) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p-1) & \gamma_{y^*}(2p-2) & \cdots & \gamma_{y^*}(p) \end{bmatrix} \cdot \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma_{y^*}(p+1) \\ \gamma_{y^*}(p+2) \\ \vdots \\ \gamma_{y^*}(2p) \end{bmatrix}, \tag{A.26}$$

or, equivalently,

$$\boldsymbol{\Gamma}(p,1)\phi_p = \gamma(p,1) \tag{A.27}$$

hence

$$\phi_p = \boldsymbol{\Gamma}(p,1)^{-1}\gamma(p,1) \tag{A.28}$$

where $\phi_p := (\phi_1, \dots, \phi_p)'$, $\gamma(p,1) = [\gamma_{y^*}(p+1), \dots, \gamma_{y^*}(2p)]'$ are vectors, and $\boldsymbol{\Gamma}(p,1)$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}(p,1) := \begin{bmatrix} \gamma_{y^*}(p) & \gamma_{y^*}(p-1) & \cdots & \gamma_{y^*}(1) \\ \gamma_{y^*}(p+1) & \gamma_{y^*}(p) & \cdots & \gamma_{y^*}(2) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p-1) & \gamma_{y^*}(2p-2) & \cdots & \gamma_{y^*}(p) \end{bmatrix}. \tag{A.29}$$

Note that (A.28) is also valid for any $j \geq 1$ such that

$$\phi_p = \boldsymbol{\Gamma}(p,j)^{-1}\gamma(p,j), \tag{A.30}$$

where $\gamma(p,j) := [\gamma_{y^*}(p+j), \dots, \gamma_{y^*}(2p+j-1)]'$ and $\boldsymbol{\Gamma}(p,j)$ is a p-dimensional Toeplitz matrices such that

$$\boldsymbol{\Gamma}(p,j) := \begin{bmatrix} \gamma_{y^*}(p+j-1) & \gamma_{y^*}(p+j-2) & \cdots & \gamma_{y^*}(j) \\ \gamma_{y^*}(p+j) & \gamma_{y^*}(p+j-1) & \cdots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & & \vdots \\ \gamma_{y^*}(2p+j-2) & \gamma_{y^*}(2p+j-3) & \cdots & \gamma_{y^*}(p+j-1) \end{bmatrix}. \tag{A.31}$$

Now, by (3.3) with $k = 0$, we have:

$$\gamma_{y^*}(0) = \phi_1\gamma_{y^*}(1) + \cdots + \phi_p\gamma_{y^*}(p) + \sigma_v^2 + \sigma_\epsilon^2, \tag{A.32}$$

hence

$$\sigma_v = [\gamma_{y^*}(0) - \sum_{j=1}^{p}\phi_j\gamma_{y^*}(j) - \pi^2/2]^{1/2} = [\gamma_{y^*}(0) - \phi_{\boldsymbol{p}}'\gamma(p) - \pi^2/2]^{1/2}, \tag{A.33}$$

where $\phi_{\boldsymbol{p}} := (\phi_1, \dots, \phi_p)'$, $\gamma(p) := [\gamma_{y^*}(1), \dots, \gamma_{y^*}(p)]'$ and $\sigma_\epsilon^2 = \psi^{(1)}(1/2) = \pi^2/2$. Now by construction,

$$\mu = \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(z_t^2)] = \log(\sigma_y^2) - 1.2704, \tag{A.34}$$

or, equivalently,

$$\sigma_y^2 = \exp(\mu + 1.2704). \tag{A.35}$$

Now from Lemma 3.3, we have:

$$\delta_p = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\tilde{\gamma}_p\Big), \tag{A.36}$$

where δ_p is a function of $\lambda_y(1) := \mathbb{E}[|y_t|y_{t-1}]$, $\tilde{\gamma}_p := \mathrm{var}(w_t) + \mathrm{cov}(w_t, w_{t-1})$ and other parameter estimates of the model $(\phi_1, \ldots, \phi_p, \sigma_v, \sigma_v)'$. Expressions of δ_p for the lower order SVL(p) models are:

$$\delta_1 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{\sigma_v^2}{1-\phi_1}\Big), \tag{A.37}$$

$$\delta_2 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{\sigma_v^2}{(1-\phi_1-\phi_2)(1+\phi_2)}\Big), \tag{A.38}$$

$$\delta_3 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{(1-\phi_3)\sigma_v^2}{(1-\phi_1-\phi_2-\phi_3)(1+\phi_1\phi_3+\phi_2-\phi_3^2)}\Big), \tag{A.39}$$

$$\delta_4 = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\Big(-\frac{1}{4}\frac{(1-\phi_3-\phi_1\phi_4-\phi_4^2)(1-\phi_1-\phi_2-\phi_3-\phi_4)^{-1}\sigma_v^2}{(1+\phi_1\phi_3+\phi_2+\phi_4+2\phi_2\phi_4+\phi_1^2\phi_4+\phi_2\phi_4^2-\phi_3^2-\phi_4^2-\phi_4^3-\phi_1\phi_3\phi_4)}\Big). \tag{A.40}$$

□

Proof of Lemma 4.1 Under Assumptions 4.1 and 4.2 with $s = 2$, the observed process $\{y_t^*\}$ is strictly stationarity and geometrically ergodic with $\mathbb{E}[y_t^*] < \infty$, $\mathbb{E}[|y_t|y_{t-1}] < \infty$ and $\mathbb{E}[y_t^* y_{t+k}^*] < \infty$. So the consistency is a simple application of the Law of Large Numbers for stationary and ergodic processes, *i.e.*, the Ergodic theorem; see Theorem 13.12 and Corollary 13.14 of Davidson (1994). □

Proof of Lemma 4.2 To establish the asymptotic normality of empirical moments, we shall use a CLT for dependent processes (see Davidson (1994), Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this CLT holds. We set

$$X_t := \begin{bmatrix} \Psi_t \\ \Lambda_t \\ \bar{\Psi}_t \end{bmatrix}, \quad \Psi_t := \log(y_t^2) - \mu, \quad \Lambda_t := [\Lambda_{t,0}, \Lambda_{t,1}, \ldots, \Lambda_{t,m}]', \tag{A.41}$$

$$\Lambda_{t,k} := y_t^* y_{t+k}^* - \gamma_{y^*}(k) = [\log(y_t^2) - \mu][\log(y_{t+k}^2) - \mu] - \gamma_{y^*}(k), \quad k = 0, 1, \ldots, m, \tag{A.42}$$

$$\bar{\Psi}_t := [|y_t|y_{t-1}] - \lambda_y(1), \tag{A.43}$$

$$S_T := \sum_{t=1}^{T} X_t = \begin{bmatrix} \sum_{t=1}^{T} \Psi_t \\ \sum_{t=1}^{T} \Lambda_t \\ \sum_{t=1}^{T} \bar{\Psi}_t \end{bmatrix}, \tag{A.44}$$

and consider the subfields $\mathcal{F}_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$. We will now show that

$$T^{-1/2} S_T \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} & C_{\mu,\lambda_y(1)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} & C_{\Gamma(m),\lambda_y(1)} \\ C_{\mu,\lambda_y(1)} & C'_{\Gamma(m),\lambda_y(1)} & V_{\lambda_y(1)} \end{bmatrix}\right), \tag{A.45}$$

which in turn yields (4.2). To do this, we will check the following conditions:

(i) $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic;

(ii) $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1;

(iii) $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\|S_T\| < \infty$, where $\|\cdot\|$ is the Euclidean norm.

(i) The fact that $\{X_t, \mathcal{F}_t\}$ is stationary and ergodic follows from results of Ahsan and Dufour (2021) regarding Stationarity, Ergodicity and Beta mixing.

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

$$\mathbb{E}|\log(y_t^2) - \mu| = \mathbb{E}|y_t^*| \le (\mathbb{E}|y_t^*|^2)^{1/2} = (\mathbb{E}[y_t^{*2}])^{1/2} = \sqrt{\gamma_{y^*}(0)} < \infty, \tag{A.46}$$

$$\begin{aligned} \mathbb{E}|y_t^* y_{t+k}^* - \gamma_{y^*}(k)| &= \mathbb{E}|y_t^* y_{t+k}^*| - |\gamma_{y^*}(k)| \le \mathbb{E}|y_t^* y_{t+k}^*| \\ &\le (\mathbb{E}|y_t^*|^2)^{1/2}(\mathbb{E}|y_{t+k}^*|^2)^{1/2} = (\mathbb{E}[y_t^{*2}])^{1/2}(\mathbb{E}[y_{t+k}^{*2}])^{1/2} \\ &= \mathbb{E}[y_t^{*2}] = \gamma_{y^*}(0) < \infty, \quad \text{for } k = 0, 1, \ldots, m, \end{aligned} \tag{A.47}$$

$$\begin{aligned} \mathbb{E}||y_t|y_{t-1} - \lambda_y(1)| &= \mathbb{E}|y_t y_{t-1}| - |\lambda_y(1)| \le \mathbb{E}|y_t y_{t-1}| \\ &\le (\mathbb{E}|y_t|^2)^{1/2}(\mathbb{E}|y_{t-1}|^2)^{1/2} \\ &= \mathbb{E}[y_t^2] = \gamma_y(0) < \infty, \end{aligned} \tag{A.48}$$

where the inequality in (A.46) is the application of Lyapunov's inequality and the inequalities in (A.47)-(A.48) follows from the Hölder's inequality.

(ii) - (2) We now show that $\{X_t, \mathcal{F}_t\}$ is a L_1-mixingale of size -1. From the discussion in Ahsan and Dufour (2021), we know that X_t is β-mixing, so it has mixing coefficients of the type $\beta_T = \psi\rho^T$, $\psi > 0$, $0 < \rho < 1$. To show that $\{X_t\}$ is of size -1, its mixing coefficients β_T must be $O(T^{-\varphi})$, with $\varphi > 1$ [see Davidson (1994, Definition 16.1, p. 247)]. Indeed,

$$\frac{\rho^T}{T^{-\varphi}} = T^\varphi \exp(T\log\rho) = \exp(\varphi\log T)\exp(T\log\rho) = \exp[\varphi(\log T) + T(\log\rho)]. \tag{A.49}$$

Since $\lim_{T\to\infty}[\varphi(\log T) + T(\log\rho)] = -\infty$, we get

$$\lim_{T\to\infty}\exp[\varphi(\log T) + T(\log\rho)] = 0. \tag{A.50}$$

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

(iii) To show that $\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\|S_T\| < \infty$, we first observe that $\mathbb{E}(S_T) = 0$ and, using the Cauchy-Schwarz inequality,

$$
\begin{aligned}
(T^{-1/2}\mathbb{E}\|S_T\|)^2 &\leq \frac{1}{T}\mathbb{E}(\|S_T\|^2) = \frac{1}{T}\mathbb{E}(S_T' S_T) = \frac{1}{T}\mathrm{tr}[\mathbb{E}(S_T S_T')] = \frac{1}{T}\mathrm{tr}[\mathrm{Var}(S_T)] \\
&= \mathrm{tr}[\mathrm{Var}(T^{-1/2}S_T)]. \qquad (A.51)
\end{aligned}
$$

It is thus sufficient to show that

$$
\limsup_{T\to\infty} \mathrm{tr}[\mathrm{Var}(T^{-1/2}S_T)] < \infty. \qquad (A.52)
$$

We now consider separately the components Ψ_t and Λ_t of X_t.

(iii) - (1) Set

$$
S_{T1} := \sum_{t=1}^{T} \Psi_t, \quad \zeta_\Psi(\tau) := \mathrm{cov}(\Psi_t, \Psi_{t+\tau}). \qquad (A.53)
$$

Then

$$
\zeta_\Psi(\tau) = \mathbb{E}[(\log(y_t^2) - \mu)(\log(y_{t+\tau}^2) - \mu)] = \mathbb{E}[y_t^* y_{t+\tau}^*] = \gamma_{y^*}(\tau), \qquad (A.54)
$$

$$
\begin{aligned}
\mathrm{Var}(T^{-1/2}S_{T1}) &= \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(\Psi_t) + \sum_{t\neq s}\mathrm{cov}(\Psi_t,\Psi_s)\right] = \frac{1}{T}\left[T\zeta_\Psi(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_\Psi(\tau)\right] \\
&= \zeta_\Psi(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_\Psi(\tau) = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau) \qquad (A.55)
\end{aligned}
$$

hence

$$
\begin{aligned}
\limsup_{T\to\infty}\mathrm{Var}(T^{-1/2}S_{T1}) &= \limsup_{T\to\infty}[\gamma_{y^*}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_{y^*}(\tau)] \\
&= \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau) = \sum_{\tau=-\infty}^{\infty}\gamma_{y^*}(\tau) \leq \sum_{\tau=-\infty}^{\infty}|\gamma_{y^*}(\tau)| < \infty. \qquad (A.56)
\end{aligned}
$$

This convergence is due to the fact that y_t^* follows a stationary ARMA(p,p) process. So y_t^* can be viewed as an MA(∞) process with absolutely summable coefficients, which implies the absolute summability of autocovariances [see Hamilton (1994, chapter 3, page 52)]. By the Cauchy-Schwarz inequality, this entails

$$
\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T1}| < \infty. \qquad (A.57)
$$

(iii) - (2) Set

$$
S_{T2} := \sum_{t=1}^{T}\Lambda_t = [S_{T2,0}, S_{T2,1}, \ldots, S_{T2,m}]', \qquad (A.58)
$$

$$
S_{T2,k} := \sum_{t=1}^{T}\Lambda_{t,k}, \quad \zeta_{\Lambda_k}(\tau) := \mathrm{cov}(\Lambda_{t,k}, \Lambda_{t+\tau,k}), \quad k = 0, 1, \ldots, m. \qquad (A.59)
$$

Then, for $k = 0, 1, \ldots, m$,

$$
\begin{aligned}
\zeta_{\Lambda_k}(\tau) &= \mathbb{E}[\big(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\big)\big(y_{t+\tau}^* y_{t+\tau+k}^* - \gamma_{y^*}(k)\big)] = \mathbb{E}[y_t^* y_{t+k}^* y_{t+\tau}^* y_{t+\tau+k}^*] - \gamma_{y^*}(k)^2 \\
&= \mathbb{E}[y_t^* y_{t+k}^*]\mathbb{E}[y_{t+\tau}^* y_{t+\tau+k}^*] + \mathrm{cov}(y_t^*, y_{t+\tau}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau+k}^*) \\
&\quad + \mathrm{cov}(y_t^*, y_{t+\tau+k}^*)\mathrm{cov}(y_{t+k}^*, y_{t+\tau}^*) - \gamma_{y^*}(k)^2 \\
&= \gamma_{y^*}(k)^2 + \gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k) - \gamma_{y^*}(k)^2
\end{aligned}
$$

$$= \quad \gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k), \tag{A.60}$$

hence

$$\mathrm{Var}(T^{-1/2}S_{T2,k}) = \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(\Lambda_{t,k}) + \sum_{t\neq s}\mathrm{cov}(\Lambda_{t,k},\Lambda_{s,k})\right] = \frac{1}{T}\left[T\zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\Lambda_k}(\tau)\right] \tag{A.61}$$

$$= \zeta_{\Lambda_k}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\Lambda_k}(\tau) \tag{A.62}$$

$$= \gamma_{y^*}(0)^2 + \gamma_{y^*}(k)\gamma_{y^*}(-k) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})[\gamma_{y^*}(\tau)^2 + \gamma_{y^*}(\tau+k)\gamma_{y^*}(\tau-k)], \tag{A.63}$$

and

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

This convergence is due to the fact that absolute summability implies square-summability. We deduce that

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\left|S_{T2,k}\right| < \infty, \quad k = 0, 1, \ldots, m. \tag{A.65}$$

(iii) - (3) Set

$$S_{T3} := \sum_{t=1}^{T}\bar{\Psi}_t, \quad \zeta_{\bar{\Psi}}(\tau) := \mathrm{cov}(\bar{\Psi}_t, \bar{\Psi}_{t+\tau}). \tag{A.66}$$

Then

$$\begin{aligned}
\zeta_{\bar{\Psi}}(\tau) &= \mathbb{E}\left[\left(|y_t|y_{t-1} - \lambda_y(1)\right)\left(|y_{t+\tau}|y_{t+\tau-1} - \lambda_y(1)\right)\right], \\
&= \mathbb{E}\left[|y_t|y_{t-1}|y_{t+\tau}|y_{t+\tau-1}\right] - \lambda_y^2(1), \\
&= \mathbb{E}\left[|y_t|y_{t-1}\right]\mathbb{E}\left[|y_{t+\tau}|y_{t+\tau-1}\right] + \mathbb{E}\left[|y_t||y_{t+\tau}|\right]\underbrace{\mathbb{E}\left[y_{t-1}y_{t+\tau-1}\right]}_{=0} + \mathbb{E}\left[|y_t|y_{t+\tau-1}\right]\mathbb{E}\left[y_{t-1}|y_{t+\tau}|\right] - \lambda_y^2(1), \\
&= \lambda_y^2(1) + 0 + \lambda_y(1-\tau)\lambda_y(1+\tau) - \lambda_y^2(1) = \lambda_y(1-\tau)\lambda_y(1+\tau),
\end{aligned} \tag{A.67}$$

where $\lambda_y(1-\tau) = 0$ for $\tau \geq 1$, hence

$$\begin{aligned}
\mathrm{Var}(T^{-1/2}S_{T3}) &= \frac{1}{T}\left[\sum_{t=1}^{T}\mathrm{Var}(\bar{\Psi}_t) + \sum_{t\neq s}\mathrm{cov}(\bar{\Psi}_t,\bar{\Psi}_s)\right] = \frac{1}{T}\left[T\zeta_{\bar{\Psi}}(0) + 2\sum_{\tau=1}^{T}(T-\tau)\zeta_{\bar{\Psi}}(\tau)\right] \\
&= \zeta_{\bar{\Psi}}(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\zeta_{\bar{\Psi}}(\tau) = \zeta_{\bar{\Psi}}(0) + 0 = \lambda_y^2(1),
\end{aligned} \tag{A.68}$$

and

$$\limsup_{T\to\infty} \mathrm{Var}(T^{-1/2}S_{T3}) = \limsup_{T\to\infty} \left[\lambda_y^2(1)\right] = \lambda_y^2(1) < \infty. \tag{A.69}$$

By the Cauchy-Schwarz inequality, this entails

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}|S_{T3}| < \infty. \tag{A.70}$$

Combining (A.57), (A.65) and (A.70), we get, for any $(m+3)\times 1$ fixed real vector $a \neq 0$,

$$\limsup_{T\to\infty} T^{-1/2}\mathbb{E}\left|a'S_T\right| < \infty. \tag{A.71}$$

It is also clear properties (i) and (ii) also hold if we replace S_T by $a'S_T$. Thus we can apply Theorem 24.5 of Davidson (1994) to $a'S_T$ to state that $T^{-1/2}(a'S_T)$ is asymptotically normal. Since this holds for any $a \neq 0$, it follows from the Cramér-Wold theorem that $T^{-1/2}\sum_{t=1}^{T} X_t$ is asymptotically multinormal:

$$T^{-1/2}S_T = T^{-1/2}\sum_{t=1}^{T} X_t = \sqrt{T}\begin{bmatrix} \hat{\mu}-\mu \\ \hat{\Gamma}(m)-\Gamma(m) \\ \hat{\lambda}_y(1)-\lambda_y(1) \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, V), \tag{A.72}$$

where

$$V = \lim_{T\to\infty} \mathbb{E}\{[T^{-1/2}S_T][T^{-1/2}S_T]'\} = \begin{bmatrix} V_\mu & C'_{\mu,\Gamma(m)} & C_{\mu,\lambda_y(1)} \\ C_{\mu,\Gamma(m)} & V_{\Gamma(m)} & C_{\Gamma(m),\lambda_y(1)} \\ C_{\mu,\lambda_y(1)} & C'_{\Gamma(m),\lambda_y(1)} & V_{\lambda_y(1)} \end{bmatrix}. \tag{A.73}$$

Using (A.56), (A.63) and (A.69), we have:

$$V_\mu = \gamma_{y^*}(0) + 2\sum_{\tau=1}^{\infty}\gamma_{y^*}(\tau), \quad V_{\lambda_y(1)} = \mathrm{Var}(\bar{\Psi}_t) + 2\sum_{\tau=1}^{\infty}\mathrm{cov}(\bar{\Psi}_t, \bar{\Psi}_{t+\tau}) = \lambda_y^2(1), \tag{A.74}$$

$$V_{\Gamma(m)} = \mathrm{Var}(\Lambda_t) + 2\sum_{\tau=1}^{\infty}\mathrm{cov}(\Lambda_t, \Lambda_{t+\tau}), \quad C_{\mu,\Gamma(m)} = [C_{\mu 0}, C_{\mu 1}, \ldots, C_{\mu m}]', \tag{A.75}$$

$$\begin{aligned} C_{\mu k} &= \sum_t \mathrm{cov}(\Psi_t, \Lambda_{t,k}) = 2\sum_{t=1}^{\infty}\mathbb{E}[\Psi_t(y_t^* y_{t+k}^* - \gamma_{y^*}(k))] \\ &= 2\sum_{t=1}^{\infty}\mathbb{E}[y_t^*(y_t^* y_{t+k}^* - \gamma_{y^*}(k))] = 2\sum_{t=1}^{\infty}[\mathbb{E}(y_t^{*2}y_{t+k}^*) - \mathbb{E}(y_t^*)\gamma_{y^*}(k)] \\ &= 2\sum_{t=1}^{\infty}\mathbb{E}(y_t^{*2}y_{t+k}^*), \quad k = 0, 1, 2, \ldots, m. \end{aligned} \tag{A.76}$$

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

$$\bar{c} := C_{\mu 0} = 2\sum_{t=1}^{\infty}\mathbb{E}(y_t^{*3}) = 2\sum_{t=1}^{\infty}[\mathbb{E}(w_t^3) + \mathbb{E}(\epsilon_t^3)] = 2\sum_{t=1}^{\infty}\mathbb{E}(\epsilon_t^3). \tag{A.77}$$

Since $\{z_t\}$ is a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables, we have $\mathbb{E}(\epsilon_t^3) = \psi^{(2)}(\frac{1}{2})$ [see (2.8)], which is equal to $-14\mathcal{Z}(3)$ where $\mathcal{Z}(\cdot)$ is Riemann's Zeta function with $\mathcal{Z}(3) = 1.20205$.[1] For $k = 1, \ldots, m$, it is easily seen that $C_{\mu k} = 0$

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

from the MA(∞) representation of w_t. So $C_{\mu,\Gamma(m)}$ is a $(m+1)\times 1$ vector given by $(\bar{c}, 0_{[m\times 1]})'$, with $\bar{c}$ is defined in (A.77). Now

$$
\begin{aligned}
C_{\mu,\lambda_y(1)} &= \sum_t \mathrm{cov}(\Psi_t, \bar{\Psi}_t) = 2\sum_{t=1}^{\infty} \mathbb{E}[y_t^*\left(|y_t|y_{t-1} - \lambda_y(1)\right)] \\
&= 2\sum_{t=1}^{\infty} \mathbb{E}[y_t^*|y_t|y_{t-1}] - \underbrace{\mathbb{E}[y_t^*]}_{=0}\lambda_y(1) \\
&= 2\sum_{t=1}^{\infty} \mathbb{E}[(w_t+\epsilon_t)|y_t|y_{t-1}] = 2\sum_{t=1}^{\infty}\left(\mathbb{E}[w_t|y_t|y_{t-1}] + \mathbb{E}[\epsilon_t|y_t|y_{t-1}]\right) \\
&= 2\sum_{t=1}^{\infty}\left(\mathbb{E}[|y_t|w_t y_{t-1}] + \mathbb{E}[|y_t|y_{t-1}\epsilon_t]\right) \\
&= 2\sum_{t=1}^{\infty}\left(2\mathbb{E}[|y_t|]\mathrm{cov}(w_t, y_{t-1}) + \sigma_y^2\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|\exp\left(\frac{w_{t-1}}{2}\right)z_{t-1}\epsilon_t\right]\right) \\
&= 2\sum_{t=1}^{\infty}\left(2\mathbb{E}[|y_t|]\mathbb{E}[w_t, y_{t-1}] + \sigma_y^2\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)\exp\left(\frac{w_{t-1}}{2}\right)z_{t-1}\right]\mathbb{E}[|z_t|\epsilon_t]\right) \\
&= 2\sum_{t=1}^{\infty}\left(2\sigma_y\underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)\right]}_{=\exp\left(\frac{\gamma_w(0)}{8}\right)}\mathbb{E}[|z_t|]\sigma_y\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right)z_{t-1}w_t\right]\right. \\
&\qquad \left. +\sigma_y^2\underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}+\frac{w_{t-1}}{2}\right)z_{t-1}\right]}_{=\frac{\delta_p\sigma_v}{2}\exp\left(\frac{\gamma_w(0)+\gamma_w(1)}{4}\right)}\left(\underbrace{\mathbb{E}[|z_t|\log(z_t^2)]}_{=\sqrt{\frac{2}{\pi}}(\log 2-\gamma^\epsilon)} - \underbrace{\mathbb{E}[|z_t|]}_{=\sqrt{\frac{2}{\pi}}}\ \underbrace{\mathbb{E}[\log(z_t^2)]}_{=\psi^{(0)}(0.5)+\log 2}\right)\right) \\
&= 2\sum_{t=1}^{\infty}\left(4\sigma_y^2\exp\left(\frac{\gamma_w(0)}{8}\right)\sqrt{\frac{2}{\pi}}\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right)\right]\underbrace{\mathrm{cov}(z_{t-1}, w_t)}_{=\mathbb{E}[z_{t-1}w_t]=\delta_p\sigma_v}\right. \\
&\qquad \left. +\sigma_y^2\frac{\delta_p\sigma_v}{2}\exp\left(\frac{\gamma_w(0)+\gamma_w(1)}{4}\right)\sqrt{\frac{2}{\pi}}\left(-\gamma^\epsilon - \psi^{(0)}(0.5)\right)\right) \\
&= 2\sum_{t=1}^{\infty}\left(4\sqrt{\frac{2}{\pi}}\delta_p\sigma_v\sigma_y^2\exp\left(\frac{\gamma_w(0)}{8}\right)\exp\left(\frac{\gamma_w(0)}{8}\right) + \sigma_y^2\frac{\delta_p\sigma_v}{2}\exp\left(\frac{\gamma_w(0)+\gamma_w(1)}{4}\right)\sqrt{\frac{2}{\pi}}\left(-\gamma^\epsilon-\psi^{(0)}(0.5)\right)\right) \\
&= 2\sum_{t=1}^{\infty}\left(4\sqrt{\frac{2}{\pi}}\delta_p\sigma_v\sigma_y^2\exp\left(\frac{\gamma_w(0)}{4}\right) + \sigma_y^2\delta_p\sigma_v\exp\left(\frac{\gamma_w(0)+\gamma_w(1)}{4}\right)\sqrt{\frac{1}{2\pi}}\left(-\gamma^\epsilon-\psi^{(0)}(0.5)\right)\right) \\
&= 2\sum_{t=1}^{\infty}\frac{\delta_p\sigma_v\sigma_y^2}{\sqrt{2\pi}}\exp\left(\frac{\gamma_w(0)}{4}\right)\left[8-\left(\gamma^\epsilon+\psi^{(0)}(0.5)\right)\exp\left(\frac{\gamma_w(1)}{4}\right)\right], \qquad \text{(A.78)}
\end{aligned}
$$

where $\gamma_w(k) = \mathrm{cov}(w_t, w_{t+k})$, $\psi^{(0)}(z)$ is the *digamma function* and γ^ϵ is the Euler-Mascheroni constant.

$$
C_{\Gamma(m),\lambda_y(1)} = [C_{\lambda 0}, C_{\lambda 1}, \dots, C_{\lambda m}]', \qquad \text{(A.79)}
$$

$$
\begin{aligned}
C_{\lambda k} &= \sum_t \mathrm{cov}(\bar{\Psi}_t, \Lambda_{t,k}) = 2\sum_{t=1}^{\infty}\mathbb{E}[\left(|y_t|y_{t-1} - \lambda_y(1)\right)\left(y_t^* y_{t+k}^* - \gamma_{y^*}(k)\right)] \\
&= 2\sum_{t=1}^{\infty}\left(\mathbb{E}\left[|y_t|y_{t-1}y_t^* y_{t+k}^*\right] - \lambda_y(1)\gamma_{y^*}(k)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= 2\sum_{t=1}^{\infty}\Big(\mathbb{E}[|y_t|]\mathrm{cov}(y_{t-1}, y_t^* y_{t+k}^*) + \mathrm{cov}(|y_t|, y_{t-1})\mathrm{cov}(y_t^*, y_{t+k}^*) \\
&\quad + \mathrm{cov}(|y_t|, y_t^*)\mathrm{cov}(y_{t-1}, y_{t+k}^*) + \mathrm{cov}(|y_t|, y_{t+k}^*)\mathrm{cov}(y_{t-1}, y_t^*) - \lambda_y(1)\gamma_{y^*}(k)\Big) \\
&= 2\sum_{t=1}^{\infty}\Big(\mathbb{E}[|y_t|]\mathbb{E}[y_{t-1} y_t^* y_{t+k}^*] + \mathbb{E}[|y_t| y_{t-1}]\mathbb{E}[y_t^*, y_{t+k}^*] \\
&\quad + \mathbb{E}[|y_t|, y_t^*]\mathbb{E}[y_{t-1}, y_{t+k}^*] + \mathbb{E}[|y_t| y_{t+k}^*]\mathbb{E}[(y_{t-1} y_t^*] - \lambda_y(1)\gamma_{y^*}(k)\Big) \\
&= 2\sum_{t=1}^{\infty}\Big(\mathbb{E}[|y_t|]\mathbb{E}[y_{t-1} y_t^* y_{t+k}^*] + \mathbb{E}[|y_t| y_t^*]\mathbb{E}[y_{t-1} y_{t+k}^*] \\
&\quad + \mathbb{E}[|y_t| y_{t+k}^*]\mathbb{E}[(y_{t-1} y_t^*]\Big), \quad k = 0, 1, 2, \ldots, m.
\end{aligned} \tag{A.80}
$$

Now

$$
\begin{aligned}
&\mathbb{E}[|y_t|]\mathbb{E}[y_{t-1} y_t^* y_{t+k}^*] = \mathbb{E}[|y_t|]\mathbb{E}\left[y_{t-1}(w_t + \epsilon_t)(w_{t+k} + \epsilon_{t+k})\right] \\
&= \mathbb{E}[|y_t|]\mathbb{E}\left[y_{t-1} w_t w_{t+k}\right] = \mathbb{E}[|y_t|]\mathbb{E}\left[\sigma_y \exp\left(\frac{w_{t-1}}{2}\right) z_{t-1} w_t w_{t+k}\right] \\
&= \sigma_y \mathbb{E}\left[\sigma_y \exp\left(\frac{w_t}{2}\right)|z_t|\right]\left[\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right) z_{t-1}\right]\mathbb{E}\left[w_t w_{t+k}\right] + \mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right) w_t\right]\mathbb{E}\left[z_{t-1} w_{t+k}\right]\right. \\
&\quad \left. + \mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right) w_{t+k}\right]\mathbb{E}\left[z_{t-1} w_t\right]\right] \\
&= \sigma_y^2 \underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)\right]}_{=\exp\left(\frac{\gamma_w(0)}{8}\right)} \underbrace{\mathbb{E}\left[|z_t|\right]}_{=\sqrt{\frac{2}{\pi}}} \left(\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right)\right]\underbrace{\mathbb{E}\left[z_{t-1}\right]}_{=0}\mathbb{E}\left[w_t w_{t+k}\right] + \underbrace{\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right) w_t\right]}_{=\frac{\gamma_w(1)}{2}\exp\left(\frac{\gamma_w(0)}{8}\right)} \underbrace{\mathbb{E}\left[z_{t-1} w_{t+k}\right]}_{=\delta_p \sigma_v \bar{\psi}_k}\right. \\
&\quad \left. + \underbrace{\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right) w_{t+k}\right]}_{=\frac{\gamma_w(k+1)}{2}\exp\left(\frac{\gamma_w(0)}{8}\right)} \underbrace{\mathbb{E}\left[z_{t-1} w_t\right]}_{=\delta_p \sigma_v}\right) \\
&= \sigma_y^2 \exp\left(\frac{\gamma_w(0)}{8}\right)\sqrt{\frac{2}{\pi}}\left(\delta_p \sigma_v \bar{\psi}_k \frac{\gamma_w(1)}{2}\exp\left(\frac{\gamma_w(0)}{8}\right) + \delta_p \sigma_v \frac{\gamma_w(k+1)}{2}\exp\left(\frac{\gamma_w(0)}{8}\right)\right) \\
&= \delta_p \sigma_v \sigma_y^2 \sqrt{\frac{1}{2\pi}}\exp\left(\frac{\gamma_w(0)}{4}\right)\left(\bar{\psi}_k \gamma_w(1) + \gamma_w(k+1)\right),
\end{aligned} \tag{A.81}
$$

where $\bar{\psi}_k$ is k-th parameter of the MA(∞) representation the latent log volatility process,

$$
\begin{aligned}
&\mathbb{E}[|y_t| y_t^*]\mathbb{E}[y_{t-1} y_{t+k}^*] = \mathbb{E}[\sigma_y \exp\left(\frac{w_t}{2}\right)|z_t|(w_t + \epsilon_t)]\mathbb{E}[y_{t-1}(w_{t+k} + \epsilon_{t+k})] \\
&= \sigma_y \mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|(w_t + \epsilon_t)\right]\left(\mathbb{E}[y_{t-1} w_{t+k}] + \mathbb{E}[y_{t-1}]\underbrace{\mathbb{E}[\epsilon_{t+k})]}_{=0}\right) \\
&= \sigma_y^2\left(\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t| w_t\right] + \mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|\log(z_t^2)\right] - \mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|\right]\mathbb{E}\left[\log(z_t^2)\right]\right)\left(\mathbb{E}[\exp\left(\frac{w_{t-1}}{2}\right) z_{t-1} w_{t+k}]\right) \\
&= \sigma_y^2 \left(\underbrace{\mathbb{E}\left[|z_t|\right]}_{=\sqrt{\frac{2}{\pi}}} \underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right) w_t\right]}_{=\frac{\gamma_w(0)}{2}\exp\left(\frac{\gamma_w(0)}{8}\right)} + \underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)\right]}_{=\exp\left(\frac{\gamma_w(0)}{8}\right)} \underbrace{\mathbb{E}\left[|z_t|\log(z_t^2)\right]}_{=\sqrt{\frac{2}{\pi}}(\log 2 - \gamma^{e})}\right.
\end{aligned}
$$

$$
-\underbrace{\mathbb{E}\left[|z_t|\right]}_{=\sqrt{\frac{2}{\pi}}}\underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)\right]}_{=\exp\left(\frac{\gamma_w(0)}{8}\right)}\underbrace{\mathbb{E}\left[\log(z_t^2)\right]}_{=\psi^{(0)}(0.5)+\log 2}\Bigg)\left(\underbrace{\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right)\right]}_{=\exp\left(\frac{\gamma_w(0)}{8}\right)}\underbrace{\operatorname{cov}(z_{t-1},w_{t+k})}_{=\mathbb{E}[z_{t-1}w_{t+k}]=\delta_p\sigma_v\bar{\psi}_k}\right)
$$

$$
= \delta_p\sigma_v\sigma_y^2\bar{\psi}_k\sqrt{\frac{2}{\pi}}\exp\left(\frac{\gamma_w(0)}{4}\right)\left(\frac{\gamma_w(0)}{2}-\gamma^\epsilon-\psi^{(0)}(0.5)\right), \tag{A.82}
$$

$$
\begin{aligned}
\mathbb{E}[|y_t|y_{t+k}^*]\mathbb{E}[y_{t-1}y_t^*] &= \mathbb{E}[\sigma_y\exp\left(\frac{w_t}{2}\right)|z_t|(w_{t+k}+\epsilon_{t+k})]\mathbb{E}[y_{t-1}(w_t+\epsilon_t)] \\
&= \sigma_y\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|(w_{t+k}+\epsilon_{t+k})\right]\left(\mathbb{E}[y_{t-1}w_t]+\mathbb{E}[y_{t-1}]\underbrace{\mathbb{E}[\epsilon_t)]}_{=0}\right) \\
&= \sigma_y^2\left(\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|w_{t+k}\right]+\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)|z_t|\right]\underbrace{\mathbb{E}[\epsilon_{t+k})]}_{=0}\right)\left(\mathbb{E}[\exp\left(\frac{w_{t-1}}{2}\right)z_{t-1}w_t]\right) \\
&= \sigma_y^2\left(\underbrace{\mathbb{E}\left[|z_t|\right]}_{=\sqrt{\frac{2}{\pi}}}\underbrace{\mathbb{E}\left[\exp\left(\frac{w_t}{2}\right)w_{t+k}\right]}_{=\frac{\gamma_w(k)}{2}\exp\left(\frac{\gamma_w(0)}{8}\right)}\right)\left(2\underbrace{\mathbb{E}\left[\exp\left(\frac{w_{t-1}}{2}\right)\right]}_{=\exp\left(\frac{\gamma_w(0)}{8}\right)}\underbrace{\operatorname{cov}(z_{t-1},w_t)}_{=\mathbb{E}[z_{t-1}w_t]=\delta_p\sigma_v}\right) \\
&= \delta_p\sigma_v\sigma_y^2\bar{\psi}_k\sqrt{\frac{2}{\pi}}\exp\left(\frac{\gamma_w(0)}{4}\right)\gamma_w(k).
\end{aligned} \tag{A.83}
$$

Combing (A.81)-(A.83), we have

$$
\begin{aligned}
C_{\lambda k} &= 2\sum_{t=1}^{\infty}\left(\frac{\delta_p\sigma_v\sigma_y^2}{\sqrt{2\pi}}\exp\left(\frac{\gamma_w(0)}{4}\right)\left(\gamma_w(0)+\bar{\psi}_k\gamma_w(1)+0.5\bar{\psi}_k\gamma_w(k)+\gamma_w(k+1)-\frac{\gamma^\epsilon+\psi^{(0)}(0.5)}{2}\right)\right), \\
& k=0,1,2,\ldots,m.
\end{aligned} \tag{A.84}
$$

where $\bar{\psi}_k$ is k-th parameter of the MA(∞) representation of the latent log volatility process, $\gamma_w(k)=\operatorname{cov}(w_t,w_{t+k})$, $\psi^{(0)}(z)$ is the *digamma function* and γ^ϵ is the Euler-Mascheroni constant. Finally, (4.2) follows on observing that

$$
\sqrt{T}\begin{bmatrix}\hat{\mu}-\mu \\ \hat{\Gamma}(m)-\Gamma(m) \\ \hat{\lambda}_y(1)-\lambda_y(1)\end{bmatrix}-T^{-1/2}S_T \xrightarrow[T\to\infty]{p} 0. \tag{A.85}
$$

□

PROOF OF THEOREM 4.3 It is easily seen that D_p is a continuously differentiable mapping of $[\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\ldots,\gamma_{y^*}(2p)]'$. The convergence result stated in (4.9) follows from the standard result for differentiable transformations of asymptotically normally distributed variables together with the application of the multivariate delta method.

In case of an SVL(1) model,

$$
D_1 := D_1(\beta) = (D_{\phi_1}, D_{\sigma_y}, D_{\sigma_v}, D_{\delta_1})', \quad \beta := [\mu,\gamma_{y^*}(0),\gamma_{y^*}(1),\gamma_{y^*}(2),\lambda_y(1)]', \tag{A.86}
$$

$$D_{\phi_1} = \gamma_{y^*}(2)/\gamma_{y^*}(1)\,, \quad D_{\sigma_y} = \exp(\mu + 1.2704)^{1/2}\,, \quad D_{\sigma_v} = [1 - (\gamma_{y^*}(2)/\gamma_{y^*}(1))^2][\gamma_{y^*}(0) - \pi^2/2]\,, \tag{A.87}$$

$$D_{\delta_1} = \frac{\sqrt{2\pi}\lambda_y(1)}{\sigma_v \sigma_y^2} \exp\left(-\frac{1}{4}\frac{\sigma_v^2}{1-\phi_1}\right), \tag{A.88}$$

$$G(\beta) := \frac{\partial D_1}{\partial \beta'} = \begin{pmatrix} 0 & 0 & -\tilde{A} & \frac{1}{\gamma_{y^*}(1)} & 0 \\ \frac{\tilde{B}}{2} & 0 & 0 & 0 & 0 \\ 0 & -\frac{\tilde{C}}{2\tilde{F}} & \frac{\gamma_{y^*}(2)^2\tilde{D}}{\gamma_{y^*}(1)^3\tilde{F}} & -\frac{\gamma_{y^*}(2)\tilde{D}}{\gamma_{y^*}(1)^2\tilde{F}} & 0 \\ -\frac{\sqrt{2}\lambda_y(1)\sqrt{\pi}e^{-\tilde{E}}e^{-(\mu+1.2704)}}{\tilde{F}} & \tilde{J} & \tilde{K} & \tilde{L} & \frac{\sqrt{2}\sqrt{\pi}e^{-\tilde{E}}e^{-(\mu+1.2704)}}{\tilde{F}} \end{pmatrix}, \tag{A.89}$$

where

$$\tilde{A} = \frac{\gamma_{y^*}(2)}{\gamma_{y^*}(1)^2}\,, \quad \tilde{B} = \sqrt{e^{\mu+1.2704}}\,, \quad \tilde{C} = \frac{\gamma_{y^*}(2)^2}{\gamma_{y^*}(1)^2} - 1\,, \quad \tilde{D} = \gamma_{y^*}(0) - \frac{\pi^2}{2}\,, \quad \tilde{E} = \frac{\tilde{C}\tilde{D}}{4(\tilde{A}-1)} \tag{A.90}$$

$$\tilde{F} = \sqrt{-\tilde{C}\tilde{D}}, \quad \tilde{G} = \left(\gamma_{y^*}(1)^2 - \gamma_{y^*}(2)^2\right), \quad \tilde{H} = 2\gamma_{y^*}(0) - \pi^2 \tag{A.91}$$

$$\tilde{I} = \gamma_{y^*}(1)^2\pi^2 - 2\gamma_{y^*}(0)\gamma_{y^*}(1)^2 + 8\gamma_{y^*}(1)\gamma_{y^*}(2) - \gamma_{y^*}(2)^2\pi^2 + 2\gamma_{y^*}(0)\gamma_{y^*}(2)^2 \tag{A.92}$$

$$\tilde{J} = -\frac{\sqrt{2}\lambda_y(1)\sqrt{\pi}e^{-(\mu+1.2704)-\tilde{E}}\tilde{G}(4\gamma_{y^*}(1) + 2\gamma_{y^*}(0)\gamma_{y^*}(1) + 2\gamma_{y^*}(0)\gamma_{y^*}(2) - \gamma_{y^*}(1)\pi^2 - \gamma_{y^*}(2)\pi^2)}{8\gamma_{y^*}(1)^3\tilde{F}^{3/2}} \tag{A.93}$$

$$\tilde{K} = -\frac{\sqrt{2}\gamma_{y^*}(2)\lambda_y(1)\sqrt{\pi}e^{-(\mu+1.2704)-\tilde{E}}\tilde{H}\tilde{I}}{16\gamma_{y^*}(1)^4\tilde{F}^{3/2}}, \quad \tilde{L} = \frac{\sqrt{2}\lambda_y(1)\sqrt{\pi}e^{-(\mu+1.2704)-\tilde{E}}\tilde{H}\tilde{I}}{16\gamma_{y^*}(1)^3\tilde{F}^{3/2}}. \tag{A.94}$$

□

B Additional ARMA-based winsorized estimators

We can achieve better stability and efficiency of the CF-ARMA estimator by using "winsorization". Winsorization substantially increases the probability of getting admissible values. Using (3.23), we propose alternative estimators of $\boldsymbol{\phi_p}$ and we call these estimators *winsorized ARMA* estimators (or *W-ARMA* estimators). Other (possibly nonlinear) averaging methods, such as the median, may also be used. We consider four types of winsorized estimators based on the expression (3.23) in the simulation section.

1. The first winsorized estimate $\hat{\phi}_p^{\mathrm{m}}$ is the arithmetic mean of sample ratios (equal weights):

$$\omega_j = 1/J, \quad j = 1, \ldots, J, \tag{B.95}$$

in (3.23). This type of winsorization is also considered by Kristensen and Linton (2006) in the context of the GARCH$(1,1)$ model estimation.

2. The second estimate $\hat{\phi}_p^{\mathrm{ld}}$ is a mean of ratios with linearly declining (LD) weights:

$$\omega_j = (2/J)[1 - (j/(J+1))], \quad j = 1, \ldots, J. \tag{B.96}$$

3. The third estimate is the median-based estimate obtained by taking the median of J estimates of each one of the p components of ϕ_p:

$$\hat{\phi}_p^{\mathrm{med}} = [\hat{\phi}_1^{\mathrm{med}}, \ldots, \hat{\phi}_p^{\mathrm{med}}]', \quad \hat{\phi}_i^{\mathrm{med}} = \mathrm{med}\left\{\hat{\boldsymbol{B}}(p, j)_i : 1 \le j \le J\right\}, \; i = 1, \ldots, p. \tag{B.97}$$

4. The fourth estimate is obtained by an OLS regression coefficient (without intercept):

$$\hat{\phi}_p^{\mathrm{ols}} = [A(p, J)' A(p, J)]^{-1} A(p, J)' e(p, J) \tag{B.98}$$

where $e(p, J)$ is a $(pJ) \times 1$ vector and $A(p, J)$ is a $(pJ) \times p$ matrix defined by

$$e(p, J) = [\hat{\gamma}(p, 1)\omega_1^{1/2}, \ldots, \hat{\gamma}(p, J)\omega_J^{1/2}]', \quad A(p, J) = [\hat{\boldsymbol{\Gamma}}(p, 1)\omega_1^{1/2}, \ldots, \hat{\boldsymbol{\Gamma}}(p, J)\omega_J^{1/2}]'. \tag{B.99}$$

Clearly, different OLS-based W-ARMA can be generated by considering different weights $\omega_1, \ldots, \omega_J$.

All these estimators are depend on J. For $J = 1$, they are equivalent to the simple CF-ARMA estimator given in (3.15).

C Additional simulation results

Table A1. Empirical power of tests for no leverage in SVL(1) model (moderate persistence & LMC not level-corrected)

	H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$											
	$\delta = -0.90$			$\delta = -0.70$			$\delta = 0.70$			$\delta = 0.90$		
T	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC
	$\phi = 0.90$, $\sigma_y = 0.10$, $\sigma_v = 0.75$											
500	61.9	73.4	20.9	42.3	51.9	12.5	42.5	50.1	11.4	63.8	73.7	19.5
1000	84.9	92.8	36.8	59.8	75.4	19.7	61.9	75.7	18.7	85.6	93.7	35.4
2000	99.1	99.2	58.7	85.6	91.5	31.7	85.6	91.6	31.9	98.7	99.3	60.3
5000	100.0	100.0	94.9	99.8	99.9	70.6	99.9	100.0	68.4	100.0	100.0	93.9
	$\phi = 0.75$, $\sigma_y = 0.10$, $\sigma_v = 1.00$											
500	96.2	98.3	62.3	80.8	87.6	35.8	80.8	87.9	35.2	96.5	98.4	63.6
1000	100.0	99.9	93.3	99.0	99.4	70.5	98.8	99.4	72.8	100.0	100.0	92.9
2000	100.0	100.0	99.7	100.0	100.0	94.5	100.0	100.0	94.4	100.0	100.0	99.9
5000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N = 99$ simulations. Asymmetric test results are level-corrected using $N = 10000$ and true parameters (*i.e.*, infeasible test in practice as it requires knowing the true DGP). LMC test results are level-corrected using $N = 1,000$ and true parameters.

Table A2. Empirical power of tests for no leverage in SVL(1) model (high persistence & LMC not level-corrected)

T	$\delta=-0.90$			$\delta=-0.70$			$\delta=0.70$			$\delta=0.90$		
T	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC
						H_0: $\delta=0$ vs. H_1: $\delta\neq 0$						
						$\phi=0.99$, $\sigma_y=0.10$, $\sigma_v=0.25$						
500	23.4	26.1	8.4	19.4	21.5	7.9	22.8	25.0	6.2	25.6	29.2	8.1
1000	34.1	40.6	12.4	26.6	31.2	10.4	23.2	29.2	8.6	29.7	35.7	9.2
2000	43.7	50.6	16.5	30.2	35.8	12.0	30.7	37.8	12.4	44.3	50.9	17.9
5000	68.8	73.1	31.7	46.9	52.8	20.0	44.5	50.7	18.1	66.0	72.6	27.5
						$\phi=0.95$, $\sigma_y=0.10$, $\sigma_v=0.50$						
500	42.9	53.2	12.8	29.6	37.2	8.0	30.9	40.1	8.1	44.9	57.4	13.2
1000	67.9	80.6	26.2	46.3	61.2	14.6	44.4	59.3	13.0	67.1	81.8	24.1
2000	87.6	95.2	40.1	62.7	76.7	20.3	65.7	79.1	22.7	89.8	94.8	43.9
5000	100.0	100.0	82.3	96.4	97.9	48.0	96.0	98.1	43.7	100.0	99.8	79.9

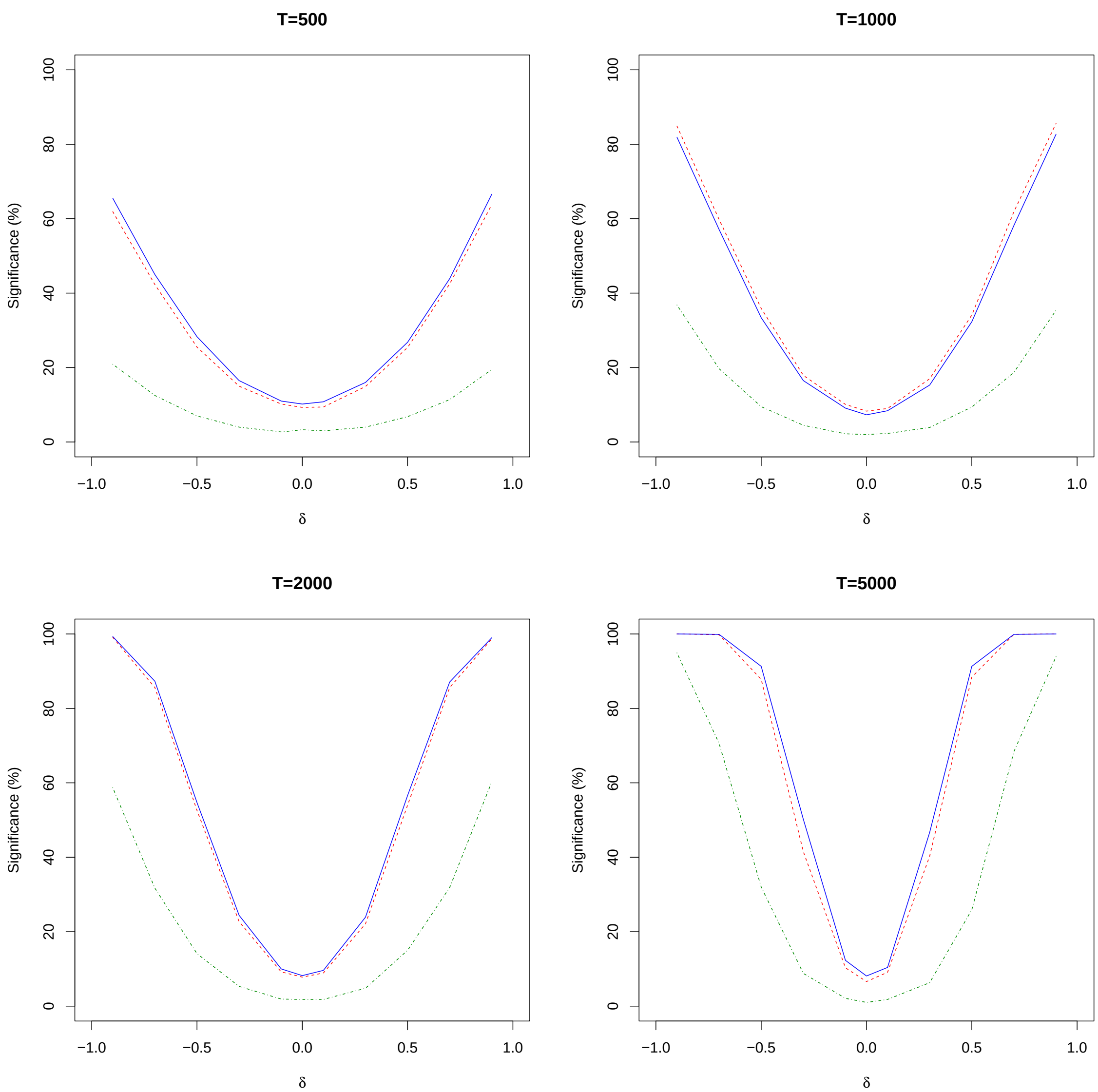
Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N=99$ simulations. Asymmetric test results are level-corrected using $N=10000$ and true parameters (*i.e.*, infeasible test in practice as it requires knowing the true DGP). LMC test results are level-corrected using $N=1{,}000$ and true parameters.

Table A3. Empirical power of tests for no leverage in SVL(2) model (LMC not level-corrected)

	$\delta=-0.90$			$\delta=-0.70$			$\delta=0.70$			$\delta=0.90$		
T	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC	Asy	LMC	MMC
						H_0: $\delta=0$ vs. H_1: $\delta\neq 0$						
						$\phi_1=0.05$, $\phi_2=0.85$, $\sigma_y=1.00$, $\sigma_v=1.00$						
500	62.4	67.5	47.0	49.4	56.9	34.1	50.0	55.2	34.3	62.1	65.8	45.2
1000	74.3	75.3	52.8	60.1	63.0	39.5	61.5	65.8	40.9	75.5	76.2	53.8
2000	90.3	88.1	67.4	80.9	80.3	52.5	80.9	80.2	49.9	91.0	87.4	66.3
5000	99.8	98.7	88.6	99.1	96.2	77.4	99.4	95.9	78.8	100.0	98.8	87.8
						$\phi_1=0.05$, $\phi_2=0.70$, $\sigma_y=1.00$, $\sigma_v=1.00$						
500	97.7	96.7	87.7	90.5	92.8	74.6	90.9	92.3	72.0	97.6	96.5	88.8
1000	100.0	99.2	96.6	99.8	98.7	92.7	99.7	98.7	92.5	100.0	99.4	96.7
2000	100.0	99.9	99.4	100.0	99.9	98.4	100.0	99.9	97.8	100.0	99.9	98.9
5000	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0

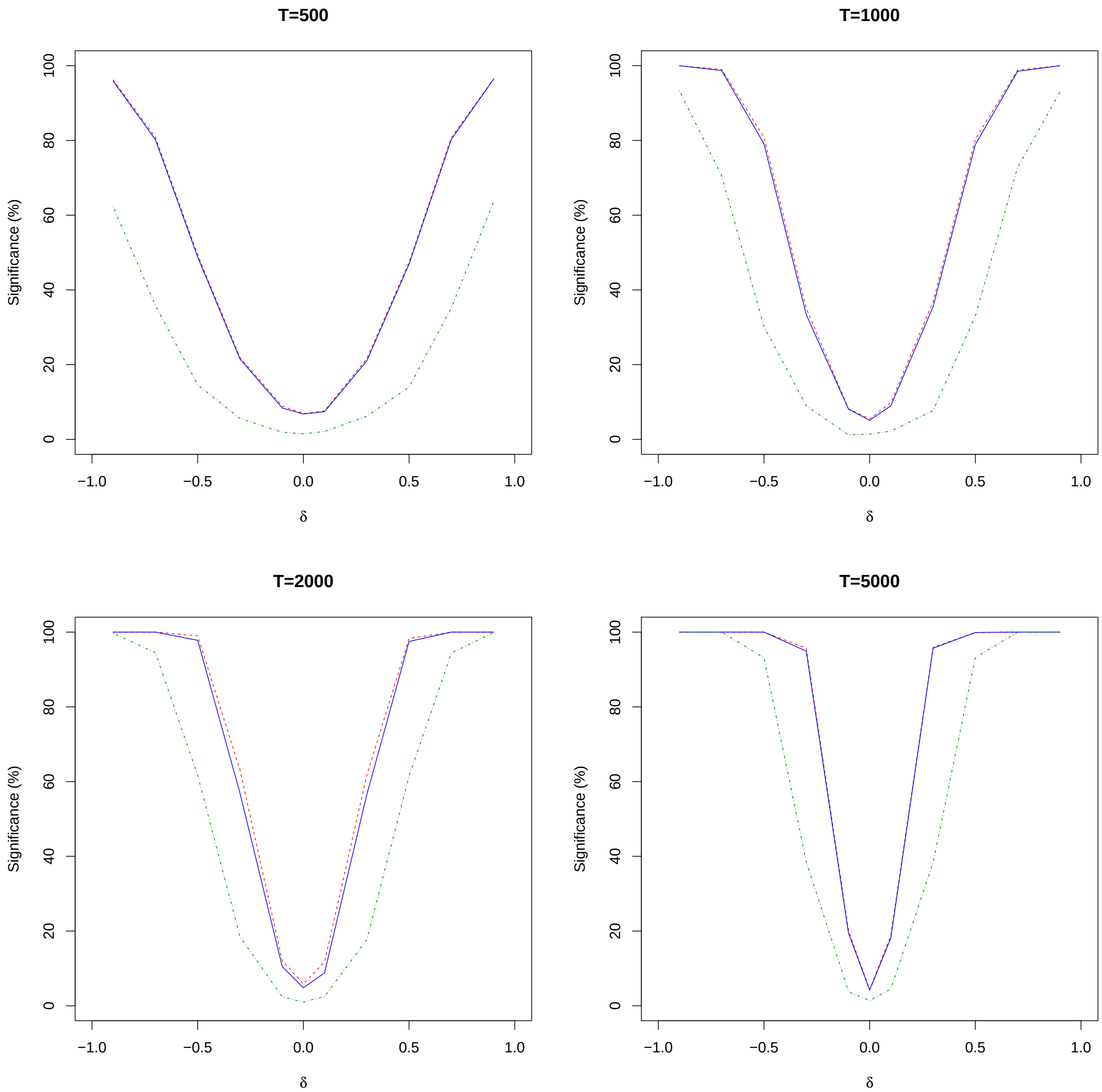
Notes: Rejection frequencies are obtained using 1000 replications. Monte Carlo tests use $N=99$ simulations. Asymmetric test results are level-corrected using $N=10000$ and true parameters (*i.e.*, infeasible test in practice as it requires knowing the true DGP). LMC test results are level-corrected using $N=1{,}000$ and true parameters.

Figure A1. Power curves of test for no leverage in SV(1) model when $\phi = 0.90$, $\sigma_y = 0.10$, and $\sigma_v = 0.75$



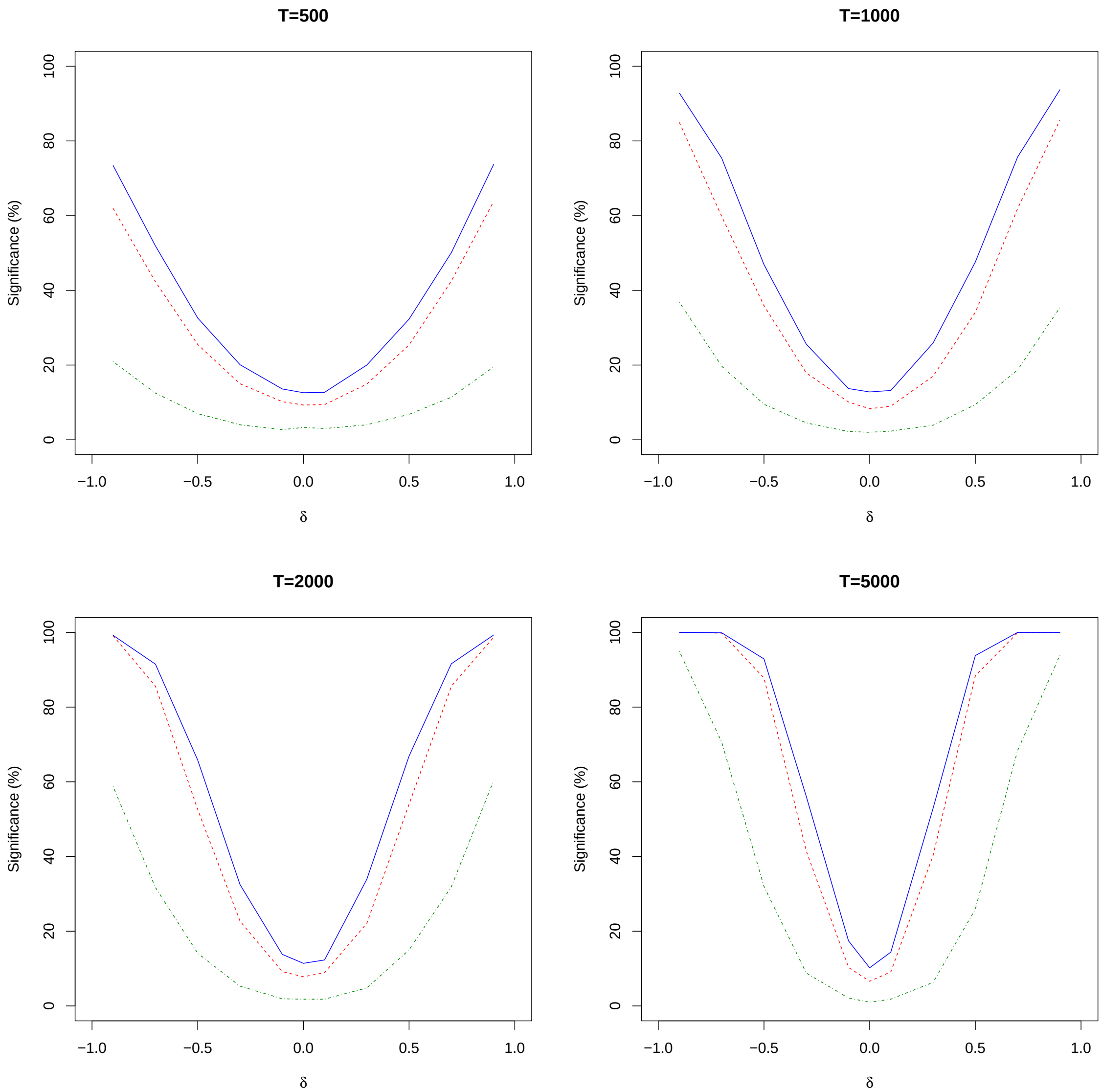
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test (level corrected), and green is the Maximized Monte Carlo test.

Figure A2. Power curves of test for no leverage in SV(1) model when $\phi = 0.75$, $\sigma_y = 0.10$, and $\sigma_v = 1.00$



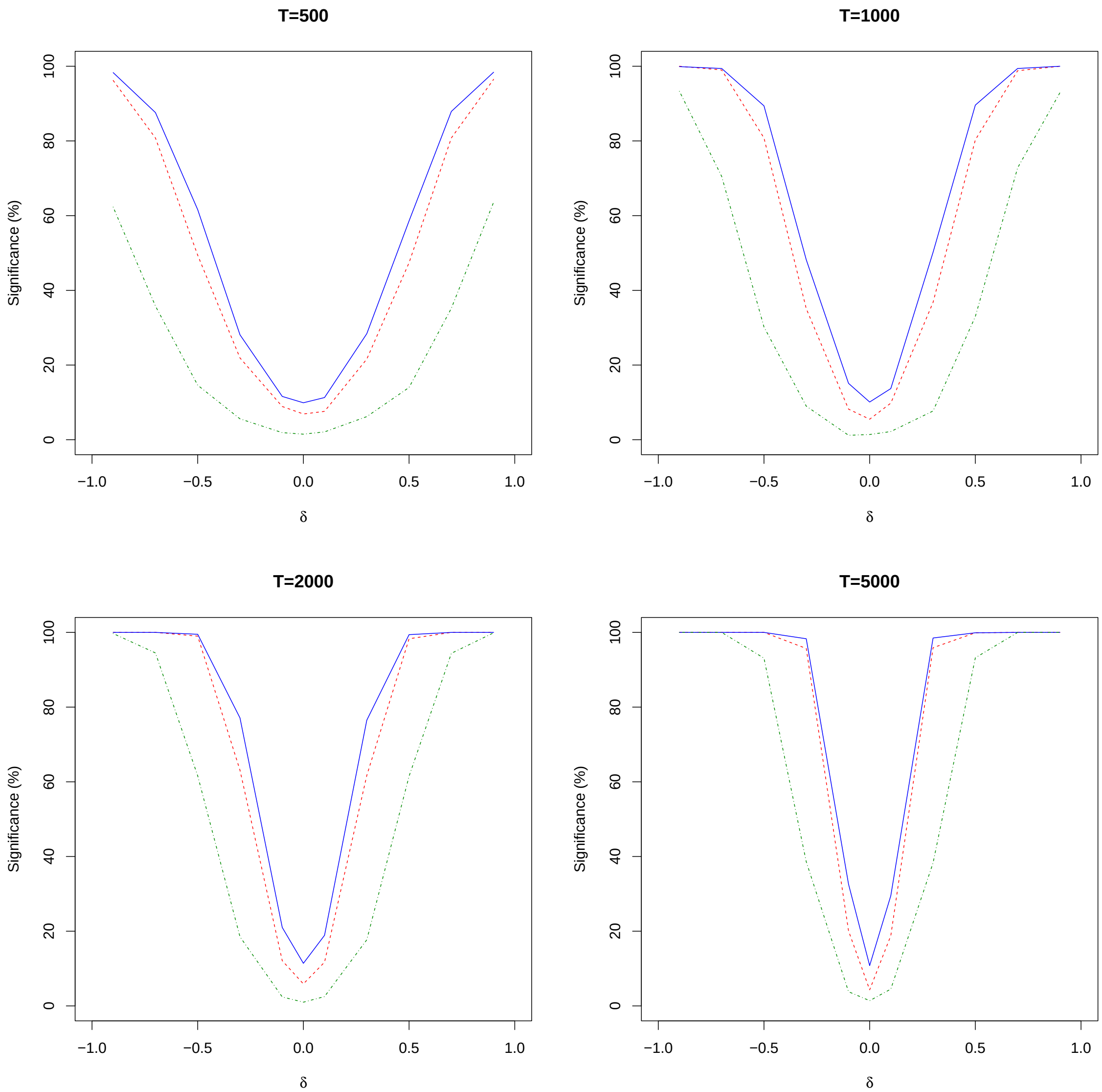
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test (level corrected), and green is the Maximized Monte Carlo test.

Figure A3. Power curves of test for no leverage in SV(1) model when $\phi = 0.90$, $\sigma_y = 0.10$, and $\sigma_v = 0.75$ (LMC not level-corrected)



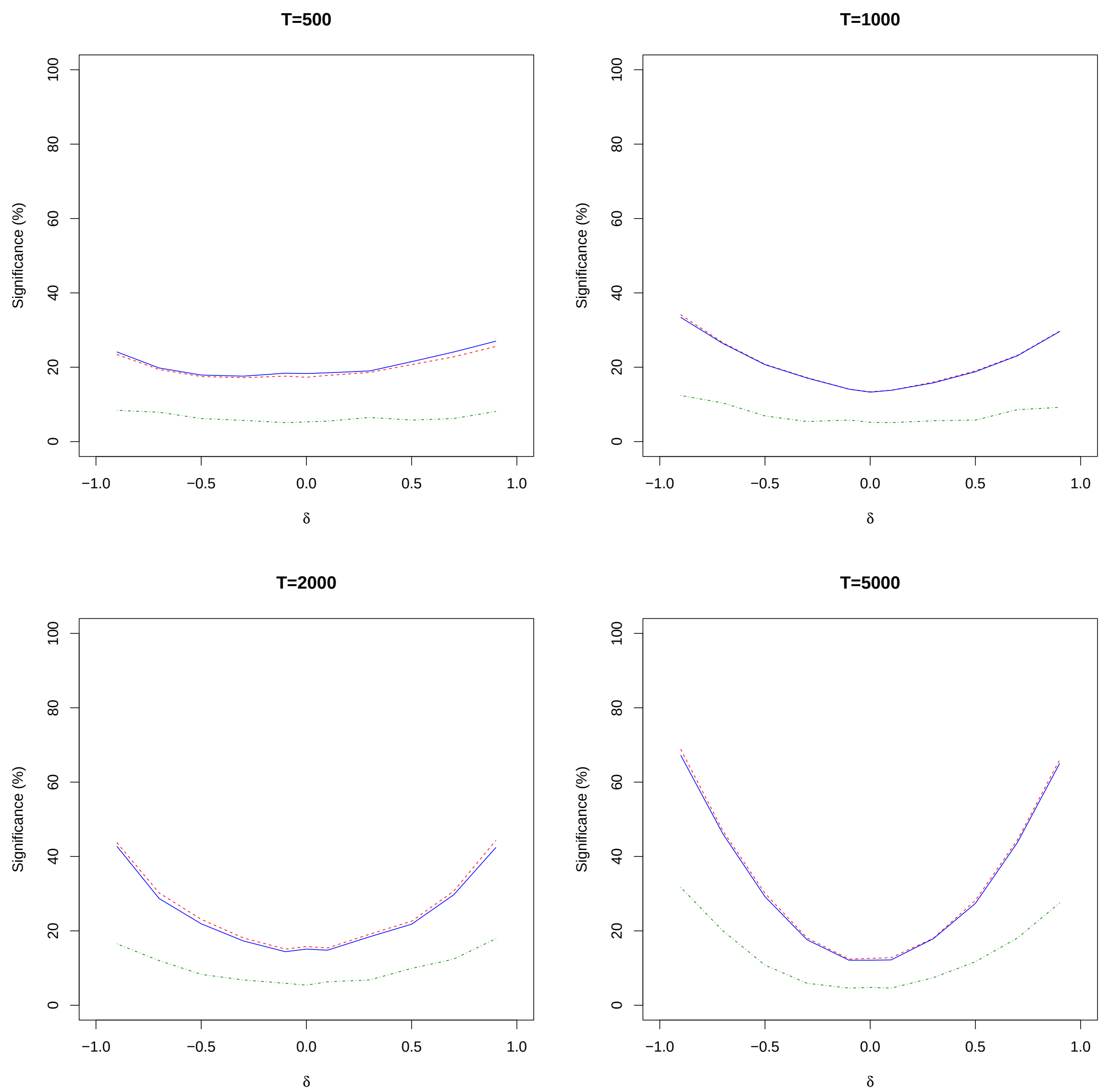
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test, and green is the Maximized Monte Carlo test.

Figure A4. Power curves of test for no leverage in SV(1) model when $\phi = 0.75$, $\sigma_y = 0.10$, and $\sigma_v = 1.00$ (LMC not level-corrected)



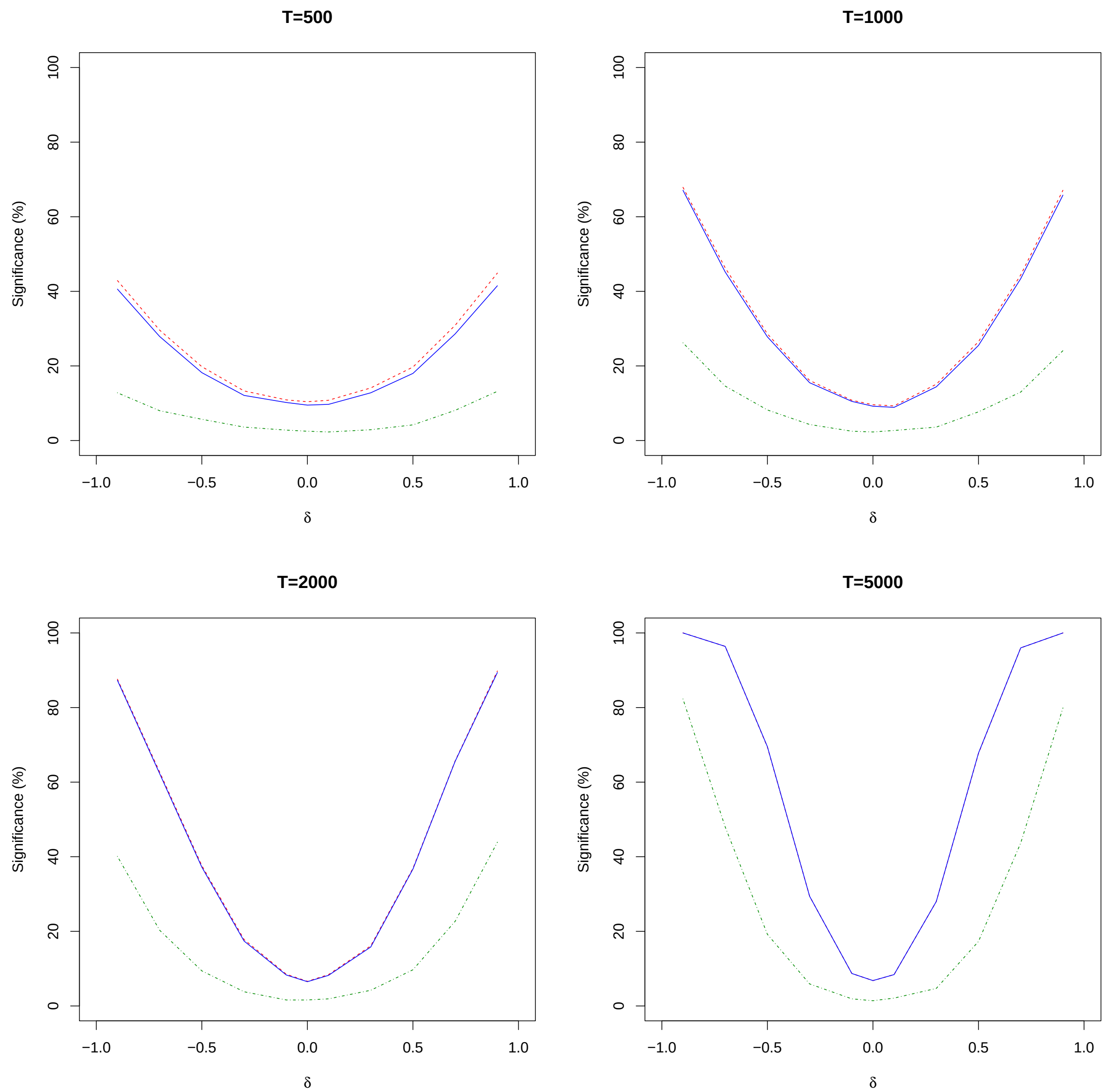
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test, and green is the Maximized Monte Carlo test.

Figure A5. Power curves of test for no leverage in SV(1) model when $\phi = 0.99$, $\sigma_y = 0.10$, and $\sigma_v = 0.25$



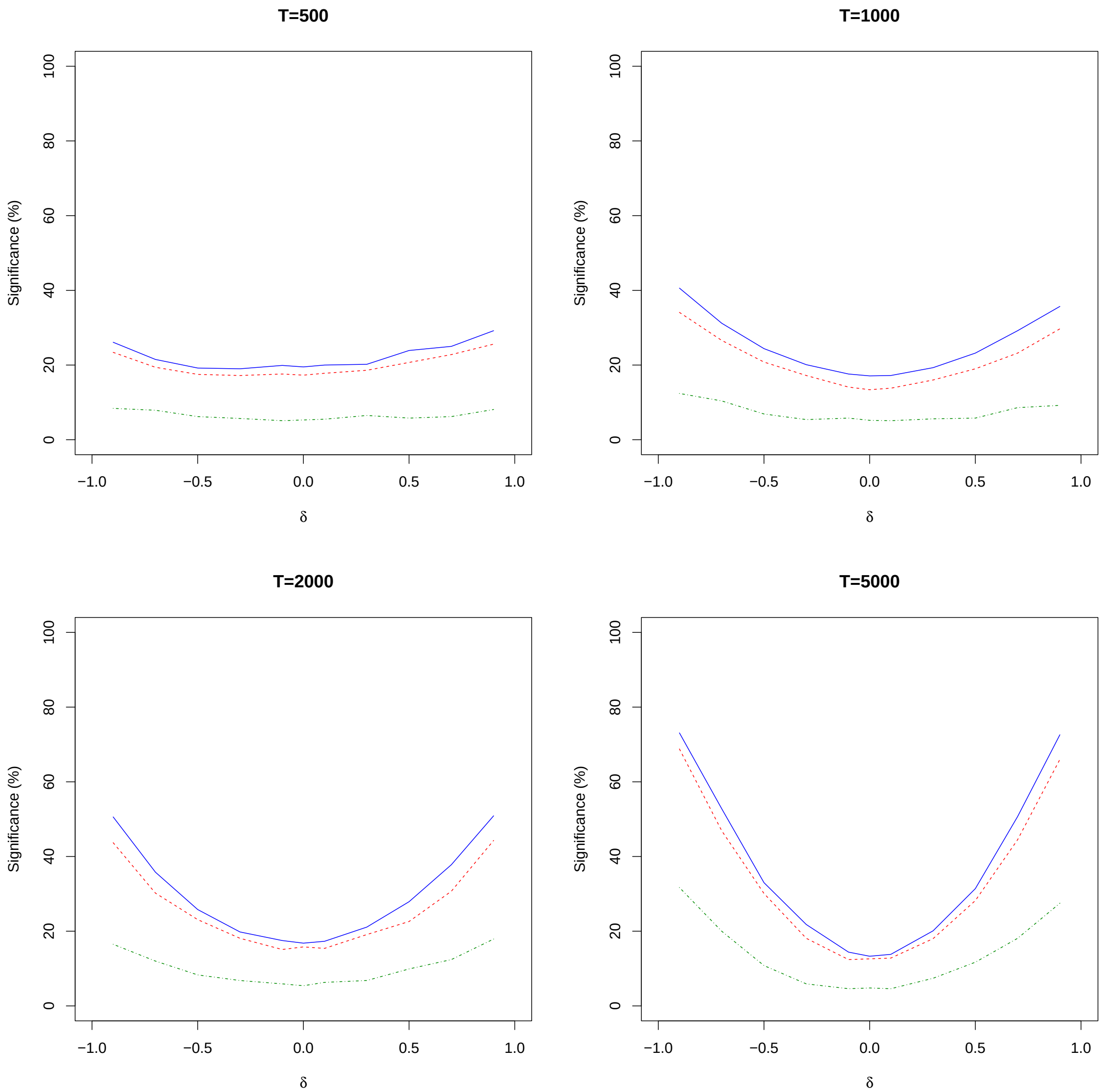
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test (level corrected), and green is the Maximized Monte Carlo test.

Figure A6. Power curves of test for no leverage in SV(1) model when $\phi = 0.95$, $\sigma_y = 0.10$, and $\sigma_v = 0.50$



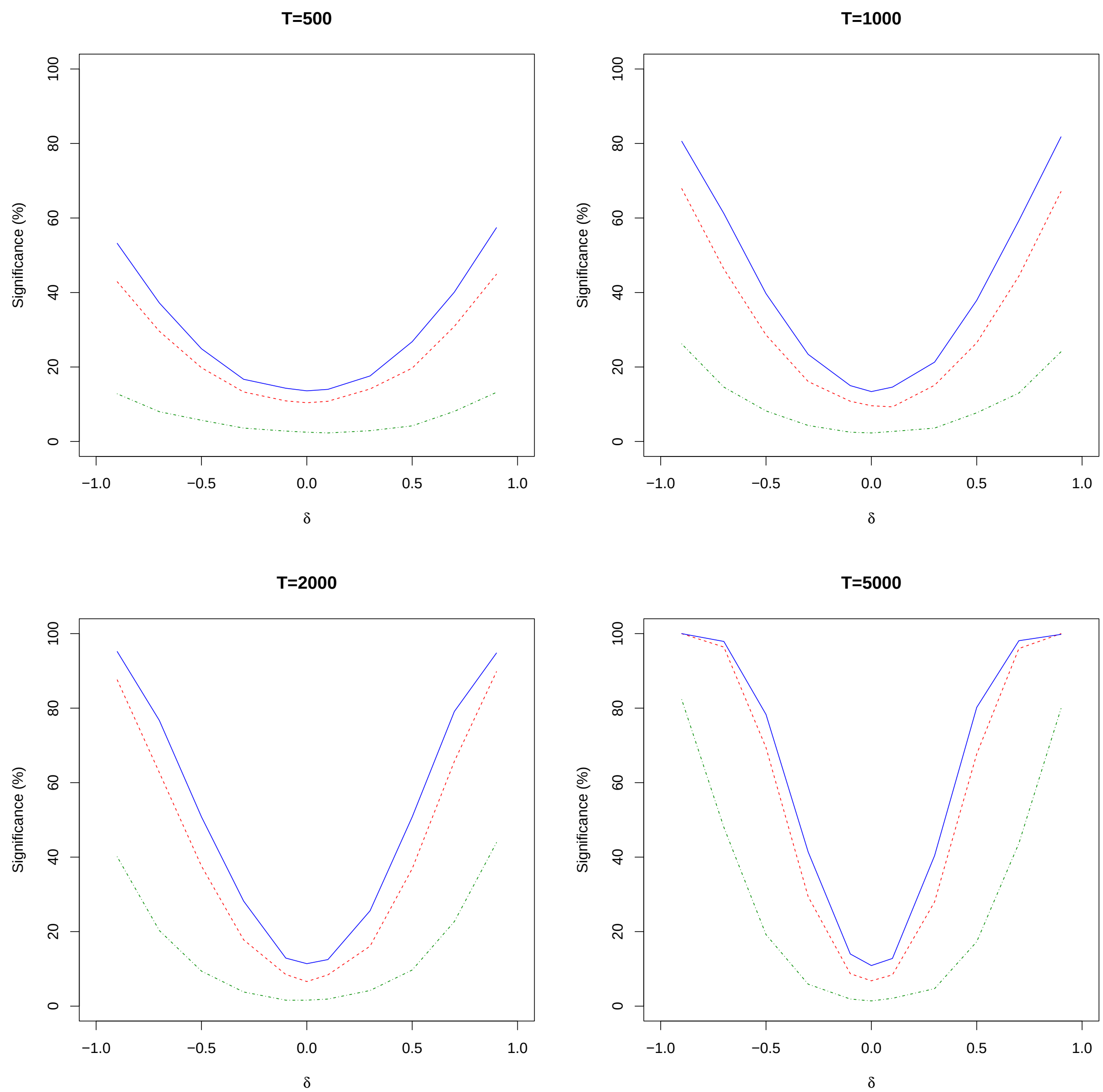
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test (level corrected), and green is the Maximized Monte Carlo test.

Figure A7. Power curves of test for no leverage in SV(1) model when $\phi = 0.99$, $\sigma_y = 0.10$, and $\sigma_v = 0.25$ (LMC not level-corrected)



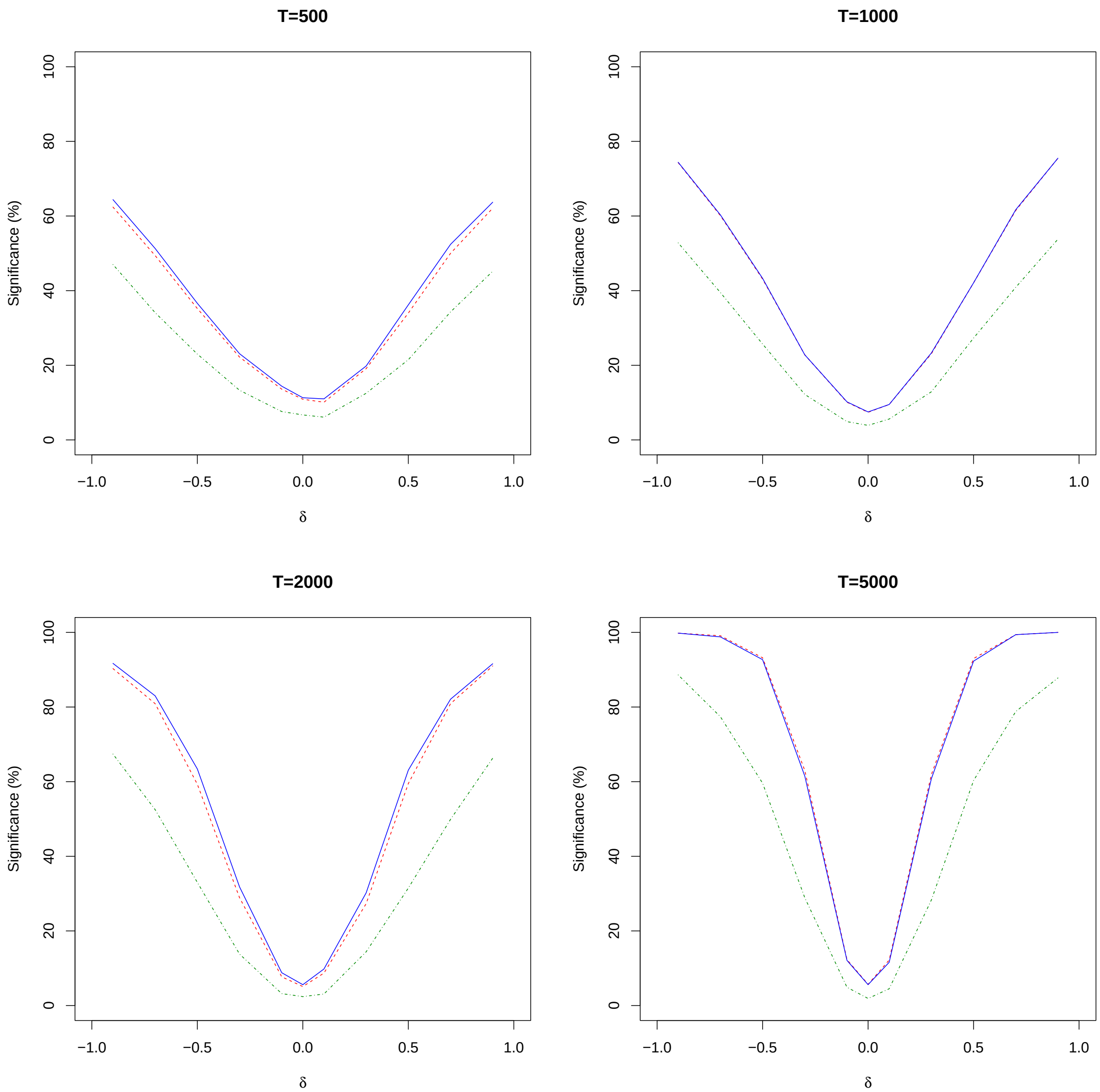
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test, and green is the Maximized Monte Carlo test.

Figure A8. Power curves of test for no leverage in SV(1) model when $\phi = 0.95$, $\sigma_y = 0.10$, and $\sigma_v = 0.50$ (LMC not level-corrected)



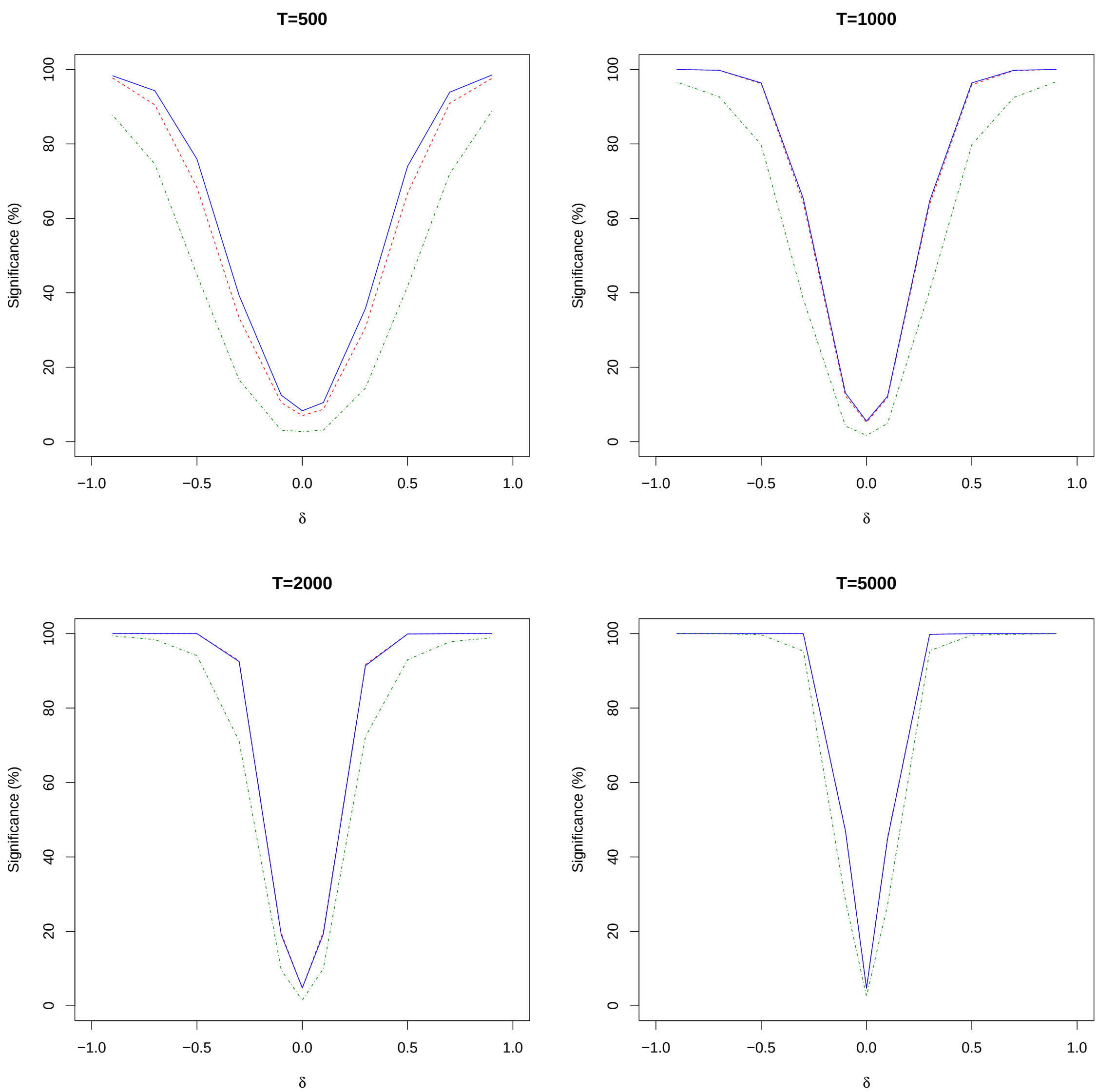
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test, and green is the Maximized Monte Carlo test.

Figure A9. Power curves of test for no leverage in SV(2) model when $\phi_1 = 0.05$, $\phi_2 = 0.85$, $\sigma_y = 1.00$, and $\sigma_v = 1.00$



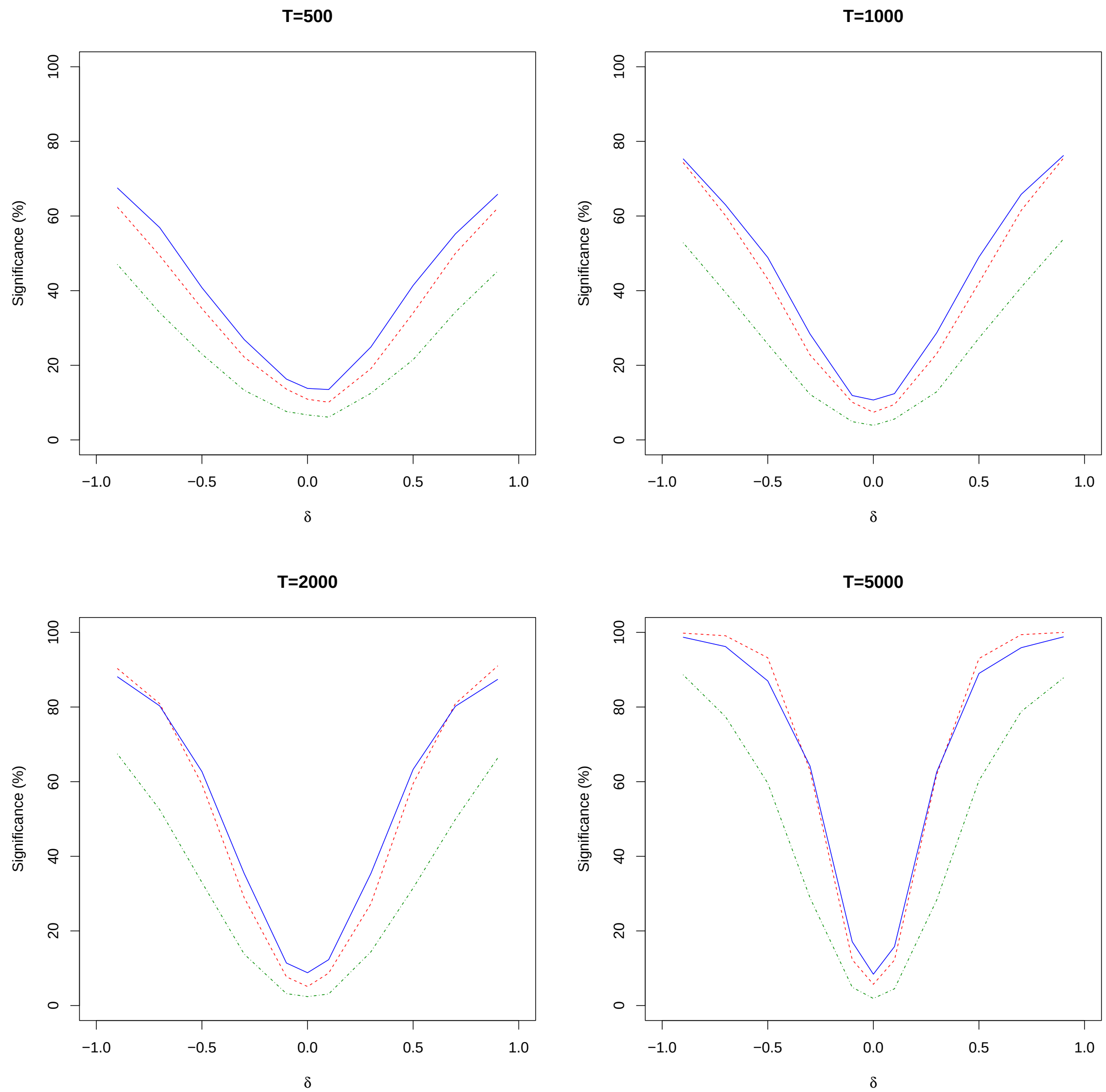
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test (level corrected), and green is the Maximized Monte Carlo test.

Figure A11. Power curves of test for no leverage in SV(2) model when $\phi_1 = 0.05$, $\phi_2 = 0.70$, $\sigma_y = 1.00$, and $\sigma_v = 1.00$



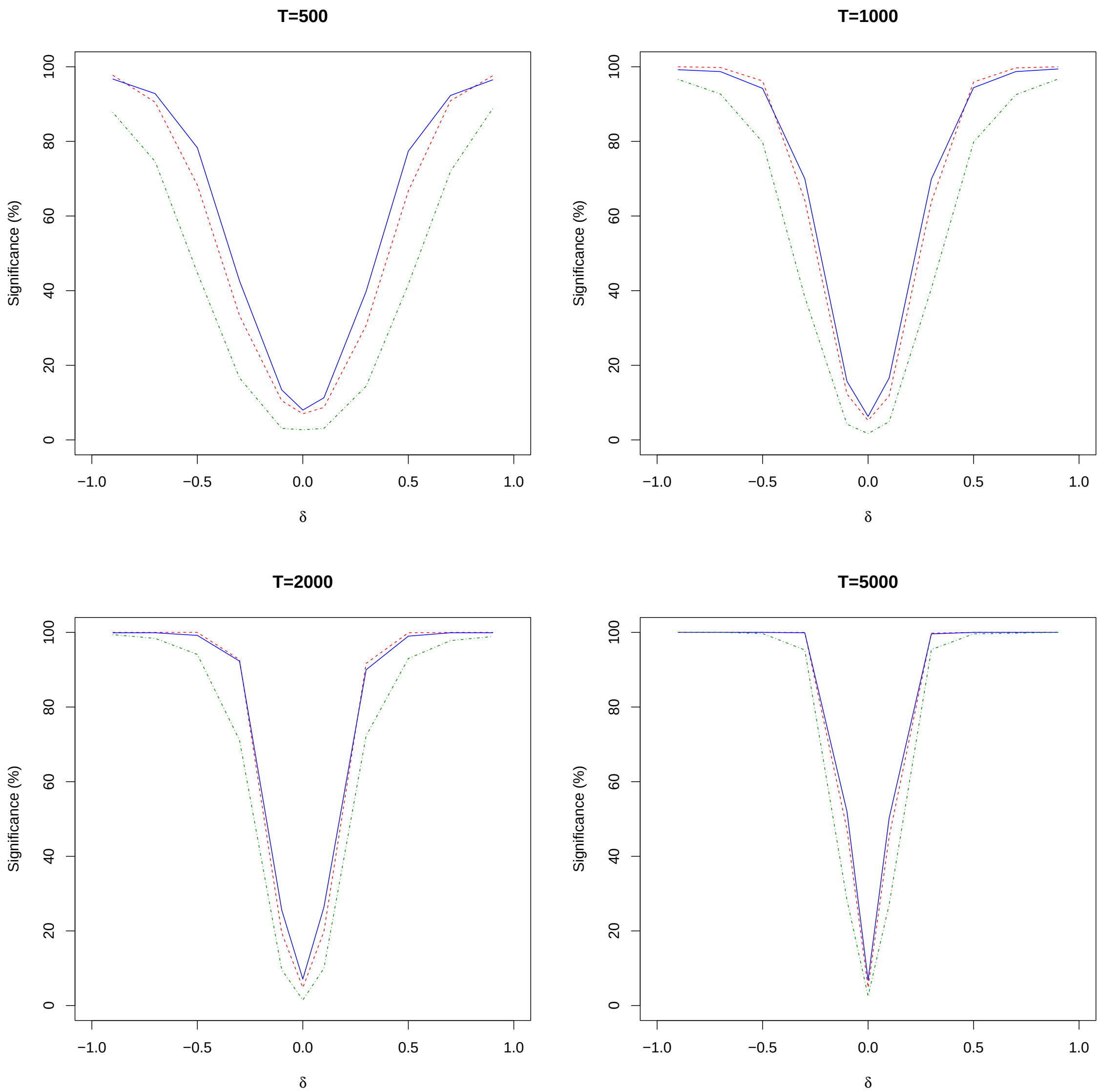
Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test (level corrected), and green is the Maximized Monte Carlo test.

Figure A11. Power curves of test for no leverage in SV(2) model when $\phi_1 = 0.05$, $\phi_2 = 0.85$, $\sigma_y = 1.00$, and $\sigma_v = 1.00$ (LMC not level-corrected)



Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test, and green is the Maximized Monte Carlo test.

Figure A12. Power curves of test for no leverage in SV(2) model when $\phi_1 = 0.05$, $\phi_2 = 0.70$, $\sigma_y = 1.00$, and $\sigma_v = 1.00$ (LMC not level-corrected)



Notes: Red is Asymptotic test (level corrected), blue is the Local Monte Carlo test, and green is the Maximized Monte Carlo test.

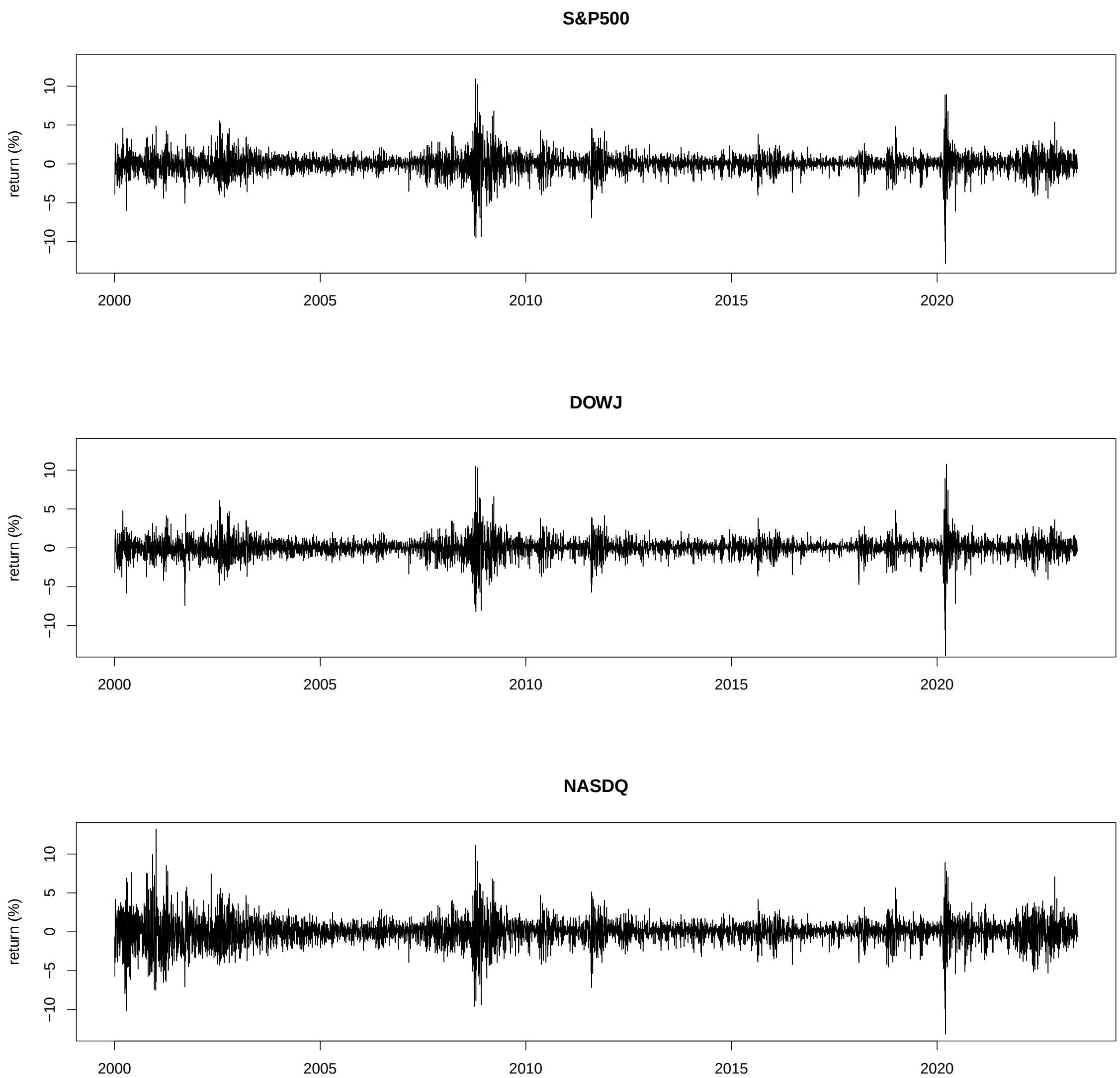
D Complementary empirical results

Table A4. Summary statistics

	Series	Mean	SD	Kurtosis	SK	Range	Min	Max	LB(10)
				Sample from 2000-Jan-04 to 2023-May-31 ($T = 5,889$)					
S&P500	y_t	0.00	1.25	10.16	-0.37	23.72	-12.78	10.94	97.32
($T = 5,889$)	y_t^2	1.56	5.43	271.31	13.58	163.41	0.00	163.41	5741.49
	y_t^*	0.00	2.58	2.47	-1.07	25.33	-18.64	6.70	1636.99
DOWJ	y_t	0.00	1.19	12.53	-0.37	24.61	-13.86	10.75	110.28
($T = 5,889$)	y_t^2	1.41	5.39	411.54	16.55	192.10	0.00	192.10	5809.00
	y_t^*	0.00	2.54	1.82	-0.98	22.37	-15.42	6.95	1551.33
NASDQ	y_t	0.00	1.60	6.06	-0.13	26.40	-13.17	13.24	63.06
($T = 5,889$)	y_t^2	2.56	7.26	161.34	10.04	175.17	0.00	175.17	4230.10
	y_t^*	0.00	2.56	2.37	-1.07	24.46	-18.28	6.19	1979.06

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

Figure A13. Time series of stock indices from 2000-Jan-04 to 2023-May-31 ($T = 5,889$)



Notes: Demeaned returns are shown (*i.e.*, $y_t = r_t - \mu$) for each stock index.

Table A5. Asymptotic and finite-sample tests for the presence of leverage in SVL(p) models (with $N = 99, N = 299$, & $N = 999$)

	H_0: $\delta = 0$ vs. H_1: $\delta \neq 0$																				
	SVL(1)							SVL(2)							SVL(3)						
	Asymptotic	LMC			MMC			Asymptotic	LMC			MMC			Asymptotic	LMC			MMC		
		$N = 99$	299	999	$N = 99$	299	999		$N = 99$	299	999	$N = 99$	299	999		$N = 99$	299	999	$N = 99$	299	999
S&P 500 ($T = 5889$)	0.00	0.01	0.01	0.00	0.02	0.01	0.01	0.07	0.79	0.84	0.83	0.92	0.93	0.95	0.00	0.28	0.34	0.32	0.96	0.65	0.70
Dow Jones ($T = 5889$)	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.78	0.82	0.84	0.93	0.94	0.94	0.00	0.39	0.41	0.37	0.92	0.76	0.46
NASDAQ ($T = 5889$)	0.00	0.02	0.01	0.00	0.01	0.00	0.01	0.56	0.86	0.85	0.87	0.91	0.90	0.92	0.00	0.62	0.66	0.70	0.81	0.89	0.89

Notes: 1. The reported values are p-values for test procedure when testing for leverage (*i.e.* $H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$). 2. The value N is the number of Monte Carlo simulations used to simulate the null distribution.

Table A6. Empirical volatility forecasting performance with competing conditional volatility models (larger estimation window)

	S&P 500			Dow Jones			NASDAQ		
Models	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
ARCH(1)	8.101	43.040	88.250	8.862	47.620	96.701	8.035	40.910	81.936
ARCH(2)	7.327	39.415	82.705	8.222	43.817	90.654	7.409	38.567	79.519
ARCH(3)	7.063	37.745	80.925	7.962	42.174	88.263	7.311	37.653	77.798
GARCH(1,1)	6.865	35.276	73.095	7.832	39.922	80.865	7.073	35.651	72.351
GARCH(2,2)	6.824	35.080	72.780	7.810	39.725	80.544	7.031	35.480	72.035
GARCH(3,3)	6.818	35.046	72.744	7.809	39.724	80.540	7.022	35.435	71.988
EGARCH(1,1)	6.568	33.881	69.999	7.636	38.723	78.100	6.827	34.425	69.701
EGARCH(2,2)	6.530	33.753	69.805	7.628	38.705	78.067	6.775	34.323	69.523
EGARCH(3,3)	6.641	34.434	71.390	7.665	38.878	78.530	6.878	34.982	71.085
GJR(1,1)	6.614	34.206	70.984	7.668	39.194	79.525	6.875	34.744	70.518
GJR(2,2)	6.585	34.073	70.791	7.662	39.128	79.404	6.846	34.619	70.326
GJR(3,3)	6.574	34.087	70.854	7.645	39.058	79.279	6.854	34.740	70.566
SV(1)	5.234	**26.906**	**56.077**	**5.932**	**29.936**	**60.345**	**5.577**	**27.924**	**56.833**
SV(2)	5.282	28.767	60.697	6.051	**31.202**	**63.594**	5.637	29.499	61.006
SV(3)	5.227	29.331	61.338	**5.940**	**31.621**	**64.082**	**5.562**	29.977	61.582
SVL(1)	**4.978**	**26.145**	**55.080**	**5.804**	**29.489**	**59.960**	**5.453**	**27.584**	**56.431**
SVL(2)	5.273	28.747	60.677	6.041	**31.177**	**63.569**	5.630	29.473	60.979
SVL(3)	**5.180**	29.252	61.258	**5.891**	**31.532**	**63.992**	**5.540**	29.930	61.534

Notes: The reported values represent the Mean Squared Error (MSE) for each model at different horizons: $h = 1$ (one day), $h = 5$ (one week), and $h = 10$ (two weeks). The values in bold indicate that the model is part of the Model Confidence Set (MCS) for that horizon and that specific index. The MCS is determined using a 5% significance level. The values in bold red indicate the models in the MCS with the lowest MSE. Sample for each index is from 2000-Jan-04 to 2023-May-31 ($T = 5,889$). Estimates are obtained using W-ARMA estimator given in (3.24) with $J = 250$. Out-of-sample forecasting is performed using a rolling window scheme of size $T_{est} = 4,880$, where the first estimation window ends at observation $4,879$, and the last estimation window starts at observation $1,001$ but ends at observation $5,879$.

Appendix References

Ahsan, M. N. and Dufour, J.-M. (2021), 'Simple estimators and inference for higher-order stochastic volatility models', *Journal of Econometrics* **224**(1), 181–197. Annals Issue: PI Day.

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

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.

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

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

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