BANK OF CANADA
BANQUE DU CANADA

Staff working paper / Document de travail du personnel—2026-8

Last updated: March 6, 2026

# Estimation and Inference for Stochastic Volatility Models with Heavy-Tailed Distributions

Gabriel Rodriguez Rondon
Canadian Economic Analysis Department
Bank of Canada
grodriguezrondon@bankofcanada.ca

Jean-Marie Dufour
McGill University
jean-marie.dufour@mcgill.ca

Md. Nazmul Ahsan
Canada Mortgage and Housing Corporation – CMHC
mahsan@cmhc-schl.gc.ca



DOI: https://doi.org/10.34989/swp-2026-8 | ISSN 1701-9397

# Estimation and Inference for Stochastic Volatility Models with Heavy-Tailed Distributions *

Md. Nazmul Ahsan† Jean-Marie Dufour‡ Gabriel Rodriguez-Rondon§

CMHC McGill University Bank of Canada

October 23, 2025

**Abstract**

Statistical inference–both estimation and testing–for stochastic volatility (SV) models is known to be challenging and computationally demanding. We propose simple and efficient estimators for SV models with conditionally heavy-tailed error distributions, particularly the Student's $t$ and Generalized Exponential Distributions (GED). The estimators rely on a small set of moment conditions derived from ARMA-type representations of SV models, with an option to apply "winsorization" to improve stability and finite-sample performance. Except for the degrees-of-freedom parameter, closed-form expressions are available for all other parameters–extending Ahsan and Dufour (2019, 2021)–thus eliminating the need for numerical optimization or initial values. We derive the estimators' asymptotic distribution and show that, due to their analytical tractability, they support reliable–and even exact–simulation-based inference via Monte Carlo or bootstrap methods. We assess their performance through extensive simulations and demonstrate their practical relevance in financial return data, which strongly reject the normality assumption in favor of heavy-tailed models.

**Key words:** Generalized method of moments, Monte Carlo tests, stochastic volatility, asymptotic distribution.

**JEL codes:** C15, C22, C53, C58

*The authors thank Torben Andersen, Russell Davidson, Hashem Pesaran, Eric Renault, Lynda Khalaf, Richard Baillie, John Galbraith, Victoria Zinde-Walsh, Denis Pelletier, Pascale Valéry, Firmin Doko Tchatoka, Masaya Takano for their valuable comments and constructive discussions. Earlier versions of this paper were presented at Bernoulli-IMS Symposium 2020. Financial support from the Fonds de recherche du Québec – Société et culture (FRQSC) Doctoral Research Scholarship (B2Z) is gratefully acknowledged.

†Canada Mortgage and Housing Corporation (CMHC), 700 Montreal Road, Ottawa, Ontario K1A 0P7, Canada; e-mail: mahsan@cmhc-schl.gc.ca.

‡William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en economie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada. e-mail: jean-marie.dufour@mcgill.ca. Web page: http://jeanmariedufour.com.

§Mailing address: Bank of Canada, 234 Wellington St. W, Ottawa, ON K1A 0G9. e-mail: grodriguezrondon@bankofcanada.ca. Wed page: https://grodriguezrondon.com.

**Résumé**

L'inférence statistique – tant pour l'estimation que pour les tests – dans les modèles de volatilité stochastique (SV) est reconnue comme étant complexe et exigeante sur le plan computationnel. Nous proposons des estimateurs simples et efficaces pour les modèles SV avec distributions d'erreurs conditionnellement à queues épaisses, notamment les distributions de Student-$t$ et exponentielle généralisée (GED). Les estimateurs reposent sur un petit nombre de conditions de moments dérivées des représentations de type ARMA des modèles SV, avec la possibilité d'appliquer une «winsorisation» afin d'améliorer la stabilité et la performance en petits échantillons. À l'exception du paramètre de degrés de liberté, des expressions en forme fermée sont disponibles pour tous les autres paramètres – prolongeant ainsi Ahsan and Dufour (2019, 2021) – ce qui élimine le besoin d'optimisation numérique ou de valeurs initiales. Nous dérivons la distribution asymptotique des estimateurs et montrons que, grâce à leur simplicité analytique, ils permettent une inférence fiable – voire exacte – fondée sur des méthodes de simulation de type Monte Carlo ou bootstrap. Nous évaluons leurs performances à l'aide de simulations extensives et démontrons leur pertinence empirique sur des données de rendements financiers, qui rejettent fortement l'hypothèse de normalité en faveur de modèles à queues épaisses.

# 1 Introduction

The statistical inference (estimation and testing) for stochastic volatility (SV) models is challenging and computationally expensive. We propose simple and efficient estimators for SV specifications with heavy-tailed distributions, especially the Student's $t$-distribution and the distribution (GED). The proposed class of estimators is based on a small set of moment equations derived from ARMA representations associated with these models, with the option of using model averaging to improve stability and efficiency. Except for the degrees of freedom parameter, closed-form expressions for all other parameters are derived, requiring neither numerical optimization nor choice of initial values. For the degrees of freedom parameter, we assume the innovation distribution belongs to a specific class (Student-$t$ or GED) and estimate its shape parameter through a profiling step. We develop the asymptotic distributional theory of the proposed estimators for the degrees of freedom parameter, and simulation studies confirm that these simple estimators perform well in practice in terms of precision.

Stochastic volatility models are widely used to describe time-varying variance in asset returns, which is crucial for asset pricing, risk management, and portfolio selection.[1] Proposed estimation methods for SV models include the generalized method of moments (GMM) [Melino and Turnbull (1990), Andersen and Sørensen (1996)]; quasi-maximum likelihood (QML) [Nelson (1988), Harvey et al. (1994), Ruiz (1994)]; the simulated method of moments (SMM) [Gallant and Tauchen (1996), Monfardini (1998), Andersen et al. (1999)]; Monte Carlo likelihood (MCL) [Sandmann and Koopman (1998)]; simulated maximum likelihood (SML) [Danielsson and Richard (1993), Danielsson (1994), Durham (2006, 2007), Richard and Zhang (2007)]; linear representation (LiR) methods [Francq and Zakoïan (2006)]; methods based on Bayesian Markov Chain Monte Carlo (MCMC) [Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Flury and Shephard (2011)]; closed-form moment-based estimator (DV) [Dufour and Valéry (2006, 2009)] and winsorized ARMA (WARMA) estimator [Ahsan and Dufour (2019, 2021); Ahsan et al. (2025a); Ahsan and Dufour (2025); Ahsan et al. (2025b).

The above estimation procedures are computationally expensive and/or statistically inefficient. SML, MCL, SMM, and Bayesian MCMC estimators are computationally expensive, inflexible across models, and may converge quite slowly [see Broto and Ruiz (2004)], while QML, GMM, LiR, and DV estimators are inefficient. The only exception is the WARMA estimator of Ahsan and Dufour

[1] See Ghysels et al. (1996), Broto and Ruiz (2004), and Shephard (2005) for reviews of the literature.

(2019) and others, which is analytically tractable, computationally simple, and extremely efficient, especially when compared with the Bayesian estimator [see Ahsan and Dufour (2021)].

*First,* we derive a system of autocovariances from the ARMA-type representations of SV models, yielding closed-form estimators for the volatility process parameters. These can be computed using only sample autocovariances, requiring no numerical optimization or initial value selection. These *simple ARMA-SV* estimators were first proposed in Ahsan and Dufour (2019). In particular, the persistence parameter $\phi$ is obtained through a simple ratio of autocovariances. Crucially, in this paper we extend this framework to handle heavy-tailed error distributions by incorporating the degrees of freedom parameter $\nu$, which plays a key role in modeling non-Gaussian behavior–a novel and central contribution of this paper.

*Second,* to improve small-sample performance and enforce stationarity, we propose winsorized versions of the simple ARMA-SV estimators (*W-ARMA-SV*), continuing the tradition of Ahsan and Dufour (2019). These combine several autocovariance-based ratios through robust weighting schemes, including the median, weighted averages, and an OLS-based scheme. The latter, in particular, offers strong empirical performance. Winsorization reduces sensitivity to outliers while preserving computational simplicity and efficiency.

*Third,* the analytical simplicity of the proposed estimators makes them particularly well-suited for constructing simulation-based hypothesis tests. In particular, they facilitate the implementation of exact Monte Carlo (MC) tests in the spirit of Dufour (2006). This avoids reliance on potentially invalid asymptotic approximations and provides level-correct inference–even in borderline or nonstationary settings, such as when the latent volatility process is nearly a unit root.

*Fourth,* we formally establish the asymptotic properties of our estimators. Under standard regularity conditions (e.g., existence of fourth moments), we show that the estimators are $\sqrt{T}$-consistent and asymptotically normal. These results hold at least for linearly winsorized versions of the estimators.

*Fifth,* we conduct extensive Monte Carlo simulations to evaluate the performance of our estimators in terms of bias and RMSE. We highlight four key findings: (1) the OLS-type W-ARMA-SV estimators consistently perform well across all designs considered; (2) even with small samples and heavy tails, the proposed estimators remain highly precise and robust; (3) estimation of $\nu$ is especially precise in the GED case, but becomes more challenging, yet still effective, as it approaches the Gaussian limit in the Student-$t$ setting; and (4) computational speed is orders of magnitude faster than simulation-based methods, making these estimators attractive for large-scale or real-time applications.

*Sixth,* we evaluate the performance of LR-type hypothesis tests for normality and heavy tails, constructed using the proposed estimators. We consider three testing procedures: (1) asymptotic chi-squared inference, (2) local Monte Carlo tests (LMC), and (3) Maximized Monte Carlo tests (MMC). We find that the asymptotic tests often suffer from size distortions, especially under moderate persistence. In contrast, the LMC and MMC procedures provide accurate level control and exhibit strong power, even in challenging settings.

*Seventh,* we apply the proposed estimators and tests to daily returns of three major U.S. stock indices (S&P 500, DOW Jones, NASDAQ) spanning over two decades. We find strong evidence of heavy tails and persistent volatility in all series. The Monte Carlo-based tests reliably reject normality and, in some cases, suggest moderate but significant deviations from extreme heavy-tailed behavior. These results highlight the empirical relevance and practical value of the proposed framework for modeling and testing volatility in financial time series.

This paper is organized as follows. Section 2 introduces the model, assumptions, and underlying motivation. Section 3 describes the proposed estimators in detail. Section 4 develops their asymptotic distribution. Section 5 presents the LR-type tests for distributional assumptions. Section 6 reports simulation results for both estimation and testing. Section 7 provides an empirical application to equity return data. We conclude in Section 8, and technical proofs are provided in the appendix.

# 2 Simple estimators for stochastic volatility model

## 2.1 Stochastic volatility model and ARMA representation

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

**Assumption 2.1** *Stochastic volatility model.*

*The process $\{y_t : t \in \mathbb{N}_0\}$ satisfies the equations*

$$y_t = \sigma_y \exp\left(\frac{w_t}{2}\right) u_t, \tag{1}$$

$$w_t = \phi w_{t-1} + \sigma_v v_t, \quad v_t \sim \mathcal{N}(0, 1), \tag{2}$$

*where $|\phi| < 1$, and $w_0 \sim \mathcal{N}\left(0, \sigma_v^2/(1 - \phi^2)\right)$. The parameters $\phi$, $\sigma_y$, and $\sigma_v$ are fixed with $\sigma_y > 0$ and $\sigma_v > 0$. The innovation sequence $\{u_t\}$ is i.i.d., independent of $\{v_t\}$, with mean zero and unit variance. It is assumed that $u_t$ follows a symmetric distribution such as Student-t or GED, with additional moment conditions imposed as needed.*

*Note also that the root of the characteristic equation of the volatility process $w_t$,*

$$\phi(z) := 1 - \phi z = 0,$$

*lies outside the unit circle ($|z| > 1$); that is, $\phi(z) \neq 0$ for all $|z| \leq 1$.*

Under the latter part of this assumption, the process $y_t$ is strictly stationary since it depends on the strictly stationary latent process $w_t$ and an i.i.d. innovation $u_t$ that is independent of $w_t$. This model consists of two stochastic processes: $y_t$ captures the dynamics of the observable variable, while $w_t := \log(\sigma_t^2)$ governs the dynamics of the latent log-volatility.[2] The process $w_t$ can be interpreted as a random flow of uncertainty shocks or information, and $\phi$ captures the persistence of volatility dynamics.

We focus on an SV(1) specification – an AR(1) process as in equation (2) – to emphasize the role of alternative distributional assumptions in contrast to previous ARMA-based estimators. However, it is straightforward to extend the approach to higher-order SV($p$) models, as described in Ahsan and Dufour (2021) and Ahsan et al. (2025a). Accordingly, we begin with a brief review of the estimators under the assumption of Gaussianity, as presented in Ahsan and Dufour (2019) for the SV(1) specification.

**Assumption 2.2** *Gaussian error.*

*The error $u_t$ is i.i.d. according to a* N[0, 1] *distribution.*

The SV model given in Assumptions 2.1–2.2 can be written in state-space form (see Ahsan and

[2] In financial applications, the $y_t$'s can be residual returns, defined as $y_t := r_t - \mu_r$ with $r_t := 100[\log(p_t) - \log(p_{t-1})]$, where $\mu_r$ is the mean return and $p_t$ is the asset price. In macroeconomic applications, $y_t$ may represent regression residuals, as in Jurado et al. (2015).

Dufour (2019)):

$$\text{State Transition Equation:} \quad w_t = \phi w_{t-1} + v_t\,, \tag{3}$$

$$\text{Measurement Equation:} \quad y_t^* = w_t + \epsilon_t\,, \tag{4}$$

where

$$y_t^* := \log(y_t^2) - \mu\,, \quad \mu := \mathbb{E}[\log(y_t^2)] = \log(\sigma_y^2) + \mathbb{E}[\log(u_t^2)]\,, \quad \epsilon_t := \log(u_t^2) - \mathbb{E}[\log(u_t^2)]\,, \tag{5}$$

with $\mathbb{E}[\log(u_t^2)] = \log(2) + \psi(1/2) \simeq -1.2704$, and the $v_t$'s are i.i.d. $\mathrm{N}(0, \sigma_v^2)$ and the $\epsilon_t$'s are i.i.d. $\log(\chi_1^2)$ with mean zero and $\sigma_\epsilon^2 = \pi^2/2$. [3]

The state-space model given in (3)–(4) has an ARMA(1, 1) representation given by:

$$y_t^* = \phi y_{t-1}^* + \eta_t - \theta \eta_{t-1} \tag{6}$$

with $\eta_t - \theta\eta_{t-1} = v_t + \epsilon_t - \phi\epsilon_{t-1}$, where the error processes $\{v_t\}$ and $\{\epsilon_t\}$ are mutually independent, the errors $v_t$ are i.i.d. $\mathrm{N}(0, \sigma_v^2)$, and the errors $\epsilon_t$ are i.i.d. according to the distribution of a $\log(\chi_1^2)$ random variable.

## 2.2 Closed-form ARMA-SV estimators

Ahsan and Dufour (2019) showed that $y_t^*$ defined in (5) has the following autocovariances

$$\mathrm{cov}(y_t^*,\, y_{t-k}^*) := \gamma_{y^*}(k) = \begin{cases} \phi\gamma_{y^*}(k-1) + \sigma_v^2 + \sigma_\epsilon^2\,, & \text{if } k = 0, \\ \phi\gamma_{y^*}(k-1) - \phi\sigma_\epsilon^2\,, & \text{if } k = 1, \\ \phi\gamma_{y^*}(k-1)\,, & \text{if } k \geq 2. \end{cases} \tag{7}$$

and the above autocovariances yield the following closed-form expressions for SV parameters:

$$\phi = \frac{\gamma_{y^*}(k+1)}{\gamma_{y^*}(k)}\,, \quad \text{for } k \geq 1\,, \tag{8}$$

$$\sigma_y^2 = \exp[\mu - \mu_2]\,, \tag{9}$$

$$\sigma_v^2 = (1 - \phi^2)[\gamma_{y^*}(0) - (\pi^2/2)], \tag{10}$$

[3] For further discussion of this representation, see Nelson (1988), Harvey et al. (1994), Ruiz (1994), Shephard (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim et al. (1998), Sandmann and Koopman (1998), Steel (1998), Chib et al. (2002), Knight et al. (2002), Francq and Zakoïan (2006), Omori et al. (2007).

where $\gamma_{y^*}(k) = \text{cov}(y_t^*, y_{t-k}^*)$, with $y_t^*$ and $\mu$ defined in (5), and $\mu_2 := \mathbb{E}[\log(u_t^2)] \simeq -1.2704$.

From (8), we see that $\phi$ can be obtained from several autocovariance ratios, which can be easily estimated with the corresponding empirical moments:

$$\hat{\gamma}_{y^*}(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} [\log(y_t^2) - \hat{\mu}][\log(y_{t+k}^2) - \hat{\mu}] , \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} \log(y_t^2) . \tag{11}$$

Of these, the ratio $\gamma_{y^*}(2)/\gamma_{y^*}(1)$ is the one for which we can use the largest number of observations. This suggests the *simple ARMA-SV* estimators:

$$\hat{\phi} = \frac{\hat{\gamma}_{y^*}(2)}{\hat{\gamma}_{y^*}(1)}, \quad \hat{\sigma}_y^2 = \exp(\hat{\mu} + 1.2704), \quad \hat{\sigma}_v^2 = (1 - \hat{\phi}^2)[\hat{\gamma}_{y^*}(0) - (\pi^2/2)] . \tag{12}$$

# 3 Simple estimators for different SV specifications

In the previous section, we discussed closed-form estimation for the log-normal SV model. However, several alternative SV specifications have been proposed in the literature to better capture empirical stylized facts–most notably, heavy-tailed return innovations. In this section, we extend the simple W-ARMA estimation approach of Ahsan and Dufour (2019, 2021) to models with heavy-tailed distributions.

## 3.1 Almost closed estimators for SV models with heavy-tailed distributions

We now assume that the innovations $u_t$ follow a non-Gaussian distribution, i.e., $u_t \sim f(u)$, where $f(u)$ is a heavy-tailed distribution. The state-space representation (equations (3)–(5)) remains valid under this more general assumption. Using the autocovariances $\gamma_{y^*}(k)$ defined in (7), which can be estimated directly from the data, we can still solve for $\phi$, $\sigma_\epsilon^2$, and $\sigma_v^2$ using (8) and

$$\sigma_\epsilon^2 = \frac{\phi\gamma_{y^*}(0) - \gamma_{y^*}(1)}{\phi} , \tag{13}$$

$$\sigma_v^2 = \gamma_{y^*}(1) - \phi\gamma_{y^*}(0) - \sigma_\epsilon^2 . \tag{14}$$

and thus obtain estimators for these parameters. That is, the Gaussian assumption is not required for these results, and both regularization and winsorization, which we discuss in more detail below, remain applicable in this setting.

However, the expression for $\log u_t^2$ will differ depending on the specific distributional assumption imposed on $u_t$.

#### 3.1.1 SV model with student-$t$ distributed error

**Assumption 3.1** *Student-t distributed error.*

*The error $u_t$ is an i.i.d. $t$ distributed random variable with $\nu$ degrees of freedom.*

The $t(\nu)$-distributed random variable $u_t$ can be expressed as $u_t = \lambda_t^{-1/2} z_t$, where $z_t \sim \mathrm{N}(0, 1)$ and $\nu\lambda_t$ is distributed independently of $z_t$ as a chi-squared variable with $\nu$ degrees of freedom, i.e., $\lambda_t \sim \chi^2_\nu/\nu$ (see Rice and Rice (2007)). Taking the logarithm of $u_t^2$ yields:

$$\log u_t^2 = \log z_t^2 - \log \lambda_t \tag{15}$$

**Lemma 3.1** *Cumulants and central moments of the* $\log \chi^2_\nu/\nu$ *random variable.*

*Let $\nu\lambda \sim \chi^2_\nu$ be a chi-square variate with $\nu$ degrees of freedom. Then the cumulants ($\kappa_m$) and central moments ($\tilde{\mu}_m$) of $\log \lambda$ are given by*

$$\kappa_m = \begin{cases} \psi\left(\frac{\nu}{2}\right) - \log\left(\frac{\nu}{2}\right), & \text{if } m = 1 \\ \psi^{(m-1)}\left(\frac{\nu}{2}\right), & \text{if } m > 1 \end{cases}$$
$$\tilde{\mu}_m = \begin{cases} 0, & \text{if } m = 1 \\ \kappa_m + \sum_{j=1}^{m-2} \binom{m-1}{j} \kappa_{m-j}\tilde{\mu}_j, & \text{if } m > 1 \end{cases}, \tag{16}$$

*where $\psi(z)$ and $\psi^{(m)}(z)$ are the digamma and polygamma function of order $m$.*

From Lemma 3.1, the mean and variance of $\log z_t^2$ and $\log \lambda_t$ are

$$\mathbb{E}(\log z_t^2) = \psi(1/2) + \log(2), \qquad Var(\log z_t^2) = \psi^{(1)}(1/2),$$
$$\mathbb{E}(\log \lambda_t) = \psi(\nu/2) - \log(\nu/2), \qquad Var(\log \lambda_t) = \psi^{(1)}(\nu/2),$$

Consequently, the mean and variance of $\log u_t^2$ are

$$\mathbb{E}(\log u_t^2) = \psi(1/2) - \psi(\nu/2) + \log(\nu), \qquad Var(\log u_t^2) = \psi^{(1)}(1/2) + \psi^{(1)}(\nu/2).$$

The above expressions yield the following expressions for SV-t parameters:

$$\begin{aligned}\phi &= \frac{\gamma_{y^*}(k+1)}{\gamma_{y^*}(k)}, \quad \text{for } k \geq 1, & (17)\\ \sigma_y^2 &= \exp[\mu - \psi(1/2) + \psi(\nu/2) - \log(\nu)], & (18)\\ \sigma_v^2 &= (1-\phi^2)[\gamma_{y^*}(0) - \psi^{(1)}(1/2) - \psi^{(1)}(\nu/2)], & (19)\end{aligned}$$

where $\gamma_{y^*}(k) = \mathrm{cov}(y_t^*, y_{t-k}^*)$, with $y_t^*$ and $\mu$ defined in (5), and $\nu$ can be obtained from the following optimization problem:

$$\nu := \arg\min_{\nu \in \mathcal{R}^+} \left\{ \left( \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\nu}{2}\right) - \gamma_{y^*}(0) + \frac{\gamma_{y^*}(1)}{\phi} \right)^2 \right\} \quad (20)$$

where $\mathcal{R}^+$ is the set of positive real numbers numbers.

#### 3.1.2 SV model with *GED* distributed error

**Assumption 3.2** *Stochastic volatility model with Generalized Exponential Distribution (GED).*

*The process $\{y_t : t \in \mathbb{N}_0\}$ satisfies the equations*

$$y_t = \sigma_y \exp\left(\frac{w_t}{2}\right)\bar{u}_t, \quad \bar{u}_t \sim \mathcal{GED}(\nu), \quad (21)$$

$$w_t = \phi w_{t-1} + \sigma_v v_t, \quad v_t \sim \mathrm{N}(0,\, 1), \quad (22)$$

*where $|\phi| < 1$ and $w_0 \sim \mathrm{N}[0, \sigma_v^2/(1-\phi^2)]$, while $\phi$, $\sigma_y$ and $\sigma_v$ are fixed parameters. The probability density function of $\bar{u}_t$ is given by*

$$f_\nu(\bar{u}) = \frac{\nu \exp\left\{ -(1/2)|\bar{u}/\lambda(\nu)|^\nu \right\}}{\lambda(\nu) 2^{1+1/\nu} \Gamma(1/\nu)}, \;\; \nu > 0, \;\; \bar{u} \in \mathbb{R}, \quad (23)$$

*where $\lambda(\nu) = \left\{2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)\right\}^{1/2}$ and $\Gamma(z)$ denotes the gamma function.*

We assume that $\bar{u}_t$ follows a standardized Generalized Exponential Distribution (GED), $\bar{u}_t \sim \mathcal{GED}(\nu)$, with $\mathbb{E}[\bar{u}_t] = 0$, $\mathbb{E}[\bar{u}_t^2] = 1$, and $\mathbb{E}[\bar{u}_t^k] < \infty$ for all $k > 0$. The $\mathcal{GED}$ includes the normal distribution as a special case when $\nu = 2$, and the double exponential distribution when $\nu = 1$, but allows for heavier tails when $\nu < 2$.[4] It is infinitely differentiable in $\nu$, except at the point $\bar{u} = 0$.

[4] The Generalized Exponential Distribution (GED) is sometimes referred to as the exponential power distribution (EPD) or the Subbotin distribution, after Subbotin (1923), who introduced the family.

The reason why the $\mathcal{GED}$ family of distributions could be an attractive alternative is because it includes the normal as an special case, and also includes distributions with thinner and fatter tails than the normal.

**Lemma 3.2** *Cumulants and central moments of the* log *squared $\mathcal{GED}$ random variable.*

*Let a random variable $\bar{u}_t$ have $\mathcal{GED}$ with parameter $\nu > 0$. Then the cumulants ($\kappa_m$) and central moments ($\tilde{\mu}_m$) of $\log(\bar{u}_t^2)$ are given by*

$$\kappa_m = \begin{cases} \frac{2}{\nu}\psi\left(\frac{1}{\nu}\right) + \log\left[\Gamma\left(\frac{1}{\nu}\right)\right] - \log\left[\Gamma\left(\frac{3}{\nu}\right)\right], & \text{if } m = 1 \\ \left(\frac{2}{\nu}\right)^m \psi^{(m-1)}\left(\frac{1}{\nu}\right), & \text{if } m > 1 \end{cases}$$

$$\tilde{\mu}_m = \begin{cases} 0, & \text{if } m = 1 \\ \kappa_m + \sum_{j=1}^{m-2} \binom{m-1}{j} \kappa_{m-j}\tilde{\mu}_j, & \text{if } m > 1 \end{cases}, \tag{24}$$

*where $\psi(z)$ and $\psi^{(m)}(z)$ are the digamma and polygamma function of order $m$.*

From Lemma 3.2, the mean and variance of $\log \bar{u}_t^2$ are

$$\mathbb{E}(\log \bar{u}_t^2) = \frac{2}{\nu}\psi\left(\frac{1}{\nu}\right) + \log\left[\Gamma\left(\frac{1}{\nu}\right)\right] - \log\left[\Gamma\left(\frac{3}{\nu}\right)\right], \qquad Var(\log \bar{u}_t^2) = \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right).$$

The above expressions yield the following expressions for $\text{SV}_{\mathcal{GED}}$ parameters:

$$\phi = \frac{\gamma_{y^*}(k+1)}{\gamma_{y^*}(k)}, \quad \text{for } k \geq 1, \tag{25}$$

$$\sigma_y^2 = \exp\left[\mu - \frac{2}{\nu}\psi\left(\frac{1}{\nu}\right) - \log\left[\Gamma\left(\frac{1}{\nu}\right)\right] + \log\left[\Gamma\left(\frac{3}{\nu}\right)\right]\right], \tag{26}$$

$$\sigma_v^2 = (1-\phi^2)\left[\gamma_{y^*}(0) - \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right)\right], \tag{27}$$

where $\gamma_{y^*}(k) = \text{cov}(y_t^*, y_{t-k}^*)$, with $y_t^*$ and $\mu$ defined in (5), and $\nu$ can be obtained from the following optimization problem:

$$\nu := \arg\min_{\nu \in \mathcal{R}^+} \left\{ \left( \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \gamma_{y^*}(0) + \frac{\gamma_{y^*}(1)}{\phi} \right)^2 \right\}, \tag{28}$$

where $\mathcal{R}^+$ is the set of positive real numbers numbers.

### 3.2 Winsorized estimator

A shortcoming of the above simple ARMA-type estimator is that it can yield inadmissible parameter values, *e.g.* with $\left|\hat{\phi}\right| \geq 1$. This issue can arise especially in small samples or in the presence of outliers. To increase stability and efficiency, Ahsan and Dufour (2019, 2021) proposed to use "winsorization" technique, which exploits the following relationship between the volatility persistence parameter $\phi$ and covariances of the process, $y_t^*$:

$$\phi = \sum_{j=1}^{\infty} w_j \frac{\gamma_{y^*}(j+1)}{\gamma_{y^*}(j)} \tag{29}$$

for any $w_j$ sequence with $\sum_{j=1}^{\infty} w_j = 1$. It is easy to see that the above relationship follows from (8). This suggests a more general class of estimators for $\phi$ can be obtained by averaging several sample analogs of the ratios $\gamma_{y^*}(j+1)/\gamma_{y^*}(j)$:

$$\tilde{\phi} = \sum_{j=1}^{J} w_j \frac{\hat{\gamma}_{y^*}(j+1)}{\hat{\gamma}_{y^*}(j)} \tag{30}$$

where $1 \leq J \leq T-2$ with $\sum_{j=1}^{J} w_j = 1$, and $T$ is the length of the time series. We call such estimators *winsorized ARMA-SV* estimators (or *W-ARMA-SV* estimators). Other (possibly nonlinear) averaging methods, such as the median, may also be used.

In simulations and empirical applications, Ahsan and Dufour (2019) find that an equal-weighted OLS-based winsorized estimator works remarkably well. The ($\hat{\phi}_{OLS}$) estimate is as follows:

$$\hat{\phi}_{OLS} = \left(\bar{a}\,'\bar{a}\,\right)^{-1}\bar{a}\,'\bar{e} \tag{31}$$

where $\bar{a} = [\hat{\gamma}_{y^*}(1)w_1^{1/2}, \ldots, \hat{\gamma}_{y^*}(J)w_J^{1/2}]'$ and $\bar{e} = [\hat{\gamma}_{y^*}(2)w_1^{1/2}, \ldots, \hat{\gamma}_{y^*}(J+1)w_J^{1/2}]'$ with $w_j = 1/J$. A simplified expression is as follows:

$$\hat{\phi}_{OLS} = \frac{\sum_{j=1}^{J} \hat{\gamma}_{y^*}(j)\,\hat{\gamma}_{y^*}(j+1)}{\sum_{j=1}^{J} \hat{\gamma}_{y^*}(j)^2}. \tag{32}$$

When $J = 1$, the above expression yields the simple ARMA-SV estimator $\hat{\phi} = \hat{\gamma}_{y^*}(2)/\hat{\gamma}_{y^*}(1)$.

For the remainder of the paper, we focus on the OLS-based winsorized estimator for both the simulation and empirical results, and report outcomes for different values of $J$ when relevant. For a detailed discussion of alternative winsorization methods and weighting schemes, we refer readers to Ahsan and Dufour (2019, 2021) and the references therein.

# 4 Asymptotic distributional theory

In this section, we derive the asymptotic distribution of the estimator for the degrees of freedom parameter $\nu$, which governs the tail behavior of the innovation distribution in our heavy-tailed stochastic volatility (SV) models.

We do not repeat the asymptotic theory for the volatility persistence parameter $\phi$ or the scale parameters $\sigma_y$ and $\sigma_\nu$, as those results are already established in Ahsan et al. (2025a). Instead, the focus here is on the novel asymptotic properties of $\nu$, whose estimation requires a distinct approach due to its implicit definition via moment-based profiling-type estimation.

Recall from Section 2 that $\nu$ is estimated by solving a nonlinear sample moment condition, derived from the identification condition of the variance of the transformed observation equation. Specifically, for the Student-$t$ case, the estimator $\hat{\nu}$ is defined as

$$\hat{\nu} := \arg\min_{\nu \in \mathcal{R}^+} \left\{ f_T^{St}(\nu) := \left( \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\nu}{2}\right) - \hat{\gamma}_{y^*}(0) + \frac{\hat{\gamma}_{y^*}(1)}{\hat{\phi}} \right)^2 \right\}, \tag{33}$$

where $\hat{\gamma}_{y^*}(k)$ denotes the sample autocovariance of $y_t^* = \log(y_t^2) - \hat{\mu}$ at lag $k$, and $\hat{\phi}$ is a consistent estimator of the persistence parameter.

We begin by stating the assumptions required for identification and inference.

**Assumption 4.1** *Moment condition identification for SV-t.*

*Define the population moment function*

$$f^{St}(\nu) := \psi^{(1)}\left(\tfrac{1}{2}\right) + \psi^{(1)}\left(\tfrac{\nu}{2}\right) - \gamma_{y^*}(0) + \frac{1}{\phi}\gamma_{y^*}(1),$$

*where* $\gamma_{y^*}(h) := \mathrm{Cov}(y_t^*, y_{t-h}^*)$ *is the autocovariance function of the transformed process. We assume:*

- *(A1)* $f^{St}(\nu)$ *is continuously differentiable on an open neighborhood* $\mathcal{N} \subset (4, \infty)$ *containing* $\nu_0$*,*
- *(A2)* $f^{St}(\nu) = 0$ *has a unique solution at* $\nu = \nu_0 \in \mathcal{N}$*,*
- *(A3)* $\partial f^{St}(\nu)/\partial\nu\big|_{\nu=\nu_0} \neq 0$*.*

**Assumption 4.2** *Fourth moment condition for SV-t.*

- *(B1)* $\mathbb{E}[\log(u_t^2)^4] < \infty$, *which holds if and only if* $\nu > 4$,
- *(B2) The long-run variance*

$$\sigma_f^2 := \mathrm{Var}\left(\hat{\gamma}_{y^*}(0) - \frac{1}{\phi}\hat{\gamma}_{y^*}(1)\right)$$

*is finite and strictly positive.*

A similar profiling-type approach is used for the GED case, namely,

$$\hat{\nu} := \arg\min_{\nu \in \mathcal{R}^+} \left\{ f_T^{\mathcal{GED}}(\nu) := \left( \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \hat{\gamma}_{y^*}(0) + \frac{\hat{\gamma}_{y^*}(1)}{\hat{\phi}} \right)^2 \right\}, \tag{34}$$

where the moment condition is derived from the expression for $\mathrm{Var}[\log u_t^2]$ under GED errors.

**Assumption 4.3** *Moment condition identification for SV-GED.*

*Define the population moment function*

$$f^{\mathcal{GED}}(\nu) := \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \gamma_{y^*}(0) + \frac{1}{\phi}\gamma_{y^*}(1),$$

*where* $\gamma_{y^*}(h) := \mathrm{Cov}(y_t^*, y_{t-h}^*)$. *We assume:*

- *(C1)* $f^{\mathcal{GED}}(\nu)$ *is continuously differentiable on an open neighborhood* $\mathcal{N} \subset (0, \infty)$ *containing* $\nu_0$,
- *(C2)* $f^{\mathcal{GED}}(\nu) = 0$ *has a unique solution at* $\nu = \nu_0 \in \mathcal{N}$,
- *(C3)* $\partial f^{\mathcal{GED}}(\nu)/\partial\nu\big|_{\nu=\nu_0} \neq 0$.

**Assumption 4.4** *Moment and variance conditions for SV-GED.*

- *(D1) GED errors admit moments of all orders, including* $\mathbb{E}[\log(u_t^2)^4] < \infty$,
- *(D2) The long-run variance*

$$\sigma_f^2 := \mathrm{Var}\left(\hat{\gamma}_{y^*}(0) - \frac{1}{\phi}\hat{\gamma}_{y^*}(1)\right)$$

*is finite and strictly positive.*

We now derive the asymptotic properties of the estimator $\hat{\nu}$. Under Assumption 2.1, the latent log-volatility process $\{w_t\}$ is strictly stationary, ergodic, and $\beta$-mixing with exponential decay, and the innovation process $\{u_t\}$ is i.i.d. and independent of $\{w_t\}$.

For the Student-$t$ case, the transformed squared innovations can be expressed as $\log(u_t^2) = \log(z_t^2) - \log(\lambda_t/\nu)$, where $z_t \sim \mathcal{N}(0,1)$ and $\lambda_t \sim \chi^2_\nu/\nu$. Hence, $\log(u_t^2)$ follows a log-$F$ distribution and has finite variance when $\nu > 2$, and finite higher moments for larger $\nu$. For the GED case, $\log(u_t^2)$ admits finite moments of all orders.

Consequently, the process $y_t^* := \log(y_t^2) - \mathbb{E}[\log(y_t^2)]$ is strictly stationary, ergodic, and $\beta$-mixing, with finite variance and higher moments under mild regularity. This follows from the fact that $y_t^*$ is a measurable function of $w_t$ and $u_t$, and that $\beta$-mixing is preserved under measurable transformations when suitable moment conditions hold; see Bradley (2005).

When $\mathbb{E}[\log(u_t^2)^4] < \infty$–which holds when $\nu > 4$ in the Student-$t$ case and always in the GED case–then the transformed process $y_t^*$ has finite fourth moments, implying that the original process $y_t$ has finite eighth moments. These moment conditions, together with the stationarity and $\beta$-mixing properties of the SV model, ensure that the sample autocovariances used in the estimation of $\nu$ admit a well-behaved asymptotic theory. In the lemmas and theorems that follow, we formally establish the consistency and asymptotic distribution of the estimator $\hat{\nu}$. The proofs draw on classical ergodic theorems and central limit theorems for dependent processes, as well as general results on the consistency and asymptotic distribution of moment-based estimators previously established in Ahsan et al. (2025a).

**Lemma 4.1** *Consistency of $\hat{\nu}$ with Student-t errors.*

*Let Assumptions 2.1 and 3.1 hold, and suppose the population moment condition*

$$f^{St}(\nu) := \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\nu}{2}\right) - \gamma_{y^*}(0) + \frac{\gamma_{y^*}(1)}{\phi}$$

*has a unique solution at the true value $\nu_0 \in \mathcal{N} \subset (4, \infty)$. Then the estimator $\hat{\nu}$ defined in (20) or (33) are consistent:*

$$\hat{\nu} \xrightarrow{p} \nu_0 \quad \text{as } T \to \infty.$$

**Theorem 4.1** *Asymptotic distribution of $\hat{\nu}$ with Student-t errors.*

*Let Assumptions 2.1, 3.1, and 4.1 –4.2 hold. Then the estimator $\hat{\nu}$ defined implicitly by*

$$f_T^{St}(\hat{\nu}) := \psi^{(1)}\left(\tfrac{1}{2}\right) + \psi^{(1)}\left(\tfrac{\hat{\nu}}{2}\right) - \hat{\gamma}_{y^*}(0) + \frac{1}{\hat{\phi}}\hat{\gamma}_{y^*}(1) = 0$$

*satisfies the asymptotic normality:*

$$\sqrt{T}(\hat{\nu} - \nu_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

*where*

$$\sigma^2 = \frac{\sigma_f^2}{\left[\frac{1}{2} \cdot \psi^{(2)} \left(\frac{\nu_0}{2}\right)\right]^2}, \quad \sigma_f^2 := \text{Var}\left(\hat{\gamma}_{y^*}(0) - \frac{1}{\phi}\hat{\gamma}_{y^*}(1)\right).$$

**Lemma 4.2** *Consistency of $\hat{\nu}$ with GED errors.*

*Let Assumptions 2.1 and 3.2 hold, and suppose the population moment condition*

$$f^{\mathcal{GED}}(\nu) := \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \gamma_{y^*}(0) + \frac{\gamma_{y^*}(1)}{\phi}$$

*has a unique root at $\nu_0 \in \mathcal{N}$. Then the estimator $\hat{\nu}$ defined in (28) or (34) are consistent:*

$$\hat{\nu} \xrightarrow{p} \nu_0 \quad \text{as } T \to \infty.$$

**Theorem 4.2** *Asymptotic distribution of $\hat{\nu}$ with GED errors.*

*Let Assumptions 2.1, 3.2, and 4.3–4.4 hold. Then the estimator $\hat{\nu}$ defined implicitly by*

$$f_T^{\mathcal{GED}}(\hat{\nu}) := \left(\frac{2}{\hat{\nu}}\right)^2 \psi^{(1)}\left(\frac{1}{\hat{\nu}}\right) - \hat{\gamma}_{y^*}(0) + \frac{1}{\hat{\phi}}\hat{\gamma}_{y^*}(1) = 0$$

*satisfies:*

$$\sqrt{T}(\hat{\nu} - \nu_0) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

*where*

$$\sigma^2 = \frac{\sigma_f^2}{\left[-\frac{8}{\nu^3} \cdot \psi^{(1)}\left(\frac{1}{\nu}\right) - \frac{4}{\nu^4} \cdot \psi^{(2)}\left(\frac{1}{\nu}\right)\right]^2}, \quad \sigma_f^2 := \text{Var}\left(\hat{\gamma}_{y^*}(0) - \frac{1}{\phi}\hat{\gamma}_{y^*}(1)\right).$$

**Remark 4.1 (Verification of Regularity and Identification Conditions)** *The continuity and differentiability conditions in Assumptions 4.1 and 4.3 follow directly from the analytic properties of the polygamma functions. In particular, the trigamma function $\psi^{(1)}(x)$ and higher-order polygamma functions are infinitely differentiable ($C^\infty$) on $\mathbb{R}^+$, and $\psi^{(1)}(x)$ is strictly decreasing for $x > 0$. As a result, the mapping*

$$f(\nu) = \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\nu}{2}\right) - \gamma_{y^*}(0) + \frac{\gamma_{y^*}(1)}{\phi}$$

*is continuously differentiable and strictly monotone on the relevant domain $\nu \in \mathcal{D}_t \subset (0, \infty)$, ensuring that a unique solution $\nu_0$ exists. The same reasoning applies to the GED case, where*

*$(2/\nu)^2\psi^{(1)}(1/\nu)$ is also smooth and strictly monotone on $\mathbb{R}^+$. Hence, the identification and regularity conditions invoked in Assumptions 4.1 and 4.3 hold automatically by construction of $f(\nu)$.*

**Remark 4.2** *Robust inference when $\nu \leq 4$ in Student-t case.*

*Asymptotic normality of $\hat{\nu}$ under the Student-t model requires $\nu > 4$ to ensure that the transformed process $y_t^* = \log(y_t^2) - \mathbb{E}[\log(y_t^2)]$ has finite fourth moments. However, in many empirically relevant applications–including our simulation and empirical results–$\nu$ may be close to or below this threshold (e.g., $\nu = 3$), in which case the asymptotic distribution of $\hat{\nu}$ may be non-normal or undefined. To address this, we complement the asymptotic theory with finite-sample simulation-based inference (e.g., Monte Carlo tests), which remain valid under heavier tails and weak regularity conditions.*

**Remark 4.3** *Robustness under GED tails.*

*Unlike the Student-t case, GED distributions admit moments of all orders regardless of the value of $\nu$. This implies that the asymptotic distribution of $\hat{\nu}$ remains valid even for strongly leptokurtic innovations. The higher-order differentiability of the GED moment condition facilitates inference without requiring supplementary simulation techniques. Nonetheless, if the process exhibits time variation in tail behavior, simulation-based inference may still be preferred.*

## 5 Hypothesis testing

Inference in stochastic volatility (SV) models can, in principle, be conducted using asymptotic tests based on likelihood ratio (LR)-type statistics. However, these tests rely on standard regularity conditions that are not always satisfied in SV settings, potentially leading to nonstandard limiting distributions and poor size control in finite samples.

To address these limitations, we complement asymptotic inference with simulation-based testing procedures, specifically Monte Carlo (MC) tests following Dufour (2006). These methods provide exact or locally level-corrected inference by directly approximating the sampling distribution of the test statistic under the null. While such procedures have traditionally been difficult to implement in SV models due to the high computational demands of alternative estimation procedures, our simple and computational efficient moment-based estimators makes these tests feasible even in large-scale applications.

Further, the asymptotic distribution of $\hat{\nu}$ derived above holds under the condition that $\mathbb{E}[\log(u_t^2)^4] < \infty$, which is satisfied for the Student-$t$ distribution only when the degrees of freedom $\nu > 4$. For cases where $\nu \leq 4$, such as $\nu = 3$ considered in our simulations and empirical application, the fourth moment of the transformed process $y_t^* := \log(y_t^2) - \mathbb{E}[\log(y_t^2)]$ is not finite, and the Central Limit Theorem for $\beta$-mixing processes does not apply. In such heavy-tailed settings, asymptotic normality may fail, motivating the use of simulation-based methods such as Monte Carlo tests discussed for valid inference on $\nu$.

In this section, we describe both the asymptotic and simulation-based versions of our proposed GMM-based LR-type test. We begin by outlining the construction of the test statistic and then describe how to implement both inference procedures. While prior work has focused on testing model dynamics–such as persistence, autoregressive order, or leverage effects in SV(1) and SV($p$) models Ahsan and Dufour (2019, 2021); Ahsan et al. (2025a)–our contribution here is to test the distributional assumptions of the innovations, particularly normality versus heavy tails.

The SV models we consider incorporate heavy-tailed error distributions and are characterized by four parameters: $\theta = (\phi, \sigma_y, \sigma_v, \nu)$, where $\nu$ is the degrees of freedom parameter and hence governs the tail behavior. The LR-type statistics we porpose using are based on the following moment-based GMM objective function:

$$M_T(\theta) := g_T(\theta)' A_T, g_T(\theta), \tag{35}$$

where $g_T(\theta)$ is a $4 \times 1$ vector of moment conditions, defined later and depending on the assumed error distribution.

Since the number of moment conditions equals the number of parameters, the model is just-identified. In this case, we could set $A_T = I_4$ and use the simplified GMM objective function

$$M_T^I(\theta) = g_T(\theta)' g_T(\theta). \tag{36}$$

However, to ensure that the LR-type statistic in (40) follows a $\chi^2$ distribution asymptotically, it is preferable to use an efficient weighting matrix, $A_T = \hat{\Omega}_*^{-1}$, where $\hat{\Omega}_*$ is a consistent estimator of the long-run variance:

$$\hat{\Omega}_* = \hat{\Gamma}_0 + \sum_{k=1}^{K(T)} \left(1 - \frac{k}{K(T)+1}\right) \left(\hat{\Gamma}_k + \hat{\Gamma}_k'\right), \tag{37}$$

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} g_t(\theta) g_t(\theta)'. \tag{38}$$

The resulting efficient GMM-type objective function is

$$M_T^*(\theta) = g_T(\theta)'\hat{\Omega}_*^{-1} g_T(\theta). \tag{39}$$

To test hypotheses on $\theta$, we compute the LR-type test statistic as the difference between the restricted and unrestricted values of the efficient GMM criterion:

$$LR_T = T \times \left[ M_T(\hat{\theta}_0) - M_T(\hat{\theta}) \right], \tag{40}$$

where $\hat{\theta}$ is the unrestricted estimator and $\hat{\theta}_0$ is the estimator under the null hypothesis. Under standard conditions, $LR_T \xrightarrow{d} \chi_r^2$, where $r$ is the number of restrictions; see Newey and West (1987), Newey and McFadden (1994), and Dufour et al. (2017). However, these standard regularity conditions may fail to hold when some parameters are unidentified under the null, or when the null lies on the boundary of the parameter space–both of which are common in SV models. In such cases, the asymptotic distribution may be nonstandard, and the weighting matrix $\hat{\Omega}_*$ may be ill-defined or near-singular. These issues motivate the use of MC-based inference to provide size-correct testing. As described in Ahsan et al. (2025a), in such cases, it may be preferable to use $A_T = I_4$. Importantly, the Monte Carlo test remains valid regardless, as it does not rely on the regularity conditions required for the $\chi_r^2$ approximation or on consistent estimation of $\hat{\theta}_0$.

The asymptotic version of the test relies on critical values from the $\chi_r^2$ distribution, where $r$ denotes the number of restrictions imposed under the null. Test statistics are evaluated by simulating data under the null or alternative, and empirical size and power are assessed by computing rejection frequencies over repeated simulations. As shown in Section 6, this approach tends to over-reject under the null. To address this, when evaluating the power of the test, we implement a locally level-corrected version of the asymptotic test. This version adjusts the critical region by replacing the standard $\chi_r^2$ quantile with an empirical critical value obtained from an approximation of the null distribution. Specifically, we simulate a large number of samples under the null using a large sample size, compute the LR statistic for each, and use the empirical $(1 - \alpha)$ quantile as the test's critical value. This yields a corrected rejection frequency that more accurately reflects the nominal size and facilitates meaningful power comparisons. While infeasible in practice–since it relies on knowledge of the true DGP–this procedure provides a useful benchmark for assessing the relative performance of alternative tests.

The Monte Carlo version of the test constructs the null distribution directly by simulating data from the model under the null hypothesis, using parameter estimates that satisfy the imposed restrictions. Specifically, for each replication, we simulate a dataset from the restricted model using the estimated parameters $\hat{\theta}_0$, compute the LR-type statistic for this dataset, and repeat the process $N$ times to generate an empirical distribution of the test statistic under the null. Given the observed test statistic $LR_0$, computed using (40) from observed data, the local MC p-value is

$$p_N^{\text{LMC}} = \frac{N + 1 - \sum_{i=1}^{N} \mathcal{I}\{LR_0 \geq LR_*^{(i)}\}}{N + 1} \tag{41}$$

where $\mathcal{I}\{\cdot\}$ is an indicator function and $LR_*^{(i)}{}_{i=1}^{N}$ is the $N \times 1$ vector of test statistics computed from the simulated null datasets. This p-value represents the rank of $LR_0$ relative to the empirical distribution of the test statistic under the null, and does not rely on asymptotic approximations or regularity conditions. The subscript $N$ indicates that the p-value is based on $N$ simulated replications of the null distribution. The superscript LMC denotes the local Monte Carlo test, in which the null distribution is simulated using parameter estimates obtained under the imposed restrictions, making the simulation local to that region of the parameter space.

Importantly, the maximized version of the Monte Carlo test (MMC) extends the local procedure by considering a broader set of parameter values consistent with the null hypothesis. This yields a test that is robust to identification issues and remains exact in finite samples. Specifically, let the vector of nuisance parameters be $\tilde{\theta} = (\phi, \sigma_y, \sigma_\nu)'$. In the local Monte Carlo (LMC) test, we use the point estimates $\hat{\tilde{\theta}}_0 = (\hat{\phi}^0, \hat{\sigma}y^0, \hat{\sigma}\nu^0)'$, obtained under the imposed restrictions, to simulate the null distribution of the test statistic. In contrast, the MMC test evaluates the p-value over a wider set of nuisance parameter values consistent with the null. Specifically, we define the constrained set:

$$\begin{aligned} C_T(\nu_0) = \Big\{ \tilde{\theta}_0 \in \tilde{\Omega}_0 : \left|\phi - \hat{\phi}^0\right| \leq 0.01, ; |\phi| \leq 0.99, ; \left|\sigma_y - \hat{\sigma}_y^0\right| \leq 0.05, ; \sigma_y \geq 0.01, ; \\ \left|\sigma_\nu - \hat{\sigma}_\nu^0\right| \leq 0.05, ; \sigma_\nu \geq 0.01 \Big\} . \end{aligned} \tag{42}$$

While it is possible to maximize over the full nuisance parameter space, restricting the search to a neighborhood around $\hat{\tilde{\theta}}_0$ helps reduce the computational burden of the optimization. For each $\tilde{\theta}_0 \in C_T(\nu_0)$, we simulate the null distribution of the test statistic, compute the corresponding Monte Carlo p-value as before, and then select the maximum p-value over this set, yielding $p_N^{\text{MMC}}$. For further details on the MMC framework and its theoretical properties, see Dufour (2006).

Unlike the asymptotic version, we do not report a locally level-corrected version of the Monte

Carlo tests, as they exhibit little to no over-rejection under the null. Consequently, in the power comparisons that follow, we evaluate the performance of the LMC and MMC procedures relative to the locally level-corrected version of the asymptotic test. This approach allows us to assess power while preserving finite-sample validity under the null.

We now turn to the moment conditions used to construct the GMM-based LR-type tests under both the asymptotic and Monte Carlo frameworks, focusing on models with Student's t and GED innovations, and on testing both normality against heavy tails and heavy tails against normality.

### 5.1 Student-$t$: Testing for Normality and Heavy Tails

For the SV(1)-$t$ model, we define the moment vector $g_T^{stu}(\theta)$ as

$$g_T^{stu}(\theta) = \begin{bmatrix} \hat{\mu} - \psi\left(\frac{1}{2}\right) + \psi\left(\frac{\hat{\nu}}{2}\right) - log(\hat{\nu}) - log(\hat{\sigma}_y^2) \\ \hat{\gamma}(1)\frac{(1-\hat{\phi}^2)}{\hat{\phi}} - \hat{\sigma}_\nu^2 \\ \hat{\gamma}_{y^*}(2) - \hat{\phi}\hat{\gamma}_{y^*}(1) \\ \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\hat{\nu}}{2}\right) - \hat{\gamma}_{y^*}(0) + \hat{\gamma}_{y^*}(1)/\hat{\phi} \end{bmatrix} \tag{43}$$

where the superscript "stu" indicates that these moments are specific to the case with Student-$t$ innovations, distinguishing them from the GED case discussed below.

Since $u_t$ is not normally distributed, the process $\epsilon_t$ is not simply i.i.d. $\log(\chi^2)$ and the moment conditions must be adjusted accordingly. The first moment is obtained by taking the logarithm of equation (18). The second moment follows from summing the Yule-Walker autocovariance identities $\hat{\gamma}(0)$ and $\hat{\gamma}(1)$, but uses the non-Gaussian expression for $\sigma_\epsilon^2$. Specifically, we substitute (13) into the autocovariance structure given in (7) to account for the heavy-tailed nature of the errors as follows

$$\begin{aligned} \gamma(0) + \gamma(1) - \phi\gamma(1) - \sigma_\nu^2 - \sigma_\epsilon^2 - \phi\gamma(0) + \phi\sigma_\epsilon^2 = 0 \\ \gamma(0)(1-\phi) + \gamma(1)(1-\phi) - \sigma_\nu^2 - \sigma_\epsilon^2(1-\phi) = 0 \\ \gamma(0)(1-\phi) + \gamma(1)(1-\phi) - (\gamma(0) - \frac{1}{\phi}\gamma(1))(1-\phi) - \sigma_\nu^2 = 0 \\ \gamma(1)(1-\phi) + \frac{(1-\phi)}{\phi}\gamma(1) - \sigma_\nu^2 = 0 \\ \gamma(1)\frac{(1-\phi^2)}{\phi} - \sigma_\nu^2 = 0 \end{aligned}$$

yielding the second moment condition. The third moment is simply the autocovariance equation in (7) when $k = 2$ and the fourth moment equation is from the optimization problem (20).

In order to estimate $\hat{\Omega}_*$, we use

$$g_t^{stu}(\theta) = \begin{bmatrix} \log(y_t^2) - \psi\left(\frac{1}{2}\right) + \psi\left(\frac{\hat{\nu}}{2}\right) - log(\hat{\nu}) - log(\hat{\sigma}_y^2) \\ y_t^* y_{t+1}^* \frac{(1-\hat{\phi}^2)}{\hat{\phi}} - \hat{\sigma}_\nu^2 \\ y_t^* y_{t+2}^* - \hat{\phi} y_t^* y_{t+1}^*) \\ \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\hat{\nu}}{2}\right) - y_t^* y_t^* + y_t^* y_{t+1}^* / \hat{\phi} \end{bmatrix} \tag{44}$$

giving the GMM-type objective function

$$M_T^{stu,*}(\theta) = g_T^{stu}(\theta)' \left(\hat{\Omega}_*^{stu}\right)^{-1} g_T^{stu}(\theta) \,. \tag{45}$$

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

$$LR_T^{stu} = T \times [M_T^{stu,*}(\hat{\theta}_0) - M_T^{stu,*}(\hat{\theta})] \tag{46}$$

Testing for normality in the Student-$t$ setting is not entirely straightforward. In theory, as the degrees of freedom parameter $\nu$ approaches infinity, the Student-$t$ distribution converges to the normal distribution. However, for sufficiently large but finite values of $\nu$, the distribution may already closely approximate normality. To account for this, we test the hypothesis:

$$H_0^{s,n} : \nu = 30 \tag{47}$$

$$H_1^{s,n} : \nu \neq 30 \tag{48}$$

where $\nu = 30$ is chosen to represent a distribution that is effectively Gaussian for practical purposes. The parameter $\nu$ appears in the first and fourth moments of the moment vector $g_T^{stu}(\theta)$ defined above. As before, the full parameter vector is given by $\theta = (\phi, \sigma_y, \sigma_\nu, \nu)'$ and the nuisance parameter vector is $\tilde{\theta} = (\phi, \sigma_y, \sigma_\nu)'$.

Notably, the second and third components of $g_T^{stu}(\theta)$ do not necessarily depend on $\nu$–for example, when using expressions such as (13) and (14)–and therefore their estimates remain unchanged under the null and alternative. In the simulation section, we consider alternatives with $\nu = 3$, 5, and 10 to evaluate the power of this test.

To test for heavy tails, we consider the following hypothesis test

$$H_0^{s,h} : \nu = 10 \tag{49}$$

$$H_1^{s,h} : \nu \neq 10 \tag{50}$$

We choose $\nu = 10$ based on prior literature that finds evidence of fat tails in SV models with Student-$t$ errors, with estimated degrees of freedom often around this value (see, for example, Jacquier et al. (2004)). However, our simulation results in the next section suggest that this test has limited power against a Gaussian alternative with $\nu = 30$, implying that $\nu = 10$ may still be too close to normality to generate meaningful separation. To address this, we also report results in the appendix for tests involving more extreme tail behavior under the null, namely:

$$H^{s,h,3}0 : \nu = 3 \quad \text{and} \quad H^{s,h,5}0 : \nu = 5 \tag{51}$$

with $\nu = 30$ as the alternative in both cases. These additional settings allow us to assess test performance when heavier tails are clearly present under the null and lead to better power performance.

### 5.2 GED: Testing for Normality and Heavy Tails

To test for normality and heavy tails under the assumption of GED errors, we use the following moment conditions for the GMM-based LR-type test:

$$g_T^{ged}(\theta) = \begin{bmatrix} \hat{\mu} - \frac{2}{\hat{\nu}}\psi\left(\frac{1}{\hat{\nu}}\right) - log\left[\Gamma\left(\frac{1}{\hat{\nu}}\right)\right] + log\left[\Gamma\left(\frac{3}{\hat{\nu}}\right)\right] - log(\hat{\sigma}_y^2) \\ \hat{\gamma}(1)\frac{(1-\hat{\phi}^2)}{\hat{\phi}} - \hat{\sigma}_\nu^2 \\ \hat{\gamma}_{y^*}(2) - \phi\hat{\gamma}_{y^*}(1) \\ \left(\frac{2}{\hat{\nu}}\right)^2 \psi^{(1)}\left(\frac{1}{\hat{\nu}}\right) - \hat{\gamma}_{y^*}(0) + \hat{\gamma}_{y^*}(1)/\hat{\phi} \end{bmatrix} \tag{52}$$

The superscript "ged" indicates that these moments are derived under the GED error assumption. The first moment is obtained by taking the logarithm of equation (26). The second moment is based on the autocovariance identity involving $\hat{\gamma}(0)$ and $\hat{\gamma}(1)$, using the expression for $\sigma_\epsilon^2$ as was shown in the Student-t case. The third moment corresponds to the autocovariance condition at lag $k = 2$, as in the Student-$t$ case. The fourth moment arises from the identification condition associated with the estimation of the shape parameter $\nu$, specifically from the optimization problem in equation (28).

In order to estimate $\hat{\Omega}_*$ in the GED error case, we use

$$g_t^{ged}(\theta) = \begin{bmatrix} \log(y_t^2) - \frac{2}{\hat{\nu}}\psi\left(\frac{1}{\hat{\nu}}\right) - log\left[\Gamma\left(\frac{1}{\hat{\nu}}\right)\right] + log\left[\Gamma\left(\frac{3}{\hat{\nu}}\right)\right] - log(\hat{\sigma}_y^2) \\ y_t^* y_{t+1}^* \frac{(1-\hat{\phi}^2)}{\hat{\phi}} - \hat{\sigma}_\nu^2 \\ y_t^* y_{t+2}^* - \phi y_t^* y_{t+1}^* \\ \left(\frac{2}{\hat{\nu}}\right)^2 \psi^{(1)}\left(\frac{1}{\hat{\nu}}\right) - y_t^* y_t^* + y_t^* y_{t+1}^* / \hat{\phi} \end{bmatrix} \tag{53}$$

giving the GMM-type objective function

$$M_T^{ged,*}(\theta) = g_T^{ged}(\theta)' \left(\hat{\Omega}_*^{ged}\right)^{-1} g_T^{ged}(\theta)\,. \tag{54}$$

and the LR-type statistic given by the objective function:

$$LR_T^{ged} = T \times [M_T^{ged,*}(\hat{\theta}_0) - M_T^{ged,*}(\hat{\theta})] \tag{55}$$

which we use to test hypotheses on $\theta$ in settings with GED errors.

In this case, testing for normality under the GED specification is relatively straightforward, as the GED distribution coincides with the normal distribution when $\nu = 2$. Accordingly, we consider the following hypothesis test:

$$H_0^{g,n} : \nu = 2 \tag{56}$$

$$H_1^{g,n} : \nu \neq 2 \tag{57}$$

This restriction is reflected in the first and fourth moments of $g_T^{ged}(\theta)$ defined above. To evaluate the size of the test, we simulate data under the null with $\nu = 2$. Meanwhile, to assess power we consider alternative data-generating processes with heavier tails, specifically $\nu = 1.5$, $\nu = 1.2$, or $\nu = 1.0$, which correspond to progressively heavier-tailed distributions and find that power increases as $\nu$ decreases, as should be expected.

To test for heavy tails, we use the same moment conditions, but consider the following hypothesis:

$$H_0^{g,h} : \nu = 1.5 \tag{58}$$

$$H_1^{g,h} : \nu \neq 1.5 \tag{59}$$

In this case, size is evaluated under a DGP with $\nu = 1.5$, while power is assessed using a more

Gaussian alternative with $\nu = 2$. While we could alternatively test null hypotheses with $\nu = 1.2$ or $\nu = 1.0$, we treat $\nu = 1.5$ as a representative upper bound for distributions with clearly heavy tails.

# 6 Simulation study

In this section, we assess–via simulation–the estimation properties of the winsorised ARMA-SV (W-ARMA) estimators with heavy-tailed distributions proposed in this paper. Similarly, we also asses the empirical performance of the asymptotic, Local Monte Carlo (LMC), and Maximized Monte Carlo (MMC) tests for normality and heavy tails.

We consider both Student-$t$ and GED specifications under two persistence regimes: moderate persistence ($\phi = 0.90$) and high persistence ($\phi = 0.95$). We also vary the sample size and the degrees of freedom parameter $\nu$ to capture different degrees of tail heaviness, including values that approximate or are equivalent to normality.

## 6.1 Estimation performance

For the estimation results, we consider both $\mathrm{SV}_t$ and $\mathrm{SV}_{\mathcal{GED}}$ models. We generate $N_{\mathrm{rep}} = 10{,}000$ replications for each design, with sample sizes $T \in \{1000, 2000, 5000\}$ and winsorisation windows $J \in \{10, 50, 100, 200\}$. For each parameter $\theta \in \{\phi, \sigma_y, \sigma_\nu, \nu\}$, we compute bias and RMSE across replications as:

$$\mathrm{Bias}(\hat{\theta}) = \frac{1}{N_{\mathrm{rep}}} \sum_{r=1}^{N_{\mathrm{rep}}} \left(\hat{\theta}^{(r)} - \theta\right), \qquad \mathrm{RMSE}(\hat{\theta}) = \sqrt{\frac{1}{N_{\mathrm{rep}}} \sum_{r=1}^{N_{\mathrm{rep}}} \left(\hat{\theta}^{(r)} - \theta\right)^2},$$

where $\hat{\theta}^{(r)}$ is the estimate of the true parameters $\theta = \{\phi, \sigma_y, \sigma_\nu, \nu\}$ from the $r^{\mathrm{th}}$ replication.

In the Student-$t$ designs, we report the RMSE for $\nu$ relative to its true value (i.e., RMSE/$\nu$) to facilitate comparison across different levels of tail heaviness, especially since estimating $\nu$ is known to be challenging in the Student-$t$ setting. This normalization is also helpful given that many prior studies report dispersion in $\nu$ using nonstandard metrics.[5]

Table 1 presents bias and RMSE results for the $\mathrm{SV}_t$ model under moderate ($\phi = 0.90$, left panel) and high ($\phi = 0.95$, right panel) persistence. We normalize $\sigma_y = 1.00$ and set $\sigma_\nu = 1.50$ to reflect relatively higher volatility in the log-volatility process. For tail thickness, we consider $\nu = 3$ in the

[5]Difficulty in estimating large values of $\nu$ has led other authors to report alternative dispersion measures. For example, Jacquier et al. (2004) report the average of the first and third quartiles from their posterior distribution.

Table 1: Estimation of SV models with Student-$t$ errors

| $T$ | | | Moderate persistence | | | | High persistence | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ |
| | Heavy tails | | 0.90 | 1.00 | 1.50 | 3.00 | 0.95 | 1.00 | 1.50 | 3.00 |
| 1000 | W-ARMA ($J = 10$) | Bias | -0.007 | -0.392 | 0.022 | 2.316 | -0.005 | -0.342 | 0.019 | 2.355 |
| | | RMSE | 0.022 | 0.421 | 0.125 | 4.301 | 0.015 | 0.477 | 0.117 | 5.221 |
| | W-ARMA ($J = 50$) | Bias | -0.004 | -0.394 | -0.001 | 2.226 | -0.005 | -0.341 | 0.018 | 2.307 |
| | | RMSE | 0.022 | 0.423 | 0.130 | 4.612 | 0.015 | 0.476 | 0.141 | 4.361 |
| | W-ARMA ($J = 100$) | Bias | 0.002 | -0.399 | -0.046 | 1.743 | -0.002 | -0.347 | -0.038 | 1.876 |
| | | RMSE | 0.022 | 0.427 | 0.143 | 4.589 | 0.014 | 0.479 | 0.150 | 4.642 |
| | W-ARMA ($J = 200$) | Bias | 0.010 | -0.406 | -0.120 | 0.765 | 0.005 | -0.357 | -0.137 | 0.895 |
| | | RMSE | 0.024 | 0.433 | 0.185 | 2.216 | 0.014 | 0.483 | 0.206 | 2.942 |
| 2000 | W-ARMA ($J = 10$) | Bias | -0.004 | -0.407 | 0.012 | 0.941 | -0.003 | -0.383 | 0.009 | 1.020 |
| | | RMSE | 0.015 | 0.421 | 0.091 | 1.833 | 0.009 | 0.440 | 0.084 | 3.303 |
| | W-ARMA ($J = 50$) | Bias | -0.002 | -0.408 | 0.000 | 0.936 | -0.003 | -0.382 | 0.010 | 0.901 |
| | | RMSE | 0.016 | 0.422 | 0.099 | 2.610 | 0.010 | 0.439 | 0.108 | 2.148 |
| | W-ARMA ($J = 100$) | Bias | 0.001 | -0.412 | -0.027 | 0.685 | -0.001 | -0.387 | -0.020 | 0.694 |
| | | RMSE | 0.016 | 0.426 | 0.105 | 2.638 | 0.010 | 0.442 | 0.111 | 1.992 |
| | W-ARMA ($J = 200$) | Bias | 0.007 | -0.418 | -0.074 | 0.358 | 0.003 | -0.394 | -0.079 | 0.384 |
| | | RMSE | 0.017 | 0.431 | 0.128 | 1.931 | 0.010 | 0.447 | 0.140 | 1.901 |
| 5000 | W-ARMA ($J = 10$) | Bias | -0.001 | -0.416 | 0.004 | 0.251 | -0.001 | -0.405 | 0.003 | 0.219 |
| | | RMSE | 0.009 | 0.421 | 0.059 | 0.990 | 0.006 | 0.426 | 0.053 | 0.338 |
| | W-ARMA ($J = 50$) | Bias | -0.001 | -0.417 | -0.001 | 0.212 | -0.001 | -0.405 | 0.004 | 0.287 |
| | | RMSE | 0.010 | 0.422 | 0.065 | 0.439 | 0.007 | 0.425 | 0.073 | 1.317 |
| | W-ARMA ($J = 100$) | Bias | 0.001 | -0.418 | -0.013 | 0.148 | -0.000 | -0.406 | -0.009 | 0.195 |
| | | RMSE | 0.010 | 0.424 | 0.067 | 0.350 | 0.006 | 0.427 | 0.074 | 0.511 |
| | W-ARMA ($J = 200$) | Bias | 0.003 | -0.422 | -0.035 | 0.045 | 0.001 | -0.411 | -0.036 | 0.075 |
| | | RMSE | 0.010 | 0.427 | 0.076 | 0.385 | 0.007 | 0.430 | 0.085 | 0.546 |
| | Normal (approx.) | | 0.90 | 1.00 | 1.50 | 30.00 | 0.95 | 1.00 | 1.50 | 30.00 |
| 1000 | W-ARMA ($J = 10$) | Bias | -0.001 | -0.082 | -0.030 | -14.085 | -0.003 | 0.002 | -0.022 | -13.602 |
| | | RMSE | 0.020 | 0.244 | 0.120 | 1.279 | 0.013 | 0.502 | 0.113 | 1.311 |
| | W-ARMA ($J = 50$) | Bias | 0.002 | -0.084 | -0.057 | -15.002 | -0.002 | 0.003 | -0.042 | -14.069 |
| | | RMSE | 0.020 | 0.245 | 0.139 | 1.242 | 0.013 | 0.505 | 0.146 | 1.307 |
| | W-ARMA ($J = 100$) | Bias | 0.008 | -0.087 | -0.105 | -16.162 | 0.002 | -0.001 | -0.096 | -15.175 |
| | | RMSE | 0.022 | 0.244 | 0.170 | 1.169 | 0.013 | 0.506 | 0.174 | 1.245 |
| | W-ARMA ($J = 200$) | Bias | 0.018 | -0.097 | -0.184 | -17.777 | 0.008 | -0.017 | -0.197 | -18.292 |
| | | RMSE | 0.027 | 0.247 | 0.232 | 1.185 | 0.015 | 0.494 | 0.251 | 1.121 |
| 2000 | W-ARMA ($J = 10$) | Bias | 0.000 | -0.076 | -0.024 | -9.998 | -0.001 | -0.037 | -0.018 | -9.512 |
| | | RMSE | 0.013 | 0.181 | 0.085 | 1.425 | 0.009 | 0.335 | 0.079 | 1.410 |
| | W-ARMA ($J = 50$) | Bias | 0.002 | -0.078 | -0.040 | -10.197 | -0.000 | -0.036 | -0.032 | -9.982 |
| | | RMSE | 0.014 | 0.181 | 0.100 | 1.424 | 0.010 | 0.337 | 0.108 | 1.399 |
| | W-ARMA ($J = 100$) | Bias | 0.006 | -0.081 | -0.067 | -11.675 | 0.002 | -0.040 | -0.060 | -11.915 |
| | | RMSE | 0.016 | 0.183 | 0.116 | 1.389 | 0.009 | 0.338 | 0.122 | 1.329 |
| | W-ARMA ($J = 200$) | Bias | 0.012 | -0.086 | -0.116 | -13.807 | 0.005 | -0.047 | -0.119 | -13.810 |
| | | RMSE | 0.019 | 0.186 | 0.153 | 1.317 | 0.011 | 0.335 | 0.164 | 1.296 |
| 5000 | W-ARMA ($J = 10$) | Bias | 0.001 | -0.061 | -0.014 | -2.888 | -0.000 | -0.042 | -0.012 | -2.572 |
| | | RMSE | 0.009 | 0.122 | 0.054 | 1.610 | 0.005 | 0.212 | 0.050 | 1.586 |
| | W-ARMA ($J = 50$) | Bias | 0.002 | -0.062 | -0.022 | -4.403 | 0.000 | -0.044 | -0.022 | -2.139 |
| | | RMSE | 0.009 | 0.123 | 0.063 | 1.542 | 0.006 | 0.211 | 0.071 | 1.693 |
| | W-ARMA ($J = 100$) | Bias | 0.003 | -0.063 | -0.034 | -5.183 | 0.001 | -0.046 | -0.033 | -4.069 |
| | | RMSE | 0.010 | 0.124 | 0.069 | 1.504 | 0.006 | 0.211 | 0.077 | 1.567 |
| | W-ARMA ($J = 200$) | Bias | 0.006 | -0.066 | -0.057 | -8.229 | 0.003 | -0.050 | -0.058 | -7.473 |
| | | RMSE | 0.011 | 0.126 | 0.084 | 1.388 | 0.007 | 0.213 | 0.093 | 1.364 |

Notes: Bias and RMSE are computed over 10,000 Monte Carlo replications. Moderate (high) persistence corresponds to $\phi = 0.90$ ($\phi = 0.95$). Variance parameters are set to $\sigma_y = 1$ and $\sigma_\nu = 1.5$. In the heavy-tail case, $\nu = 3.0$, while $\nu = 30.0$ is used to approximate normality. RMSE for $\nu$ is reported as RMSE/$\nu$. J denotes the winsorisation parameter.

Table 2: Estimation of SV models with Student-$\mathcal{GED}$ errors

| $T$ | | | Moderate persistence | | | | High persistence | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ |
| | Heavy tails | | 0.90 | 1.00 | 1.50 | 1.50 | 0.95 | 1.00 | 1.50 | 1.50 |
| 1000 | W-ARMA ($J=10$) | Bias | -0.007 | 0.025 | 0.017 | 0.158 | -0.005 | 0.115 | 0.015 | 0.134 |
| | | RMSE | 0.022 | 0.262 | 0.129 | 0.580 | 0.015 | 0.578 | 0.118 | 0.526 |
| | W-ARMA ($J=50$) | Bias | -0.003 | 0.033 | -0.011 | 0.110 | -0.005 | 0.115 | 0.010 | 0.141 |
| | | RMSE | 0.022 | 0.265 | 0.136 | 0.516 | 0.015 | 0.580 | 0.146 | 0.556 |
| | W-ARMA ($J=100$) | Bias | 0.003 | 0.048 | -0.062 | 0.040 | -0.001 | 0.134 | -0.052 | 0.056 |
| | | RMSE | 0.022 | 0.271 | 0.154 | 0.454 | 0.014 | 0.592 | 0.160 | 0.472 |
| | W-ARMA ($J=200$) | Bias | 0.013 | 0.072 | -0.146 | -0.060 | 0.006 | 0.167 | -0.161 | -0.068 |
| | | RMSE | 0.026 | 0.281 | 0.206 | 0.382 | 0.014 | 0.617 | 0.226 | 0.375 |
| 2000 | W-ARMA ($J=10$) | Bias | -0.004 | 0.013 | 0.009 | 0.072 | -0.003 | 0.057 | 0.007 | 0.063 |
| | | RMSE | 0.015 | 0.181 | 0.092 | 0.331 | 0.009 | 0.372 | 0.084 | 0.306 |
| | W-ARMA ($J=50$) | Bias | -0.002 | 0.017 | -0.004 | 0.055 | -0.003 | 0.057 | 0.006 | 0.069 |
| | | RMSE | 0.015 | 0.183 | 0.099 | 0.322 | 0.010 | 0.372 | 0.109 | 0.324 |
| | W-ARMA ($J=100$) | Bias | 0.002 | 0.026 | -0.033 | 0.018 | -0.001 | 0.067 | -0.027 | 0.028 |
| | | RMSE | 0.016 | 0.185 | 0.107 | 0.297 | 0.010 | 0.377 | 0.114 | 0.295 |
| | W-ARMA ($J=200$) | Bias | 0.008 | 0.041 | -0.086 | -0.044 | 0.004 | 0.085 | -0.091 | -0.042 |
| | | RMSE | 0.018 | 0.190 | 0.137 | 0.265 | 0.010 | 0.387 | 0.149 | 0.265 |
| 5000 | W-ARMA ($J=10$) | Bias | -0.001 | 0.004 | 0.003 | 0.028 | -0.001 | 0.020 | 0.003 | 0.028 |
| | | RMSE | 0.009 | 0.114 | 0.058 | 0.178 | 0.006 | 0.226 | 0.053 | 0.171 |
| | W-ARMA ($J=50$) | Bias | -0.001 | 0.006 | -0.002 | 0.022 | -0.001 | 0.020 | 0.003 | 0.033 |
| | | RMSE | 0.010 | 0.115 | 0.064 | 0.178 | 0.006 | 0.226 | 0.072 | 0.183 |
| | W-ARMA ($J=100$) | Bias | 0.001 | 0.009 | -0.014 | 0.007 | -0.000 | 0.023 | -0.009 | 0.018 |
| | | RMSE | 0.010 | 0.116 | 0.067 | 0.173 | 0.006 | 0.227 | 0.074 | 0.177 |
| | W-ARMA ($J=200$) | Bias | 0.004 | 0.016 | -0.038 | -0.020 | 0.001 | 0.031 | -0.037 | -0.012 |
| | | RMSE | 0.011 | 0.117 | 0.077 | 0.167 | 0.006 | 0.230 | 0.085 | 0.169 |
| | Normal | | 0.90 | 1.00 | 1.50 | 2.00 | 0.95 | 1.00 | 1.50 | 2.00 |
| 1000 | W-ARMA ($J=10$) | Bias | -0.006 | 0.036 | 0.006 | 0.241 | -0.005 | 0.128 | 0.007 | 0.233 |
| | | RMSE | 0.021 | 0.266 | 0.122 | 0.920 | 0.014 | 0.593 | 0.115 | 0.908 |
| | W-ARMA ($J=50$) | Bias | -0.002 | 0.044 | -0.022 | 0.174 | -0.005 | 0.130 | 0.000 | 0.230 |
| | | RMSE | 0.021 | 0.269 | 0.132 | 0.870 | 0.014 | 0.595 | 0.141 | 0.925 |
| | W-ARMA ($J=100$) | Bias | 0.005 | 0.058 | -0.074 | 0.057 | -0.000 | 0.148 | -0.059 | 0.100 |
| | | RMSE | 0.021 | 0.275 | 0.154 | 0.789 | 0.013 | 0.609 | 0.157 | 0.833 |
| | W-ARMA ($J=200$) | Bias | 0.015 | 0.082 | -0.161 | -0.113 | 0.007 | 0.180 | -0.168 | -0.112 |
| | | RMSE | 0.026 | 0.288 | 0.214 | 0.676 | 0.014 | 0.634 | 0.229 | 0.685 |
| 2000 | W-ARMA ($J=10$) | Bias | -0.003 | 0.011 | 0.006 | 0.178 | -0.002 | 0.052 | 0.006 | 0.169 |
| | | RMSE | 0.015 | 0.178 | 0.089 | 0.688 | 0.009 | 0.365 | 0.083 | 0.666 |
| | W-ARMA ($J=50$) | Bias | -0.001 | 0.016 | -0.009 | 0.140 | -0.002 | 0.051 | 0.005 | 0.185 |
| | | RMSE | 0.015 | 0.180 | 0.097 | 0.665 | 0.010 | 0.365 | 0.109 | 0.710 |
| | W-ARMA ($J=100$) | Bias | 0.003 | 0.025 | -0.039 | 0.060 | -0.000 | 0.061 | -0.027 | 0.100 |
| | | RMSE | 0.015 | 0.182 | 0.107 | 0.596 | 0.010 | 0.370 | 0.115 | 0.637 |
| | W-ARMA ($J=200$) | Bias | 0.009 | 0.041 | -0.093 | -0.067 | 0.004 | 0.079 | -0.090 | -0.042 |
| | | RMSE | 0.018 | 0.188 | 0.140 | 0.509 | 0.010 | 0.380 | 0.149 | 0.544 |
| 5000 | W-ARMA ($J=10$) | Bias | -0.001 | 0.005 | 0.003 | 0.078 | -0.001 | 0.023 | 0.002 | 0.076 |
| | | RMSE | 0.009 | 0.114 | 0.057 | 0.384 | 0.006 | 0.225 | 0.052 | 0.362 |
| | W-ARMA ($J=50$) | Bias | -0.000 | 0.007 | -0.004 | 0.065 | -0.001 | 0.023 | 0.002 | 0.086 |
| | | RMSE | 0.010 | 0.115 | 0.063 | 0.384 | 0.006 | 0.225 | 0.072 | 0.393 |
| | W-ARMA ($J=100$) | Bias | 0.001 | 0.011 | -0.017 | 0.032 | -0.000 | 0.027 | -0.011 | 0.053 |
| | | RMSE | 0.010 | 0.115 | 0.066 | 0.363 | 0.006 | 0.227 | 0.074 | 0.371 |
| | W-ARMA ($J=200$) | Bias | 0.004 | 0.019 | -0.042 | -0.027 | 0.002 | 0.035 | -0.040 | -0.012 |
| | | RMSE | 0.011 | 0.117 | 0.078 | 0.330 | 0.007 | 0.230 | 0.086 | 0.335 |

Notes: Bias and RMSE are computed over 10,000 Monte Carlo replications. Moderate (high) persistence corresponds to $\phi = 0.90$ ($\phi = 0.95$). Variance parameters are set to $\sigma_y = 1$ and $\sigma_\nu = 1.5$. In the heavy-tail case, $\nu = 1.5$, while $\nu = 2.0$ is used for normality. J denotes the winsorisation parameter.

heavy-tailed case and $\nu = 30$ for approximate normality.[6]

As expected, both bias and RMSE decline monotonically with the sample size $T$. The largest gains occur between $T = 1{,}000$ and $T = 2{,}000$, with further–often though relatively smaller–improvements between $T = 2{,}000$ and $T = 5{,}000$, particularly for $\nu$. Increasing the winsorisation parameter $J$ from 10 to 50 yields substantial reductions in RMSE for the scale parameters $\sigma_y$ and $\sigma_\nu$. Further increasing to $J = 100$ provides marginal improvements, and results beyond $J = 100$ show negligible gains, suggesting that $J \approx 100$ is a practical and efficient choice.

High-persistence designs ($\phi = 0.95$) exhibit slightly higher RMSEs for $\sigma_y$ and $\sigma_\nu$, although the differences are modest. In contrast, RMSE for $\phi$ tends to be lower in the high-persistence design, indicating more precise estimation of the persistence parameter. Across all parameters, the qualitative ranking across $J$ and the rate at which precision improves with $T$ remain stable. For $\nu$, the relative RMSE (i.e., RMSE/$\nu$) is lowest under strong heavy tails ($\nu = 3$) and highest in the near-Gaussian case ($\nu = 30$), reflecting the intrinsic difficulty of estimating a nearly normal distribution.

Appendix Table 7 presents results for intermediate values of $\nu$, specifically $\nu = 5$ and $\nu = 10$. These results reinforce the same general patterns: bias shrinks with $T$, RMSE decreases up to $J \approx 100$, and estimation error for $\nu$ increases as the distribution becomes closer to normal.

Table 2 reports results for the $\text{SV}_{\mathcal{GED}}$ model. The main findings from the Student-$t$ case largely carry over. In particular, RMSE for $\nu$ increases as the distribution approaches normality ($\nu = 2.0$), and declines for heavier-tailed cases ($\nu = 1.5$). However, the RMSE for $\nu$ is generally lower than in the Student-$t$ setting, despite not being normalized in the GED case. Appendix Table 8 provides further results for $\text{SV}_{\mathcal{GED}}$ models with stronger tail behavior ($\nu = 1.2$ and $\nu = 1.0$). These findings also align with the results presented here, confirming that RMSE for $\nu$ decreases as tail thickness increases.

Overall, the bias and RMSE we observe are quite small, suggesting that the proposed W-ARMA estimators for SV models with Student-$t$ and GED errors perform remarkably well–even at moderate sample sizes and relatively small values of $J$.

[6] Recall that the asymptotic distribution of $\hat{\nu}$ requires $\mathbb{E}[\log(u_t^2)^4] < \infty$, which holds for Student-$t$ innovations only if $\nu > 4$. For smaller $\nu$, such as $\nu = 3$, the fourth moment of $y_t^*$ is infinite, invalidating standard asymptotic theory. Nevertheless we present simulation results with $\nu = 3$ as it is used for hypothesis testing, motivates the use of Monte Carlo tests considered here, and is empirically relevant as seen in the Empirical section. Further, these simulation results showcase the estimation properties of our estimation procedure even in these settings where asymptotic theory isn't available.

## 6.2 Test performance

We now show empirical size and power for the asymptotic (Asy), Local Monte Carlo (LMC), and Maximized Monte Carlo (MMC) tests described above. Specifically, we consider again the same two persistence regimes used in Section 6.1, namely the moderate persistence ($\phi = 0.90$) and high persistence ($\phi = 0.95$), and focus on testing the null hypothesis of normality and heavy tails in SV models with Student-$t$ or GED error distributions. Here, we add additional sample sizes, where now $T \in \{500, 1000, 2000, 5000, 10000\}$. All tests are based on 1,000 replications, with winsorisation window $J = 100$, and variance parameters $\sigma_y = 1$ and $\sigma_\nu = 1.5$.

Table 3: Empirical size of Asymptotic and Monte Carlo LR-type tests

| $T$ | Test for normality | | | | | | Test for heavy tails | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi = 0.90$ | | | $\phi = 0.95$ | | | $\phi = 0.90$ | | | $\phi = 0.95$ | | |
| | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC |
| $SV_{\mathcal{GED}}$ | $H_0$: $\nu = 2.0$ vs. $H_1$: $\nu \neq 2.0$ | | | | | | $H_0$: $\nu = 1.5$ vs. $H_1$: $\nu \neq 1.5$ | | | | | |
| 500 | 6.1 | 6.8 | 3.7 | 3.2 | 5.8 | 2.0 | 9.4 | 4.4 | 2.1 | 5.9 | 4.5 | 2.0 |
| 1000 | 7.4 | 5.0 | 2.1 | 3.6 | 3.9 | 1.4 | 11.2 | 5.0 | 2.9 | 7.2 | 4.4 | 1.9 |
| 2000 | 8.5 | 5.0 | 1.4 | 6.0 | 4.3 | 1.3 | 9.5 | 5.2 | 2.3 | 6.4 | 4.6 | 1.4 |
| 5000 | 10.5 | 5.8 | 1.9 | 6.3 | 4.3 | 1.6 | 10.4 | 5.4 | 1.5 | 7.7 | 5.2 | 1.6 |
| 10000 | 11.2 | 6.3 | 1.7 | 8.2 | 4.5 | 1.7 | 10.1 | 4.7 | 1.3 | 7.1 | 4.3 | 1.3 |
| $SV_t$ | $H_0$: $\nu = 30$ vs. $H_1$: $\nu \neq 30$ | | | | | | $H_0$: $\nu = 10$ vs. $H_1$: $\nu \neq 10$ | | | | | |
| 500 | 8.4 | 5.5 | 1.6 | 2.3 | 4.3 | 0.9 | 5.8 | 6.6 | 1.4 | 1.0 | 4.2 | 0.9 |
| 1000 | 12.7 | 5.0 | 1.4 | 5.6 | 5.5 | 2.3 | 8.2 | 5.7 | 1.8 | 2.9 | 4.8 | 1.9 |
| 2000 | 12.2 | 6.6 | 2.1 | 5.4 | 4.7 | 2.0 | 9.0 | 6.0 | 1.8 | 5.2 | 6.7 | 2.8 |
| 5000 | 9.8 | 6.2 | 1.8 | 5.1 | 4.4 | 1.5 | 9.3 | 6.7 | 1.4 | 4.5 | 6.2 | 2.1 |
| 10000 | 7.5 | 4.2 | 0.7 | 4.9 | 5.0 | 1.2 | 7.2 | 6.0 | 0.9 | 2.4 | 5.3 | 0.3 |

Notes: The level of the test considered here is $\alpha = 5\%$. Empirical performance is computed over $1,000$ replications. Moderate (high) persistence corresponds to $\phi = 0.90$ ($\phi = 0.95$). The variance parameters are set to $\sigma_y = 1$ and $\sigma_\nu = 1.5$. Throughout, the winsorisation parameter is fixed at $J = 100$. The value of $\nu$ corresponds to the null hypothesis being tested: in the GED case (top panels), $\nu = 2.0$ when testing for normality (left) and $\nu = 1.5$ when testing for heavy tails (right); in the Student-$t$ case (bottom panels), $\nu = 30$ when testing for normality (left) and $\nu = 10$ when testing for heavy tails (right).

Table 3 reports empirical size for all three tests in the Student-$t$ (bottom) and GED (top) settings, while Table 4 reports empirical power. The null value of $\nu$ varies by test, as specified in Section 5. To summarize: when testing for normality in the GED error setting, we consider $H_0^{g,n}: \nu = 2.0$, as this is equivalent to normality, and consider alternatives of $\nu = 1.0$, $\nu = 1.2$, and $\nu = 1.5$ when evaluating power. When testing for normality in the Student-$t$ case, we consider a null hypothesis $H_0^{s,n}: \nu = 30$ to approximate normality and consider alternatives of $\nu = 3$, $\nu = 5$, and $\nu = 10$ for power analysis. When testing for heavy tails in the GED error setting, we test the null hypothesis $H_0^{g,h} : \nu = 1.5$ and consider the alternative $\nu = 2.0$ to evaluate power. Finally, when testing for heavy tails in the Student-$t$ case, we consider the null $H_0^{s,h} : \nu = 10$ and use $\nu = 30$ as the alternative to evaluate power against approximate normality. As mentioned in Section 5, this configuration does not appear to offer much power. For this reason, we also consider setting the null hypothesis as $H_0^{s,h,3} : \nu = 3$ and $H_0^{s,h,5} : \nu = 5$, and test against the alternative $\nu = 30$ in both cases to evaluate power when the null is further from normality. As described below, the case where $H_0^{s,h,3} : \nu = 5$ leads to relatively better power, but the strongest performance achieved when the null is $H_0^{s,h,3} : \nu = 3$.

**Size control.** The empirical size results in Table 3 suggest that the asymptotic LR test often exhibits notable size distortions, particularly under moderate persistence. In the Student-$t$ case, distortion is especially pronounced when the null distribution approximates normality, i.e., $H_0 : \nu = 30$. In contrast, distortions are generally smaller under the high persistence scenarios. The LMC procedure yields rejection rates close to the nominal 5% level across all designs, while the MMC test yields rejection frequencies that are at or below the nominal level, as expected from theory.

Table 9 provides additional results for Student-$t$ models with heavier tails under the null ($\nu_0 \in \{3, 5\}$). The observed patterns here are broadly consistent with those described above, although we do observe that for these heavier-tailed DGPs, the asymptotic test displays more pronounced over-rejection when persistence is high–though still less than in the moderate persistence cases.

Overall, we find that the LMC test performs best in terms of controlling the size of the test, and that the MMC test performs well in terms of maintaining the rejection frequencies at or below the level of the test, as should be expected. The asymptotic test, by contrast, often fails to maintain appropriate size, particularly under moderate persistence.

**Power patterns.** Power increases with sample size across all designs, with one key exception: the case of testing for heavy tails in the SV model with Student-$t$ errors, i.e., testing $H_0^{s,h} : \nu = 10$ against the alternative $\nu = 30$. For such moderate deviations, power remains low even at $T = 10{,}000$, underscoring the difficulty of distinguishing nearly normal distributions from those with moderately heavy tails.

Table 4: Empirical power of Asymptotic and Monte Carlo LR-type tests

| $T$ | Test for normality | | | | | | | | | | | | | | | | | | Test for heavy tails | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi = 0.90$ | | | $\phi = 0.95$ | | | $\phi = 0.90$ | | | $\phi = 0.95$ | | | $\phi = 0.90$ | | | $\phi = 0.95$ | | | $\phi = 0.90$ | | | $\phi = 0.95$ | | |
| | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC |
| $SV_{\mathcal{GED}}$ | $H_0$: $\nu = 2.0$ vs. $H_1$: $\nu \neq 2.0$ | | | | | | | | | | | | | | | | | | $H_0$: $\nu = 1.5$ vs. $H_1$: $\nu \neq 1.5$ | | | | | |
| | $\nu = 1.0$ | | | | | | $\nu = 1.2$ | | | | | | $\nu = 1.5$ | | | | | | $\nu = 2.0$ | | | | | |
| 500 | 66.8 | 64.8 | 53.2 | 59.2 | 65.1 | 50.1 | 41.3 | 41.5 | 30.7 | 34.4 | 40.7 | 26.0 | 14.1 | 15.6 | 10.0 | 11.4 | 15.4 | 7.4 | 8.8 | 9.0 | 4.4 | 6.0 | 6.9 | 4.1 |
| 1000 | 89.5 | 87.0 | 77.9 | 84.0 | 87.2 | 73.9 | 60.4 | 58.0 | 39.3 | 47.7 | 55.2 | 34.6 | 21.6 | 20.8 | 13.3 | 14.1 | 19.3 | 8.8 | 16.6 | 16.5 | 11.1 | 13.3 | 17.9 | 12.0 |
| 2000 | 99.1 | 99.1 | 96.4 | 98.4 | 98.7 | 93.2 | 80.3 | 78.3 | 63.0 | 69.1 | 75.9 | 52.4 | 27.6 | 26.8 | 14.5 | 17.7 | 25.5 | 9.8 | 28.6 | 28.0 | 18.7 | 24.4 | 29.7 | 17.9 |
| 5000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.2 | 98.9 | 94.2 | 98.3 | 98.9 | 92.6 | 58.4 | 56.0 | 29.2 | 44.1 | 49.7 | 26.3 | 58.2 | 57.6 | 40.6 | 57.1 | 62.4 | 41.2 |
| 10000 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 99.9 | 100.0 | 100.0 | 99.8 | 85.3 | 85.2 | 58.1 | 80.2 | 82.4 | 56.6 | 86.2 | 84.9 | 66.9 | 83.0 | 85.6 | 66.2 |
| $SV_t$ | $H_0$: $\nu = 30$ vs. $H_1$: $\nu \neq 30$ | | | | | | | | | | | | | | | | | | $H_0$: $\nu = 10$ vs. $H_1$: $\nu \neq 10$ | | | | | |
| | $\nu = 3$ | | | | | | $\nu = 5$ | | | | | | $\nu = 10$ | | | | | | $\nu = 30$ | | | | | |
| 500 | 16.2 | 16.5 | 5.0 | 3.9 | 13.8 | 3.6 | 10.7 | 10.6 | 3.4 | 2.4 | 8.3 | 3.2 | 6.0 | 7.4 | 2.4 | 0.8 | 5.7 | 1.8 | 4.0 | 4.0 | 0.8 | 0.9 | 3.4 | 0.9 |
| 1000 | 35.0 | 35.6 | 16.1 | 18.8 | 36.1 | 20.5 | 14.1 | 15.7 | 6.6 | 5.5 | 12.7 | 5.8 | 8.3 | 9.2 | 4.0 | 3.0 | 9.1 | 4.1 | 3.4 | 3.6 | 0.7 | 0.6 | 3.7 | 1.3 |
| 2000 | 62.7 | 61.0 | 40.3 | 48.7 | 62.6 | 47.1 | 29.0 | 28.3 | 14.0 | 12.1 | 24.2 | 13.7 | 9.4 | 10.8 | 4.1 | 4.1 | 12.2 | 5.5 | 2.8 | 3.9 | 1.4 | 1.3 | 3.0 | 1.0 |
| 5000 | 95.9 | 95.4 | 84.6 | 92.3 | 96.1 | 88.1 | 50.8 | 50.0 | 26.1 | 33.8 | 47.6 | 29.6 | 15.0 | 17.3 | 5.4 | 6.7 | 13.7 | 5.5 | 1.5 | 2.7 | 0.5 | 0.5 | 1.8 | 0.7 |
| 10000 | 100.0 | 99.9 | 99.2 | 99.8 | 100.0 | 99.6 | 77.1 | 76.0 | 47.8 | 63.6 | 75.3 | 53.6 | 19.3 | 20.2 | 4.9 | 12.3 | 20.5 | 8.0 | 0.4 | 0.4 | 0.0 | 0.7 | 2.0 | 0.1 |

Notes: The level of the test considered here is $\alpha = 5\%$. Empirical performance is computed over 1,000 replications. The true DGP is such that $\sigma_y^2 = 1$ and $\sigma_\nu^2 = 1.5$, $\phi = 0.90$ in the moderate persistence scenario and $\phi = 0.95$ is the high persistence scenario. Throughout, the winsorisation parameter is fixed at $J = 100$. The value of $\nu$ is specified above, following an alternative hypothesis (i.e. $\nu = \{1.0, 1.2, 1.5, 2.0\}$) for GED case or $\nu = \{3, 5, 10, 30\}$) in the Student-t case).

As noted earlier, Table 9 explores power in settings where the null hypothesis is instead $\nu = 5$ or $\nu = 3$. Power improves considerably when the null is set at $\nu = 3$, which corresponds to stronger tail behavior. When testing for heavy tails in the GED case, both the LMC and MMC tests perform well, with high power across all sample sizes and tail thickness levels.

It is worth highlighting again that the asymptotic test results shown here are locally level-corrected, and therefore not feasible in practice, as they require knowledge of the true DGP. Nevertheless, the fact that the LMC test closely mirrors the performance of this infeasible benchmark further validates its practical value.

In both the GED and Student-$t$ settings, power increases as the alternative hypothesis moves further from the null–for example, as $\nu \to 1.0$ in the GED case, or $\nu \to 3$ in the Student-$t$ case. In general, the Monte Carlo-based tests show strong power properties, especially at larger sample sizes. The only clear exception remains the Student-$t$ setting when testing $H_0^{s,h} : \nu = 10$ or $H_0^{s,h,5} : \nu = 5$ against $\nu = 30$, where moderate tail differences yield weak power.

**Practical guidance.** Among the three procedures, the LMC test offers the best trade-off between size accuracy and statistical power. It consistently outperforms the asymptotic test and is computationally simpler than the MMC. That said, the MMC test may be preferable in situations where identification is weak or the null lies on the boundary of the parameter space–scenarios in which both the asymptotic and LMC tests may underperform. Taken together, the results strongly support the use of simulation-based inference methods in SV models with non-Gaussian errors, particularly when standard asymptotic approximations are unreliable.

## 7 Applications to stock price volatilities

We now apply the proposed W-ARMA estimators for SV models with Student-$t$ and GED errors, along with the three hypothesis testing procedures previously developed, to daily returns from three major U.S. equity indices: the S&P 500, the Dow Jones Industrial Average (DOW Jones), and the NASDAQ Composite. The sample spans from 4 January 2000 to 31 May 2023, yielding 5,889 observations per series. Appendix Table 10 provides summary statistics for the raw demeaned returns ($y_t$), squared demeaned returns ($y_t^2$), and demeaned log-squared returns ($\log(y_t^2)$). The series exhibit large standard deviations, excess kurtosis, and significant autocorrelation in volatility, as evidenced by elevated Ljung–Box (LB) statistics on $\log(y_t^2)$. Appendix Figure 1 shows time series plots of $y_t$ for each index, clearly revealing large shocks, persistence in the shocks, and volatility clustering.

We estimate both $\text{SV}_t$ and $\text{SV}_{\mathcal{GED}}$ models using the W-ARMA estimator with a winsorisation parameter of $J = 100$. Table 5 presents the results: the top panel reports estimates from the $\text{SV}_{\mathcal{GED}}$ model, while the bottom panel reports those from the $\text{SV}_t$ model. All three indices exhibit highly persistent volatility processes, with $\hat{\phi} \approx 0.98$. The estimated degrees of freedom indicate strong evidence of heavy tails in all three assets under both specifications: $\hat{\nu} \in (3, 5)$ for $\text{SV}_t$ and $\hat{\nu} \in (1.2, 1.5)$ for $\text{SV}_{\mathcal{GED}}$. Across models and indices, $\hat{\sigma}_\nu \in (0.14, 0.19)$ suggests stable latent volatility noise. Values in parentheses are standard errors obtained via the implicit standard error procedure described in Section 6.3 of Ahsan et al. (2025a).

Table 5: Empirical W-ARMA estimates of $\text{SV}_{\mathcal{GED}}$ and $\text{SV}_t$ models

| | S&P 500 | | | | DOW Jones | | | | NASDAQ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\phi}$ | $\hat{\sigma}_y$ | $\hat{\sigma}_\nu$ | $\hat{\nu}$ | $\hat{\phi}$ | $\hat{\sigma}_y$ | $\hat{\sigma}_\nu$ | $\hat{\nu}$ | $\hat{\phi}$ | $\hat{\sigma}_y$ | $\hat{\sigma}_\nu$ | $\hat{\nu}$ |
| | | | | | | $SV_{\mathcal{GED}}$ | | | | | | |
| est. | 0.984 | 0.973 | 0.172 | 1.342 | 0.982 | 0.895 | 0.186 | 1.473 | 0.989 | 1.274 | 0.146 | 1.408 |
| SE | (0.002) | (0.070) | (0.018) | (0.086) | (0.002) | (0.064) | (0.014) | (0.097) | (0.001) | (0.100) | (0.016) | (0.091) |
| | | | | | | $SV_t$ | | | | | | |
| est. | 0.984 | 0.725 | 0.172 | 3.488 | 0.982 | 0.720 | 0.186 | 4.535 | 0.989 | 0.988 | 0.146 | 3.956 |
| SE | (0.002) | (0.085) | (0.020) | (0.423) | (0.002) | (0.046) | (0.017) | (0.713) | (0.001) | (0.074) | (0.014) | (0.562) |

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 (31) with $J = 100$. Standard errors of parameters are obtained using the implicit standard errors procedure described in Section 6.3 of Ahsan et al. (2025a). Top panel shows results from estimating an $\text{SV}_{\mathcal{GED}}$ model and bottom panel show results from using an $\text{SV}_t$ model.

To assess distributional assumptions, we test for normality and for heavy tails using the GMM-based LR-type tests previously discussed. Table 6 reports $p$-values for the asymptotic, LMC, and MMC versions of the test, with Monte Carlo sample sizes $N \in \{99, 299, 999\}$.

To test for normality, we set $H_0 : \nu = 2$ in the $\text{SV}_{\mathcal{GED}}$ case and $H_0 : \nu = 30$ in the $\text{SV}_t$ case. For all indices, all three tests reject the null of normality ($p < 0.05$) across both types of models.

To test for heavy tails, we use $H_0 : \nu = 1.5$ in the GED model and $H_0 : \nu = 3$ in the Student-$t$ model. These choices reflect null hypotheses of very heavy tails. As shown in Table 5, the estimates of $\nu$ are near these values, especially in the GED case. Accordingly, all three tests fail to reject the null of heavy tails in the $\text{SV}_{\mathcal{GED}}$ model across indices.

In contrast, for the $\text{SV}_t$ model, the findings are more nuanced. For the S&P 500, the asymptotic test rejects the null of heavy tails, whereas both LMC and MMC tests do not reject at conventional levels when $\hat{\nu} \approx 3.48$. For the other two indices, the estimated $\nu$ values are further from 3 ($\hat{\nu} \approx 4.54$

for DOWJ and ≈ 3.96 for NASDAQ), suggesting relatively lighter tails. Accordingly, both the asymptotic and LMC tests reject the null for these indices. However, the MMC test only rejects for the DOWJ; for NASDAQ, the null is not rejected. Taken together with the results from testing for normality, this suggests that the Dow Jones index exhibits heavy tails, though not to the same extent as the other two stock indices considered here.

Table 6: Asymptotic and Monte Carlo tests of error distribution in $SV_{\mathcal{GED}}$ and $SV_t$ models

| | Asy | LMC | | | MMC | | | Asy | LMC | | | MMC | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $N = 99$ | 299 | 999 | $N = 99$ | 299 | 999 | | $N = 99$ | 299 | 999 | $N = 99$ | 299 | 999 |
| Test for normality | | | | | | | | | | | | | | |
| | $SV_{\mathcal{GED}}$ - $H_0 : \nu = 2.0$ vs. $\nu \neq 2.0$ | | | | | | | $SV_t$ - $H_0 : \nu = 30$ vs. $\nu \neq 30$ | | | | | | |
| S&P 500 ($T = 5889$) | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.02 | 0.05 | 0.04 |
| Dow Jones ($T = 5889$) | 0.00 | 0.02 | 0.02 | 0.01 | 0.02 | 0.04 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.05 | 0.03 |
| NASDAQ ($T = 5889$) | 0.00 | 0.02 | 0.03 | 0.02 | 0.03 | 0.04 | 0.04 | 0.00 | 0.05 | 0.05 | 0.03 | 0.05 | 0.05 | 0.05 |
| Test for heavy tails | | | | | | | | | | | | | | |
| | $SV_{\mathcal{GED}}$ - $H_0 : \nu = 1.5$ vs. $\nu \neq 1.5$ | | | | | | | $SV_t$ - $H_0 : \nu = 3$ vs. $\nu \neq 3$ | | | | | | |
| S&P 500 ($T = 5889$) | 0.13 | 0.36 | 0.30 | 0.28 | 0.36 | 0.48 | 0.28 | 0.02 | 0.16 | 0.16 | 0.17 | 0.70 | 0.89 | 0.17 |
| Dow Jones ($T = 5889$) | 0.77 | 0.90 | 0.86 | 0.86 | 0.90 | 0.86 | 0.86 | 0.00 | 0.04 | 0.01 | 0.02 | 0.04 | 0.01 | 0.02 |
| NASDAQ ($T = 5889$) | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.05 | 0.03 | 0.03 | 0.17 | 0.16 | 0.17 |

Notes: 1. The reported values are p-values for test procedure when testing for normality. That is, $H_0 : \nu = 2.0$ vs. $H_1 : \nu \neq 2.0$ for $SV_{\mathcal{GED}}$ models (left side) and $H_0 : \nu = 30$ vs. $H_1 : \nu \neq 30$ for $SV_t$ models (right side). 2. The value $N$ is the number of Monte Carlo simulations used to simulate the null distribution, where multiple values are used to confirm the robustness of the results. 3. 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 (31) with $J = 100$.

Overall, these results support the use of heavy-tailed distributions when modeling financial return volatility. Heavy-tailed SV models provide a better fit than Gaussian specifications, with evidence of persistent volatility and non-Gaussian errors across these three major U.S. stock indices.

## 8 Conclusion

This paper proposes a new class of moment-based estimators for stochastic volatility (SV) models with conditionally heavy-tailed errors, specifically the Student's $t$ and GED. Building on ARMA-type moment representations, we derive simple, closed-form expressions for key model parameters–most notably the persistence and volatility parameters–without relying on numerical optimization or initial value selection. For the degrees-of-freedom parameter governing tail behavior, we introduce a profiling-based approach, enabling tractable inference in models that would otherwise require simulation-based maximum likelihood or Bayesian methods.

A key contribution is the development of winsorized ARMA estimators dealing with *heavy-tailed* error distributions, which significantly improve the robustness and stability of the estimation procedure in small samples and under heavy-tailed shocks. These estimators are easy to implement, computationally inexpensive, and highly efficient, making them particularly well-suited for empirical work involving large datasets or models that require repeated estimation.

The simplicity of the proposed estimators also opens the door to reliable hypothesis testing procedures. In particular, we construct likelihood-ratio-type (LR-type) tests for assessing the presence of normality or heavy tails in SV models with Student-$t$ or GED errors. We evaluate three versions of the LR-type test: the standard asymptotic approximation, the Local Monte Carlo (LMC) test, and the Maximized Monte Carlo (MMC) test. Simulation results indicate that the asymptotic test can suffer from size distortions, particularly in moderately persistent models, while the LMC test performs well in terms of both size control and power. The MMC test provides an identification-robust procedure that is also valid in finite-samples. An important direction for future research involves developing and evaluating one-sided tests, such as in the Student-$t$ setting where values of $\nu > 30$ can be regarded as consistent with normality. Addressing this requires adapting the proposed test statistic to a one-sided framework, which is the focus of ongoing research.

Our extensive simulation study confirms the strong performance of the W-ARMA estimators in in terms of providing precise estimates across a wide range of settings. Bias and RMSE are consistently low–even when the model features heavy-tailed innovations–and performance improves with sample size. The LR-type tests also perform well, with the LMC and MMC tests providing robust alternatives when the standard asymptotic procedure fails to control size.

Finally, we apply our methods to daily return data for three major U.S. equity indices: the S&P 500, Dow Jones Industrial Average, and NASDAQ Composite. The estimation results provide strong evidence of heavy tails in all series, with degrees-of-freedom estimates ranging from 3 to 5 for the Student-$t$ models and 1.2 to 1.5 for the GED models. The LR-type tests decisively reject the hypothesis of normality and, in some cases, reject the presence of very heavy tails in favor of moderately heavy tails. These findings underscore the relevance of heavy-tailed SV models for financial time series and confirm the practical value of the proposed methods.

Overall, this paper offers new tools for efficient estimation and reliable inference in SV models with heavy-tailed errors. The proposed methods provide both theoretical and practical advantages, particularly for applied researchers seeking computationally efficient yet statistically valid tools for inference and testing in high-frequency or heavy-tailed financial data.

# References


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

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.

Ahsan, M. N. and Dufour, J.-M. (2025). Practical estimation methods for high-dimensional multivariate stochastic volatility models. Working Paper 5081221, Social Science Research Network.

Ahsan, M. N., Dufour, J.-M., and Rodriguez-Rondon, G. (2025a). Estimation and inference for higher-order stochastic volatility models with leverage. *Journal of Time Series Analysis*.

Ahsan, M. N., Dufour, J.-M., and Rodriguez-Rondon, G. (2025b). Volatility forecasting with higher-order stochastic volatility models. Technical report, McGill University.

Andersen, T. G., Chung, H.-J., and Sørensen, B. E. (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of econometrics*, 91(1):61–87.

Andersen, T. G. and Sørensen, B. E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics*, 14(3):328–352.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144.

Breidt, F. J. and Carriquiry, A. L. (1996). Improved quasi-maximum likelihood estimation for stochastic volatility models. In Johnson, W. O., Lee, J. C., and Zellner, A., editors, *Modelling and Prediction: Honouring Seymour Geisel*, pages 228–247. Springer-Verlag, New York.

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

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

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

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

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., 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.

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

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

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

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

Flury, T. and Shephard, N. (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory*, 27(05):933–956.

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

Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric theory*, 12(4):657–681.

Ghysels, E., Harvey, A., and Renault, E. (1996). Stochastic volatility. In Maddala, G. S. and Rao, C. R., editors, *Handbook of Statistics: Statistical Methods in Finance*, volume 14, pages 119–191. North-Holland, Amsterdam.

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

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.

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.

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.

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

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.

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

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

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

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

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(3):777–787.

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.

Rice, J. A. and Rice, J. A. (2007). *Mathematical statistics and data analysis*, volume 371. Thomson/Brooks/Cole Belmont, CA.

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

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

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

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

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

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

Subbotin, M. T. (1923). On the law of frequency of error. *Matematicheskii Sbornik*, 31(2):296–301. In Russian.

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

Wishart, J. (1947). The cumulants of the $z$ and of the logarithmic $\chi^2$ and $t$ distributions. *Biometrika*, 34(1/2):170–178.

# 9 Appendix

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

1. Mathematical Proofs.
2. Asymptotic Tests
3. Complementary simulation results.
4. Complementary empirical results.

### 9.1 Mathematical proofs

**Proof 9.1** *Proof of Lemma 3.1. The cumulant generating function of* $\log(\chi^2_\nu/\nu)$ *distribution is:*

$$\begin{aligned} M(s) &= \log \mathbb{E}\big[\exp\big(s\log(\chi^2_\nu/\nu)\big)\big] = \log\left[\mathbb{E}\left(\chi^2_\nu/\nu\right)^s\right] = \log\left[\frac{2^s\,\Gamma\big((\nu/2)+s\big)}{\nu^s\Gamma(\nu/2)}\right] \\ &= s\log(2/\nu) + \log[\Gamma\big((\nu/2)+s\big)] - \log[\Gamma(\nu/2)]\,, \quad \textit{for } s \geq 0\,, \end{aligned} \tag{60}$$

*where* $\Gamma(z) := \int_0^\infty x^{z-1}e^{-x}\,dx$ *is the* gamma function*; see Wishart (1947). The* $m^{th}$ *cumulant of the* $\log(\chi^2_\nu/\nu)$ *random variable is the* $m^{th}$ *derivative of* $M(s)$ *evaluated at* $s=0$*. Thus, the corresponding cumulants* ($\kappa_m$) *and central moments* ($\tilde{\mu}_m$) *are:*

$$\kappa_m = \begin{cases} \psi\big(\frac{\nu}{2}\big) - \log(\nu/2), & \textit{if } m = 1 \\ \psi^{(m-1)}\big(\frac{\nu}{2}\big), & \textit{if } m > 1 \end{cases},$$

$$\tilde{\mu}_m = \begin{cases} 0, & \textit{if } m = 1 \\ \kappa_m + \sum_{j=1}^{m-2}\binom{m-1}{j}\kappa_{m-j}\tilde{\mu}_j, & \textit{if } m > 1 \end{cases}, \tag{61}$$

*where*

$$\psi(z) := \frac{d}{dz}\log\left[\Gamma(z)\right] = \frac{\Gamma'(z)}{\Gamma(z)} \tag{62}$$

*is the* digamma function *and*

$$\psi^{(m)}(z) := \frac{d^m}{dz^m}\psi(z) = \frac{d^{m+1}}{dz^{m+1}}\log[\Gamma(z)] \tag{63}$$

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

**Proof 9.2** *Proof of Lemma 3.2. Since* $\bar{u}$ *is symmetric about zero, we have* $\mathbb{E}\big[\bar{u}^k\big] = 0$ *for* $k$ *is odd. When* $k$ *is even,* $k = 2m$ *say,*

$$\mathbb{E}\big[\bar{u}^{2m}\big] = \left[\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^m \left[\frac{\Gamma\big(\frac{2m}{\nu}+\frac{1}{\nu}\big)}{\Gamma(1/\nu)}\right],\ m \geq 1. \tag{64}$$

*The cumulant generating function of* $\log(\bar{u}_t^2)$ *distribution is:*

$$\begin{aligned} M(s) &= \log \mathbb{E}\big[\exp\big(s\log(\bar{u}_t^2)\big)\big] = \log\Big[\mathbb{E}\left(\bar{u}_t\right)^{2s}\Big] = \log\left(\left[\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^s \left[\frac{\Gamma\big(\frac{2s}{\nu}+\frac{1}{\nu}\big)}{\Gamma(1/\nu)}\right]\right) \\ &= s\log\Gamma(1/\nu) - s\log\Gamma(3/\nu) + \log\left[\Gamma\left(\frac{2s}{\nu}+\frac{1}{\nu}\right)\right] - \log[\Gamma(1/\nu)]\,, \quad \textit{for } s \geq 0\,, \end{aligned} \tag{65}$$

*where* $\Gamma(z)$ *is the* gamma function. *The* $m^{th}$ *cumulant of the* log *squared* $\mathcal{GED}$ *random variable is the* $m^{th}$ *derivative of* $M(s)$ *evaluated at* $s = 0$. *Thus, the corresponding cumulants* ($\kappa_m$) *and central moments* ($\tilde{\mu}_m$) *are:*

$$\begin{aligned} \kappa_m &= \begin{cases} \frac{2}{\nu}\psi\big(\frac{1}{\nu}\big) + \log\Big[\Gamma\big(\frac{1}{\nu}\big)\Big] - \log\Big[\Gamma\big(\frac{3}{\nu}\big)\Big], & \textit{if } m = 1 \\ \big(\frac{2}{\nu}\big)^m \psi^{(m-1)}\big(\frac{1}{\nu}\big), & \textit{if } m > 1 \end{cases}, \\ \tilde{\mu}_m &= \begin{cases} 0, & \textit{if } m = 1 \\ \kappa_m + \sum_{j=1}^{m-2}\binom{m-1}{j}\kappa_{m-j}\tilde{\mu}_j, & \textit{if } m > 1 \end{cases}, \end{aligned} \tag{66}$$

*where* $\psi(z)$ *and* $\psi^{(m)}(z)$ *are the digamma and polygamma function of order* $m$.

**Proof 9.3** *Proof of Lemma 4.1.*

*Define the sample moment function:*

$$f_T^{St}(\nu) := \psi^{(1)}\left(\frac{1}{2}\right) + \psi^{(1)}\left(\frac{\nu}{2}\right) - \hat{\gamma}_{y^*}(0) + \frac{\hat{\gamma}_{y^*}(1)}{\hat{\phi}}.$$

*Under Assumptions 2.1 and 3.1, the process* $y_t^*$ *is strictly stationary and ergodic with finite fourth moments. It follows by Lemma 3 and Theorem 1 of Ahsan et al. (2025a) that:*

$$\hat{\gamma}_{y^*}(h) \xrightarrow{p} \gamma_{y^*}(h), \quad \textit{for } h = 0, 1, \quad \textit{and} \quad \hat{\phi} \xrightarrow{p} \phi.$$

*Therefore,*

$$f_T^{St}(\nu) \xrightarrow{p} f^{St}(\nu) \quad \textit{pointwise in } \nu.$$

*Next, note that* $f_T^{St}(\nu)$ *is continuous in* $\nu$ *for each* $T$, *and* $f^{St}(\nu)$ *is continuous with a unique root at* $\nu = \nu_0$. *Moreover, over any compact subinterval* $\mathcal{N}_0 \subset \mathcal{N}$, *the convergence is uniform by the stochastic equicontinuity of* $f_T^{St}(\nu)$, *which follows from the continuity of the polygamma function* *and the uniform convergence of sample moments.*

*By the standard theory of Z-estimators (see, e.g., Newey and McFadden, 1994), it follows that*

$$\hat{\nu} := \arg\min_{\nu \in \mathcal{N}} \left| f_T^{St}(\nu) \right| \xrightarrow{p} \nu_0.$$

**Proof 9.4** *Proof of Theorem 4.1.*

*Let $\nu_0 \in \mathcal{N} \subset (4, \infty)$ denote the true value of the tail parameter. Define the sample moment function:*

$$f_T^{St}(\nu) := \psi^{(1)}\left(\tfrac{1}{2}\right) + \psi^{(1)}\left(\tfrac{\nu}{2}\right) - \hat{\gamma}_{y^*}(0) + \frac{1}{\hat{\phi}}\hat{\gamma}_{y^*}(1),$$

*which satisfies $f_T^{St}(\hat{\nu}) = 0$ by definition of $\hat{\nu}$. The corresponding population moment function is*

$$f^{St}(\nu) := \psi^{(1)}\left(\tfrac{1}{2}\right) + \psi^{(1)}\left(\tfrac{\nu}{2}\right) - \gamma_{y^*}(0) + \frac{1}{\phi}\gamma_{y^*}(1),$$

*with a unique root at $\nu = \nu_0$ by Assumption 4.1.*

*Under Assumptions 2.1 and 3.1, the process $y_t^*$ is strictly stationary and $\beta$-mixing with finite fourth moments. As a result, by Lemma 4 of Ahsan et al. (2025a), the sample autocovariances satisfy:*

$$\sqrt{T}\begin{pmatrix} \hat{\gamma}_{y^*}(0) - \gamma_{y^*}(0) \\ \hat{\gamma}_{y^*}(1) - \gamma_{y^*}(1) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

*where $\Sigma$ is the long-run variance-covariance matrix of the vector $(\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1))$.*

*By the continuous mapping theorem,*

$$\sqrt{T}\left[f_T^{St}(\nu_0) - f^{St}(\nu_0)\right] = \sqrt{T}\left[-(\hat{\gamma}_{y^*}(0) - \gamma_{y^*}(0)) + \frac{1}{\phi}(\hat{\gamma}_{y^*}(1) - \gamma_{y^*}(1))\right] \xrightarrow{d} \mathcal{N}(0, \sigma_f^2),$$

*with*

$$\sigma_f^2 := \mathrm{Var}\left(\hat{\gamma}_{y^*}(0) - \frac{1}{\phi}\hat{\gamma}_{y^*}(1)\right) = \mathbf{c}^\top \Sigma \mathbf{c}, \quad \textit{where } \mathbf{c} = \begin{pmatrix} 1 \\ -1/\phi \end{pmatrix}.$$

*Since $f_T^{St}(\hat{\nu}) = 0$, a first-order Taylor expansion around $\nu_0$ yields:*

$$0 = f_T^{St}(\hat{\nu}) = f_T^{St}(\nu_0) + (\hat{\nu} - \nu_0) \cdot \left[\left.\frac{\partial f_T^{St}(\nu)}{\partial \nu}\right|_{\nu = \bar{\nu}}\right] + o_p(T^{-1/2}),$$

*for some* $\bar{\nu} \in (\nu_0, \hat{\nu})$. *Solving for* $\hat{\nu} - \nu_0$, *we obtain:*

$$\sqrt{T}(\hat{\nu} - \nu_0) = -\left[\frac{\partial f_T^{St}(\nu)}{\partial \nu}\bigg|_{\nu=\bar{\nu}}\right]^{-1} \cdot \sqrt{T} f_T^{St}(\nu_0) + o_p(1).$$

*By Assumptions 4.1–4.2, the derivative*

$$\frac{\partial f_T^{St}(\nu)}{\partial \nu} \xrightarrow{p} \frac{\partial f_T^{St}(\nu)}{\partial \nu}\bigg|_{\nu=\nu_0} = \frac{1}{2} \cdot \psi^{(2)}\left(\frac{\nu_0}{2}\right) \neq 0.$$

*By Lemma 4.1,* $\hat{\nu} \xrightarrow{p} \nu_0$, *so* $\bar{\nu} \xrightarrow{p} \nu_0$ *as well. Hence, by Slutsky's theorem,*

$$\sqrt{T}(\hat{\nu} - \nu_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_f^2}{\left[\frac{1}{2} \cdot \psi^{(2)}\left(\frac{\nu_0}{2}\right)\right]^2}\right).$$

**Proof 9.5** *Proof of Lemma 4.2.*

*The argument mirrors that of Lemma 4.1. Under Assumptions 2.1 and 3.2, the process* $y_t^*$ *is strictly stationary and ergodic with finite moments of all orders. By Lemma 3 and Theorem 1 of Ahsan et al. (2025a), we have:*

$$\hat{\gamma}_{y^*}(h) \xrightarrow{p} \gamma_{y^*}(h), \quad \text{for } h = 0, 1, \quad \text{and} \quad \hat{\phi} \xrightarrow{p} \phi.$$

*Thus, the sample moment function:*

$$f_T^{\mathcal{GED}}(\nu) := \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \hat{\gamma}_{y^*}(0) + \frac{\hat{\gamma}_{y^*}(1)}{\hat{\phi}}$$

*converges pointwise in probability to:*

$$f^{\mathcal{GED}}(\nu) := \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \gamma_{y^*}(0) + \frac{\gamma_{y^*}(1)}{\phi}.$$

*Continuity of* $f_T^{\mathcal{GED}}(\nu)$ *and* $f^{\mathcal{GED}}(\nu)$ *over* $\mathcal{N}$, *together with the uniqueness of the root* $\nu_0$, *implies that* $\hat{\nu} \xrightarrow{p} \nu_0$ *by standard Z-estimator arguments (e.g., Newey and McFadden, 1994).*

**Proof 9.6** *Proof of Theorem 4.2.*

*Let* $\nu_0 \in \mathcal{N} \subset (0, \infty)$ *denote the true tail parameter. Define the sample moment function:*

$$f_T^{\mathcal{GED}}(\nu) := \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \hat{\gamma}_{y^*}(0) + \frac{1}{\hat{\phi}}\hat{\gamma}_{y^*}(1),$$

*which satisfies* $f_T^{\mathcal{GED}}(\hat{\nu}) = 0$ *by definition of* $\hat{\nu}$*. The corresponding population moment function is*

$$f^{\mathcal{GED}}(\nu) := \left(\frac{2}{\nu}\right)^2 \psi^{(1)}\left(\frac{1}{\nu}\right) - \gamma_{y^*}(0) + \frac{1}{\phi}\gamma_{y^*}(1),$$

*with a unique root at* $\nu = \nu_0$ *by Assumption 4.3.*

*Under Assumptions 2.1 and 3.2, the process* $y_t^*$ *is strictly stationary and* $\beta$*-mixing with finite fourth moments (since all moments exist for the GED distribution). As a result, by Lemma 4 of Ahsan et al. (2025a), the sample autocovariances satisfy:*

$$\sqrt{T}\begin{pmatrix}\hat{\gamma}_{y^*}(0) - \gamma_{y^*}(0) \\ \hat{\gamma}_{y^*}(1) - \gamma_{y^*}(1)\end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

*where* $\Sigma$ *is the long-run variance-covariance matrix of* $(\hat{\gamma}_{y^*}(0), \hat{\gamma}_{y^*}(1))$*.*

*It then follows from the continuous mapping theorem that:*

$$\sqrt{T}\left[f_T^{\mathcal{GED}}(\nu_0) - f^{\mathcal{GED}}(\nu_0)\right] = \sqrt{T}\left[-(\hat{\gamma}_{y^*}(0) - \gamma_{y^*}(0)) + \frac{1}{\phi}(\hat{\gamma}_{y^*}(1) - \gamma_{y^*}(1))\right] \xrightarrow{d} \mathcal{N}(0, \sigma_f^2),$$

*with*

$$\sigma_f^2 := \mathrm{Var}\left(\hat{\gamma}_{y^*}(0) - \frac{1}{\phi}\hat{\gamma}_{y^*}(1)\right) = \mathbf{c}^\top \Sigma \mathbf{c}, \quad \textit{where } \mathbf{c} = \begin{pmatrix} 1 \\ -1/\phi \end{pmatrix}.$$

*Since* $f_T^{\mathcal{GED}}(\hat{\nu}) = 0$*, a first-order Taylor expansion around* $\nu_0$ *gives:*

$$0 = f_T^{\mathcal{GED}}(\hat{\nu}) = f_T^{\mathcal{GED}}(\nu_0) + (\hat{\nu} - \nu_0) \cdot \left[\left.\frac{\partial f_T^{\mathcal{GED}}(\nu)}{\partial \nu}\right|_{\nu=\bar{\nu}}\right] + o_p(T^{-1/2}),$$

*for some* $\bar{\nu} \in (\nu_0, \hat{\nu})$*. Solving for* $\hat{\nu} - \nu_0$*, we obtain:*

$$\sqrt{T}(\hat{\nu} - \nu_0) = -\left[\left.\frac{\partial f_T^{\mathcal{GED}}(\nu)}{\partial \nu}\right|_{\nu=\bar{\nu}}\right]^{-1} \cdot \sqrt{T} f_T^{\mathcal{GED}}(\nu_0) + o_p(1).$$

*By Assumptions 4.3–4.4, and by continuity of* $f_T^{\mathcal{GED}}(\nu)$*, we have:*

$$\frac{\partial f_T^{\mathcal{GED}}(\nu)}{\partial \nu} \xrightarrow{p} \left.\frac{\partial f^{\mathcal{GED}}(\nu)}{\partial \nu}\right|_{\nu=\nu_0} = -\frac{8}{\nu_0^3} \cdot \psi^{(1)}\left(\frac{1}{\nu_0}\right) - \frac{4}{\nu_0^4} \cdot \psi^{(2)}\left(\frac{1}{\nu_0}\right) \neq 0.$$

*Finally, by Slutsky's theorem:*

$$\sqrt{T}(\hat{\nu} - \nu_0) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma_f^2}{\left[-\frac{8}{\nu_0^3} \cdot \psi^{(1)}\left(\frac{1}{\nu_0}\right) - \frac{4}{\nu_0^4} \cdot \psi^{(2)}\left(\frac{1}{\nu_0}\right)\right]^2}\right).$$

## 9.2 Complementary simulation results

Table 7: Estimation of SV models with Student-$t$ errors and intermediate heavy-tails

| $T$ | | | Moderate persistence | | | | High persistence | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ |
| | Heavy Tails (Fin. app.) | | 0.90 | 1.00 | 1.50 | 5.00 | 0.95 | 1.00 | 1.50 | 5.00 |
| | W-ARMA ($J = 10$) | Bias | -0.005 | -0.214 | -0.001 | 5.011 | -0.005 | -0.142 | 0.000 | 4.630 |
| | | RMSE | 0.021 | 0.293 | 0.120 | 5.347 | 0.014 | 0.454 | 0.113 | 5.008 |
| 1000 | W-ARMA ($J = 50$) | Bias | -0.001 | -0.216 | -0.027 | 4.286 | -0.004 | -0.141 | -0.009 | 4.573 |
| | | RMSE | 0.021 | 0.295 | 0.132 | 4.961 | 0.014 | 0.455 | 0.141 | 5.067 |
| | W-ARMA ($J = 100$) | Bias | 0.005 | -0.221 | -0.075 | 3.220 | -0.000 | -0.148 | -0.067 | 3.256 |
| | | RMSE | 0.022 | 0.298 | 0.155 | 4.489 | 0.013 | 0.454 | 0.161 | 4.136 |
| | W-ARMA ($J = 200$) | Bias | 0.014 | -0.231 | -0.152 | 1.631 | 0.006 | -0.162 | -0.168 | 1.387 |
| | | RMSE | 0.025 | 0.303 | 0.208 | 3.349 | 0.014 | 0.452 | 0.229 | 3.193 |
| | W-ARMA ($J = 10$) | Bias | -0.002 | -0.219 | 0.001 | 4.463 | -0.002 | -0.187 | 0.001 | 4.281 |
| | | RMSE | 0.014 | 0.262 | 0.086 | 4.898 | 0.009 | 0.340 | 0.081 | 4.567 |
| 2000 | W-ARMA ($J = 50$) | Bias | -0.001 | -0.222 | -0.014 | 3.826 | -0.002 | -0.187 | -0.003 | 4.212 |
| | | RMSE | 0.015 | 0.264 | 0.096 | 4.146 | 0.010 | 0.340 | 0.106 | 4.411 |
| | W-ARMA ($J = 100$) | Bias | 0.003 | -0.225 | -0.041 | 2.984 | -0.000 | -0.190 | -0.034 | 3.360 |
| | | RMSE | 0.015 | 0.266 | 0.106 | 3.984 | 0.010 | 0.342 | 0.114 | 3.758 |
| | W-ARMA ($J = 200$) | Bias | 0.009 | -0.234 | -0.090 | 1.980 | 0.004 | -0.202 | -0.094 | 1.983 |
| | | RMSE | 0.017 | 0.273 | 0.137 | 3.623 | 0.010 | 0.345 | 0.149 | 3.429 |
| | W-ARMA ($J = 10$) | Bias | -0.001 | -0.219 | 0.002 | 2.520 | -0.001 | -0.205 | 0.001 | 2.300 |
| | | RMSE | 0.009 | 0.238 | 0.056 | 3.196 | 0.006 | 0.270 | 0.051 | 2.652 |
| 5000 | W-ARMA ($J = 50$) | Bias | -0.001 | -0.220 | -0.004 | 2.206 | -0.001 | -0.205 | 0.001 | 2.739 |
| | | RMSE | 0.010 | 0.239 | 0.063 | 2.477 | 0.006 | 0.270 | 0.071 | 3.108 |
| | W-ARMA ($J = 100$) | Bias | 0.001 | -0.223 | -0.016 | 1.918 | -0.000 | -0.208 | -0.012 | 2.256 |
| | | RMSE | 0.010 | 0.241 | 0.066 | 2.664 | 0.006 | 0.272 | 0.073 | 3.001 |
| | W-ARMA ($J = 200$) | Bias | 0.004 | -0.228 | -0.040 | 1.155 | 0.002 | -0.213 | -0.039 | 1.453 |
| | | RMSE | 0.011 | 0.246 | 0.077 | 2.029 | 0.006 | 0.275 | 0.085 | 2.435 |
| | Light-Tails/Near Normal | | 0.90 | 1.00 | 1.50 | 10.00 | 0.95 | 1.00 | 1.50 | 10.00 |
| | W-ARMA ($J = 10$) | Bias | -0.003 | -0.129 | -0.017 | 3.431 | -0.004 | -0.051 | -0.012 | 3.699 |
| | | RMSE | 0.020 | 0.254 | 0.118 | 3.219 | 0.013 | 0.475 | 0.111 | 3.174 |
| 1000 | W-ARMA ($J = 50$) | Bias | 0.001 | -0.130 | -0.045 | 2.893 | -0.003 | -0.052 | -0.027 | 3.103 |
| | | RMSE | 0.021 | 0.254 | 0.134 | 2.972 | 0.014 | 0.474 | 0.142 | 2.916 |
| | W-ARMA ($J = 100$) | Bias | 0.007 | -0.133 | -0.091 | 1.720 | 0.001 | -0.056 | -0.082 | 1.706 |
| | | RMSE | 0.022 | 0.256 | 0.161 | 2.887 | 0.013 | 0.473 | 0.166 | 2.675 |
| | W-ARMA ($J = 200$) | Bias | 0.016 | -0.140 | -0.168 | -0.331 | 0.007 | -0.064 | -0.182 | 0.295 |
| | | RMSE | 0.026 | 0.259 | 0.218 | 2.278 | 0.014 | 0.477 | 0.237 | 2.727 |
| | W-ARMA ($J = 10$) | Bias | -0.001 | -0.123 | -0.014 | 5.849 | -0.002 | -0.083 | -0.010 | 5.836 |
| | | RMSE | 0.014 | 0.199 | 0.085 | 3.629 | 0.009 | 0.327 | 0.079 | 3.485 |
| 2000 | W-ARMA ($J = 50$) | Bias | 0.001 | -0.125 | -0.030 | 4.993 | -0.001 | -0.086 | -0.022 | 5.367 |
| | | RMSE | 0.015 | 0.199 | 0.097 | 3.260 | 0.010 | 0.326 | 0.106 | 3.368 |
| | W-ARMA ($J = 100$) | Bias | 0.005 | -0.128 | -0.056 | 3.622 | 0.001 | -0.088 | -0.050 | 4.192 |
| | | RMSE | 0.015 | 0.201 | 0.112 | 3.125 | 0.009 | 0.326 | 0.118 | 3.246 |
| | W-ARMA ($J = 200$) | Bias | 0.010 | -0.134 | -0.104 | 1.839 | 0.005 | -0.097 | -0.108 | 2.023 |
| | | RMSE | 0.018 | 0.205 | 0.145 | 2.822 | 0.010 | 0.327 | 0.157 | 2.903 |
| | W-ARMA ($J = 10$) | Bias | -0.000 | -0.113 | -0.006 | 7.593 | -0.001 | -0.097 | -0.005 | 7.907 |
| | | RMSE | 0.009 | 0.152 | 0.054 | 3.635 | 0.005 | 0.219 | 0.050 | 3.706 |
| 5000 | W-ARMA ($J = 50$) | Bias | 0.001 | -0.114 | -0.013 | 7.229 | -0.000 | -0.097 | -0.010 | 7.944 |
| | | RMSE | 0.009 | 0.153 | 0.062 | 3.651 | 0.006 | 0.220 | 0.070 | 3.685 |
| | W-ARMA ($J = 100$) | Bias | 0.002 | -0.116 | -0.025 | 6.103 | 0.001 | -0.099 | -0.022 | 6.397 |
| | | RMSE | 0.010 | 0.155 | 0.067 | 3.419 | 0.006 | 0.220 | 0.074 | 3.270 |
| | W-ARMA ($J = 200$) | Bias | 0.005 | -0.120 | -0.048 | 4.761 | 0.002 | -0.104 | -0.048 | 4.677 |
| | | RMSE | 0.011 | 0.158 | 0.080 | 3.226 | 0.007 | 0.222 | 0.088 | 3.083 |

Notes: Bias and RMSE are computed over 10,000 Monte Carlo replications. Moderate (high) persistence corresponds to $\phi = 0.90$ ($\phi = 0.95$). Variance parameters are set to $\sigma_y = 1$ and $\sigma_\nu = 1.5$. The intermediate heavy-tail cases correspond to, $\nu = 5.0$ (top) and $\nu = 10.0$ (bottom). RMSE for $\nu$ is reported as RMSE/$\nu$. $J$ denotes the winsorisation parameter.

Table 8: Estimation of SV models with Student-$\mathcal{GED}$ errors and heavier-tails

| $T$ | | | Moderate persistence | | | | High persistence | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ | $\phi$ | $\sigma_y$ | $\sigma_\nu$ | $\nu$ |
| | Heavier tails | | 0.90 | 1.00 | 1.50 | 1.20 | 0.95 | 1.00 | 1.50 | 1.20 |
| 1000 | W-ARMA ($J = 10$) | Bias | -0.007 | 0.024 | 0.020 | 0.081 | -0.005 | 0.114 | 0.016 | 0.067 |
| | | RMSE | 0.023 | 0.264 | 0.134 | 0.328 | 0.015 | 0.577 | 0.121 | 0.290 |
| | W-ARMA ($J = 50$) | Bias | -0.004 | 0.031 | -0.006 | 0.058 | -0.005 | 0.114 | 0.012 | 0.069 |
| | | RMSE | 0.022 | 0.266 | 0.138 | 0.302 | 0.015 | 0.578 | 0.147 | 0.303 |
| | W-ARMA ($J = 100$) | Bias | 0.002 | 0.045 | -0.054 | 0.018 | -0.001 | 0.133 | -0.049 | 0.022 |
| | | RMSE | 0.022 | 0.271 | 0.153 | 0.263 | 0.014 | 0.592 | 0.160 | 0.266 |
| | W-ARMA ($J = 200$) | Bias | 0.011 | 0.066 | -0.131 | -0.037 | 0.006 | 0.163 | -0.156 | -0.048 |
| | | RMSE | 0.025 | 0.280 | 0.198 | 0.229 | 0.014 | 0.613 | 0.224 | 0.228 |
| 2000 | W-ARMA ($J = 10$) | Bias | -0.003 | 0.012 | 0.009 | 0.034 | -0.003 | 0.055 | 0.007 | 0.031 |
| | | RMSE | 0.015 | 0.184 | 0.093 | 0.188 | 0.009 | 0.376 | 0.085 | 0.179 |
| | W-ARMA ($J = 50$) | Bias | -0.002 | 0.016 | -0.003 | 0.026 | -0.003 | 0.055 | 0.008 | 0.035 |
| | | RMSE | 0.015 | 0.185 | 0.100 | 0.184 | 0.010 | 0.376 | 0.110 | 0.187 |
| | W-ARMA ($J = 100$) | Bias | 0.002 | 0.023 | -0.030 | 0.006 | -0.001 | 0.064 | -0.023 | 0.013 |
| | | RMSE | 0.015 | 0.187 | 0.107 | 0.175 | 0.010 | 0.380 | 0.115 | 0.177 |
| | W-ARMA ($J = 200$) | Bias | 0.007 | 0.037 | -0.078 | -0.027 | 0.003 | 0.081 | -0.085 | -0.027 |
| | | RMSE | 0.017 | 0.191 | 0.132 | 0.164 | 0.010 | 0.389 | 0.146 | 0.165 |
| 5000 | W-ARMA ($J = 10$) | Bias | -0.001 | 0.006 | 0.003 | 0.014 | -0.001 | 0.024 | 0.002 | 0.015 |
| | | RMSE | 0.009 | 0.116 | 0.060 | 0.111 | 0.006 | 0.227 | 0.054 | 0.106 |
| | W-ARMA ($J = 50$) | Bias | -0.001 | 0.007 | -0.001 | 0.011 | -0.001 | 0.023 | 0.003 | 0.017 |
| | | RMSE | 0.010 | 0.116 | 0.065 | 0.111 | 0.007 | 0.228 | 0.072 | 0.111 |
| | W-ARMA ($J = 100$) | Bias | 0.001 | 0.011 | -0.013 | 0.003 | -0.000 | 0.027 | -0.009 | 0.008 |
| | | RMSE | 0.010 | 0.117 | 0.067 | 0.108 | 0.006 | 0.229 | 0.074 | 0.108 |
| | W-ARMA ($J = 200$) | Bias | 0.003 | 0.017 | -0.035 | -0.013 | 0.001 | 0.034 | -0.036 | -0.009 |
| | | RMSE | 0.011 | 0.118 | 0.076 | 0.105 | 0.007 | 0.231 | 0.085 | 0.105 |
| | Laplace (double-exp.) | | 0.90 | 1.00 | 1.50 | 1.00 | 0.95 | 1.00 | 1.50 | 1.00 |
| 1000 | W-ARMA ($J = 10$) | Bias | -0.008 | 0.024 | 0.024 | 0.048 | -0.006 | 0.116 | 0.019 | 0.040 |
| | | RMSE | 0.024 | 0.265 | 0.139 | 0.202 | 0.015 | 0.579 | 0.125 | 0.186 |
| | W-ARMA ($J = 50$) | Bias | -0.005 | 0.031 | 0.002 | 0.036 | -0.005 | 0.116 | 0.015 | 0.041 |
| | | RMSE | 0.023 | 0.267 | 0.142 | 0.191 | 0.015 | 0.580 | 0.150 | 0.190 |
| | W-ARMA ($J = 100$) | Bias | 0.001 | 0.043 | -0.041 | 0.013 | -0.001 | 0.134 | -0.043 | 0.012 |
| | | RMSE | 0.023 | 0.271 | 0.152 | 0.176 | 0.014 | 0.592 | 0.160 | 0.173 |
| | W-ARMA ($J = 200$) | Bias | 0.009 | 0.062 | -0.112 | -0.021 | 0.005 | 0.164 | -0.149 | -0.033 |
| | | RMSE | 0.026 | 0.278 | 0.190 | 0.159 | 0.015 | 0.613 | 0.222 | 0.156 |
| 2000 | W-ARMA ($J = 10$) | Bias | -0.003 | 0.013 | 0.010 | 0.020 | -0.003 | 0.055 | 0.007 | 0.018 |
| | | RMSE | 0.015 | 0.182 | 0.095 | 0.128 | 0.010 | 0.369 | 0.086 | 0.121 |
| | W-ARMA ($J = 50$) | Bias | -0.002 | 0.016 | -0.000 | 0.015 | -0.003 | 0.054 | 0.008 | 0.020 |
| | | RMSE | 0.016 | 0.183 | 0.101 | 0.125 | 0.010 | 0.369 | 0.111 | 0.125 |
| | W-ARMA ($J = 100$) | Bias | 0.001 | 0.023 | -0.025 | 0.003 | -0.001 | 0.063 | -0.023 | 0.006 |
| | | RMSE | 0.016 | 0.184 | 0.107 | 0.120 | 0.010 | 0.374 | 0.115 | 0.119 |
| | W-ARMA ($J = 200$) | Bias | 0.006 | 0.035 | -0.069 | -0.017 | 0.003 | 0.080 | -0.083 | -0.020 |
| | | RMSE | 0.017 | 0.188 | 0.129 | 0.114 | 0.010 | 0.383 | 0.147 | 0.114 |
| 5000 | W-ARMA ($J = 10$) | Bias | -0.001 | 0.002 | 0.003 | 0.009 | -0.001 | 0.016 | 0.002 | 0.010 |
| | | RMSE | 0.009 | 0.115 | 0.060 | 0.076 | 0.006 | 0.225 | 0.053 | 0.073 |
| | W-ARMA ($J = 50$) | Bias | -0.001 | 0.003 | -0.000 | 0.007 | -0.001 | 0.015 | 0.004 | 0.012 |
| | | RMSE | 0.010 | 0.116 | 0.065 | 0.076 | 0.006 | 0.225 | 0.072 | 0.076 |
| | W-ARMA ($J = 100$) | Bias | 0.000 | 0.006 | -0.011 | 0.002 | -0.000 | 0.019 | -0.009 | 0.006 |
| | | RMSE | 0.010 | 0.116 | 0.067 | 0.075 | 0.006 | 0.226 | 0.073 | 0.075 |
| | W-ARMA ($J = 200$) | Bias | 0.003 | 0.012 | -0.032 | -0.007 | 0.001 | 0.026 | -0.036 | -0.006 |
| | | RMSE | 0.010 | 0.117 | 0.075 | 0.074 | 0.006 | 0.228 | 0.084 | 0.073 |

Notes: Bias and RMSE are computed over $10,000$ Monte Carlo replications. Moderate (high) persistence corresponds to $\phi = 0.90$ ($\phi = 0.95$). Variance parameters are set to $\sigma_y = 1$ and $\sigma_\nu = 1.5$. In the heavier-tail cases correspond to $\nu = 1.20$ (top) and $\nu = 1.00$ (bottom). $J$ denotes the winsorisation parameter.

Table 9: Empirical performance of hypothesis tests for heavy tails with Student-t distribution and alternative null hypothesis

| $T$ | $H_0$: $\nu = 3$ vs. $H_1$: $\nu \neq 3$ | | | | | | $H_0$: $\nu = 5$ vs. $H_1$: $\nu \neq 5$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi = 0.90$ | | | $\phi = 0.95$ | | | $\phi = 0.90$ | | | $\phi = 0.95$ | | |
| | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC | Asy | LMC | MMC |
| | Empirical size | | | | | | | | | | | |
| 500 | 3.0 | 4.4 | 0.5 | 0.7 | 4.3 | 0.4 | 0.9 | 5.8 | 1.2 | 0.1 | 4.5 | 0.5 |
| 1000 | 7.3 | 5.8 | 1.3 | 3.2 | 3.8 | 0.9 | 4.4 | 5.2 | 1.1 | 1.9 | 6.0 | 1.5 |
| 2000 | 8.8 | 5.8 | 2.1 | 4.8 | 4.1 | 1.4 | 6.9 | 5.8 | 2.0 | 2.4 | 4.0 | 1.6 |
| 5000 | 11.3 | 4.0 | 0.9 | 6.3 | 3.3 | 1.1 | 11.3 | 4.0 | 0.4 | 6.2 | 3.9 | 0.7 |
| 10000 | 9.4 | 4.0 | 0.3 | 7.0 | 4.0 | 0.9 | 10.8 | 3.9 | 0.4 | 7.4 | 3.9 | 0.7 |
| | Empirical power | | | | | | | | | | | |
| 500 | 13.8 | 9.9 | 1.0 | 5.2 | 8.9 | 0.7 | 2.4 | 2.9 | 0.3 | 0.6 | 1.7 | 0.4 |
| 1000 | 23.0 | 20.8 | 6.4 | 13.1 | 22.3 | 5.2 | 1.3 | 1.5 | 0.2 | 0.2 | 1.3 | 0.2 |
| 2000 | 58.5 | 54.7 | 31.6 | 44.6 | 51.9 | 28.2 | 1.1 | 1.8 | 0.0 | 0.1 | 2.0 | 0.0 |
| 5000 | 93.8 | 93.8 | 82.2 | 90.0 | 92.2 | 80.5 | 23.7 | 23.7 | 0.9 | 7.4 | 20.7 | 1.8 |
| 10000 | 99.8 | 99.8 | 98.4 | 99.9 | 99.9 | 98.8 | 59.4 | 57.5 | 22.0 | 49.4 | 59.3 | 27.6 |

Notes: Empirical performance are computed over 1,000 replications. Moderate (high) persistence corresponds to $\phi = 0.90$ ($\phi = 0.95$). Variance parameters are set to $\sigma_y = 1$ and $\sigma_\nu = 1.5$. Throughout, the winsorisation parameter is fixed at $J = 100$. The parameter $\nu$ takes the value specified by the null hypothesis being when evaluating the empirical size (i.e. $\nu = \{3, 5\}$) - top panels) or $\nu = 30$ when evaluating the empirical power of the test (bottom panels).

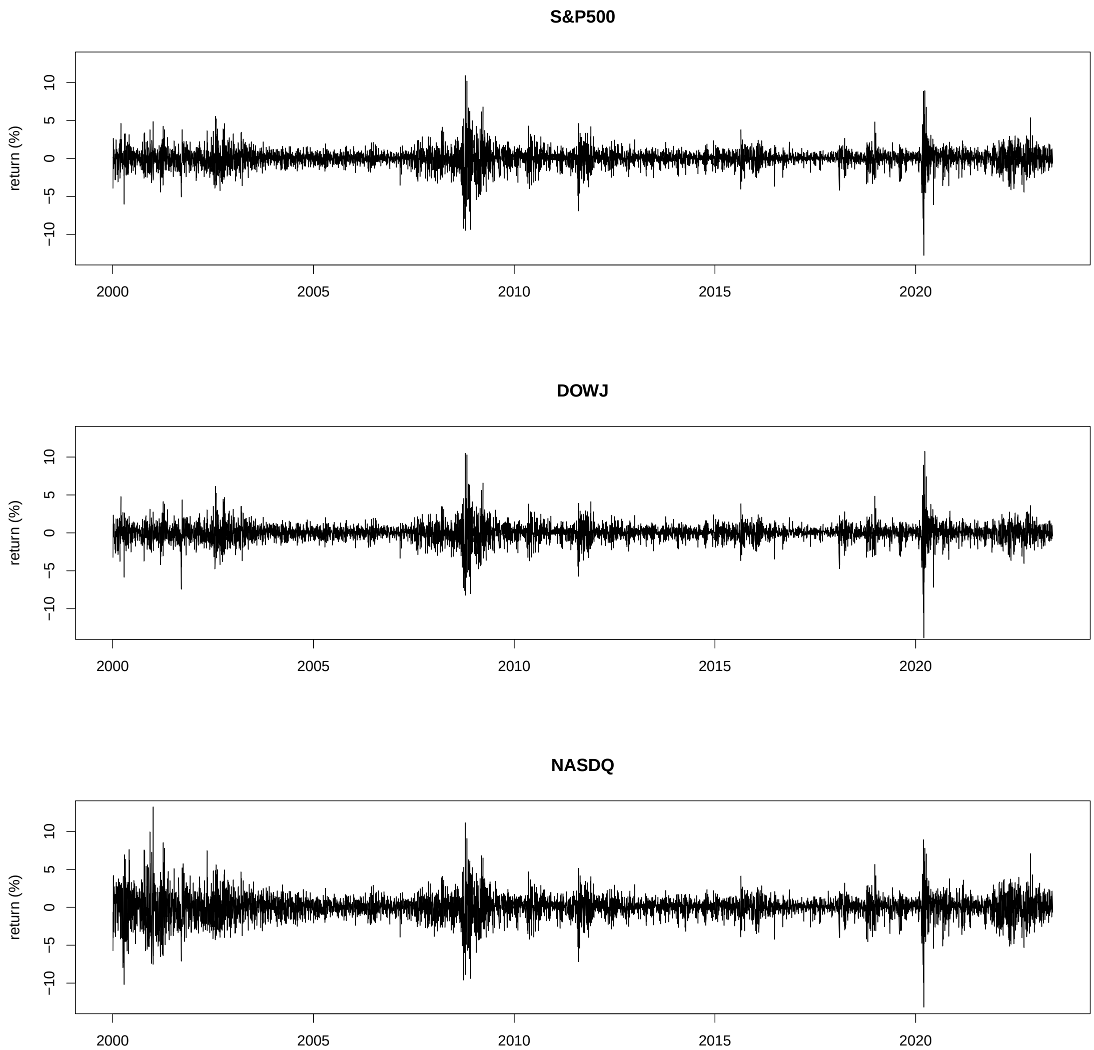
## 9.3 Complementary empirical results

Table 10: Summary statistics

| | Sample from 2000-Jan-04 to 2023-May-31 ($T = 5,889$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Series | Mean | SD | Kurtosis | SK | Range | Min | Max | LB(10) |
| 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 1: 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.