Estimating Stochastic Volatility Models with Moving Average Components: Methodology and Applications*

Meilin Tong† Jean-Marie Dufour‡
McGill University McGill University

October 2026

* We thank John W. Galbraith, Victoria Zinde-Walsh, Silvia Gonçalves, Md. Nazmul Ahsan, Gabriel Rodriguez-Rondon, Marine Carrasco, Kensuke Sakamoto, and participants at the 59th Annual Meetings of the Canadian Economics Association, the CIREQ Lunch Seminar, the 20th CIREQ–CIRANO Ph.D. Students' Conference, and the 18th Annual Society for Financial Econometrics (SoFiE) Conference for helpful comments and suggestions. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. E-mail: meilin.tong@mail.mcgill.ca.

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

ABSTRACT

Moving-average dynamics provide a parsimonious way to capture short-run time-series patterns that may otherwise require additional autoregressive lags. Yet they are rarely incorporated into stochastic volatility models because latent volatility complicates estimation and model selection. We develop a framework for a broad class of stochastic volatility models in which latent log-volatility follows an ARMA(p, q) process, denoted SV(p, q). For this class of models, we show that appropriately transformed returns admit an ARMA representation. The proposed estimation procedure then consists of two steps: it first estimates the ARMA parameters by regression and then recovers the structural parameters through autocovariance matching. Consistency and asymptotic normality are established for the regression estimator. To guide empirical implementation, we develop a two-stage procedure for selecting the model orders. Monte Carlo simulations assess the method's finite-sample performance. In an application to daily S&P 500 returns from 1962 to 2024, the model-selection procedure favors an SV$(1, 1)$ specification, supporting the inclusion of a moving-average component in latent volatility. The empirical analysis also evaluates volatility forecasts at multiple horizons against standard GARCH-family benchmarks. Together, these contributions provide a practical framework for estimation, model selection, and forecasting with richer volatility dynamics.

Key words: Stochastic volatility; ARMA log-volatility; Hannan–Rissanen estimation; backward mapping; model selection; volatility forecasting.

JEL classification: C22; C58; G17.

1 Introduction

Asset price volatility is a central state variable in finance: it governs risk premia and tail exposures, enters portfolio choice and risk management, and is a key input for derivative valuation. Its importance also extends to macroeconomics, where fluctuations in uncertainty can affect investment and hiring, while changes in the volatility of economic shocks matter for macroeconomic forecasting (Bloom, 2009; Stock and Watson, 2007). An important feature of volatility is that it evolves over time: shocks may generate substantial short-run changes in risk, while their effects can persist long after the initial disturbance. Understanding these dynamics is therefore essential not only for measuring current risk, but also for assessing how risk evolves and forecasting its future path. A natural research question is therefore how to build volatility models that are economically interpretable, flexible enough to capture both short-run adjustments and persistent effects, and practical for estimation, model selection, and forecasting from observed returns.

Stochastic volatility (SV) models (Taylor, 1982) provide a natural framework for addressing this question and offer three important advantages that motivate their use alongside standard GARCH models (Bollerslev, 1986). First, unlike standard GARCH models, where conditional variance is determined by past returns and conditional variances, SV models describe volatility as a latent stochastic process driven by its own innovations. This allows volatility shocks to be modeled directly and provides a state-space representation suitable for filtering and forecasting (Harvey et al., 1994; Shephard and Andersen, 2009). Second, SV models can be easily linked to continuous-time diffusion frameworks used in asset pricing. In particular, the standard lognormal SV model can be viewed as a discrete-time counterpart of a diffusion process for latent log-volatility (Asai, 2008; Lindner, 2009; López-Pérez et al., 2025). Third, certain broader classes of SV models are closed under temporal aggregation, meaning that returns aggregated to a lower sampling frequency retain a representation within the same model class (Meddahi and Renault, 2004), whereas conventional strong GARCH models do not generally possess this property (Drost and Nijman, 1993; Engle and Patton, 2001).

Choosing an SV framework still leaves a central modeling question: how should latent log-volatility depend on its past and on volatility shocks? The SV(p) models studied by Ahsan and Dufour (2021) specify autoregressive dynamics for latent log-volatility, allowing shocks to have persistent effects. We extend this autoregressive specification to SV(p, q) by allowing past volatility innovations to enter directly through moving-average (MA) terms. This provides additional flexibility to distinguish the initial adjustment to a volatility shock from the persistence of its remaining effect. For example, in SV(1), a single autoregressive coefficient governs both the response one period after a shock and its subsequent decay.

Adding an MA component allows that response to fall sharply even when the remaining effect persists. Although higher-order autoregressions can also capture richer responses, MA terms offer a parsimonious alternative. This motivation parallels the move from ARCH to GARCH, where introducing lagged conditional variances allows richer dynamics without requiring a large number of ARCH lags (Bollerslev, 1986). The extension also has a structural motivation: sampling certain continuous-time linear systems, including continuous-time ARMA (CARMA) models for latent log-volatility, can generate discrete-time ARMA representations with MA terms (Chambers, 1999; Chambers and Thornton, 2012; Lindner, 2009).

The flexibility of SV models comes with an econometric cost: likelihood-based inference typically requires integrating over an unobserved volatility process, often through computationally intensive numerical methods or simulation. To address this challenge, we develop a unified framework for estimating SV(p, q) models and selecting their structural orders, without requiring likelihood optimization or simulation of latent volatility paths. Existing approaches include the MCMC framework of Kim et al. (1998) for likelihood-based SV inference and the sequential Monte Carlo methods of Urteaga et al. (2017) for latent ARMA state and parameter estimation. These methods rely on repeated simulation of unobserved states, which can be computationally demanding, particularly when multiple candidate models must be evaluated. Our approach instead builds on observable ARMA representations of transformed returns used in the SV estimation methods of Francq and Zakoïan (2006); Ahsan and Dufour (2019, 2021); Ahsan et al. (2025a). We extend the strategy to latent ARMA volatility dynamics, using the observable representation to recover structural parameters and distinguish competing autoregressive and moving-average orders.

The key step in our approach is the log-squared return transformation, which provides an observable measure of latent log-volatility. Under the maintained assumption of Gaussian return innovations independent of volatility innovations, centered log-squared returns equal latent log-volatility plus measurement noise whose distribution, and hence variance, is known. We show that the transformed series admits a weak ARMA(p, m) representation, with $m = \max\{p, q\}$ under the maintained model conditions, including a noncancellation restriction. We estimate this observable reduced form using a three-step Hannan–Rissanen regression procedure (Hannan and Rissanen, 1982; Dufour and Pelletier, 2022), and then recover the structural volatility parameters by matching the autocovariances implied by the observable ARMA and latent SV representations, accounting for the known measurement noise and imposing admissibility restrictions. For model selection, the observable ARMA orders do not always reveal the MA order of latent volatility. For example, both SV(1) and SV$(1, 1)$ can generate an ARMA$(1, 1)$ representation for transformed returns. We therefore develop a two-stage model-selection procedure: an information criterion first selects the

observable ARMA orders, while a bootstrap-weighted, penalized minimum-distance criterion distinguishes candidate structural MA orders when necessary.

The paper makes three main contributions. First, we characterize the relationship between latent ARMA volatility dynamics and the observable reduced-form representation, establishing conditions for the exact reduced-form order and unique admissible structural parameter recovery for a given MA order. Second, we extend regression-based estimation to SV(p, q) models, connecting observable reduced-form estimates to the structural volatility parameters. We further establish consistency and asymptotic normality of the Hannan–Rissanen reduced-form coefficient estimator under fixed, correctly specified orders. Third, we develop a structural model-selection procedure that exploits covariance restrictions to distinguish SV specifications sharing the same observable ARMA orders. This provides a systematic basis for assessing whether, and at what order, moving-average dynamics are supported in latent volatility.

Monte Carlo experiments assess the finite-sample performance of the proposed estimation and model-selection procedures. Estimation accuracy generally improves with sample size, and the two-stage procedure selects the correct structural order in 94.4–100% of replications across ten designs at $T = 10{,}000$. In the application to daily S&P 500 returns from 1962 to 2024, the procedure selects SV$(1, 1)$, with the result unchanged across the three bootstrap block lengths considered. The estimated MA component implies a sharp initial adjustment followed by a smaller, persistent effect: following a volatility shock normalized to raise conditional volatility by 10% on impact, the increase falls to approximately 1.4% after one trading day and then decays gradually. A recursive pseudo out-of-sample forecasting exercise further demonstrates that the proposed estimation, structural recovery, and filtering procedures can be implemented repeatedly to produce multi-horizon variance forecasts. SV$(1, 1)$ delivers forecast accuracy broadly comparable to standard GARCH-family benchmarks, although it does not achieve the lowest out-of-sample MSE in the evaluation period.

The framework may also be relevant for macroeconomic uncertainty measurement, where stochastic volatility models capture time-varying innovation variances and forecast-error uncertainty (Stock and Watson, 2007; Jurado et al., 2015; Ahsan et al., 2025b). Exploring this application is left for future research.

The remainder of the paper is organized as follows. Section 2 introduces the SV(p, q) framework and establishes the reduced-form representation and parameter correspondence. Section 3 presents the estimation methods, and Section 4 develops asymptotic results for the Hannan–Rissanen estimator. Section 5 presents the two-stage model-selection procedure. Section 6 reports Monte Carlo results, Section 7 presents the S&P 500 application and forecasting exercise, and Section 8 concludes.

2 Framework

Let $\{y_t\}_{t\in\mathbb{Z}}$ denote (demeaned) log returns and let $w_t := \log(\sigma_t^2)$ be the latent log-variance process. We study a discrete-time stochastic volatility model in which returns are conditionally Gaussian given w_t, and the latent log-volatility follows an ARMA(p, q) law. This specification nests the SV(p) framework in Ahsan and Dufour (2021) as the special case $q = 0$ and accommodates moving-average effects motivated by discretization arguments (e.g., Chambers (1999); Chambers and Thornton (2012)).

Assumption 2.1 (Stochastic Volatility (p,q))**.** The observed process $\{y_t : t \in N_0\}$ satisfies

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

$$w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + v_t - \sum_{j=1}^{q} \theta_j v_{t-j} \tag{2.2}$$

where $z_t \overset{i.i.d.}{\sim} N(0,1)$, $v_t \overset{i.i.d.}{\sim} N(0,\sigma_v^2)$, with the two sequences mutually independent. Moreover, $\phi_p \neq 0$ when $p \geq 1$, and $\theta_q \neq 0$ when $q \geq 1$. We normalize $\mathbb{E}[w_t] = 0$; equivalently, a nonzero intercept in the log-volatility equation can be absorbed into σ_y.

Assumption 2.2 (Stability and invertibility)**.** The latent log-volatility process $\{w_t\}$ is stable and invertible.

Assumption 2.3 (Minimality)**.** The latent log-volatility process $\{w_t\}$ is minimal, with p and q denoting its exact autoregressive and moving-average orders. That is, $\phi(z)$ and $\theta(z)$ have no common zeros.

Assumption 2.2 imposes the standard stability and invertibility conditions for w_t: the roots of the autoregressive polynomial $\phi(z)$ and the moving average polynomial $\theta(z)$ all lie outside the unit circle. Stability guarantees that the volatility equation has a unique causal and stationary solution. Since $\{v_t\}$ is i.i.d. Gaussian and $\{z_t\}$ is i.i.d. and independent of $\{v_t\}$, the joint process $(y_t, w_t)'$ is then strictly stationary. Invertibility selects the conventional, uniquely parameterized moving-average representation and is used later in recovering the structural parameters from the reduced form. Assumption 2.3 rules out cancellations that would undermine minimality and complicate identification.

A key device in reduced-form SV estimation is the log-squared transformation (see, e.g., Taylor (1982); Harvey et al. (1994); Kim et al. (1998); Broto and Ruiz (2004); Ahsan and Dufour (2021)). From (2.1),

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

Let $c_z := \mathbb{E}[\log(z_t^2)] = \log 2 + \psi(1/2)$ and define the centered measurement error

$$\epsilon_t := \log(z_t^2) - c_z, \qquad \mathbb{E}[\epsilon_t] = 0, \qquad \mathrm{Var}(\epsilon_t) = \psi'(1/2) = \pi^2/2,$$

where $\psi(\cdot)$ is the digamma function. Define the centered log-squared return series

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

Then (2.3) implies the measurement equation

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

where $\{\epsilon_t\}$ is i.i.d., independent of $\{v_t\}$, and its variance $\sigma_\epsilon^2 = \pi^2/2$ is known. This "known-variance measurement noise" is a central advantage of the transformation: it turns the SV model into an observable time-series problem with a structured error component. The SV(p,q) model can be written in the state-space form:

$$\begin{aligned} \text{State Transition Equation: } & w_t = \sum_{j=1}^{p} \phi_j w_{t-j} + v_t - \sum_{j=1}^{q} \theta_j v_{t-j} \\ \text{Measurement Equation: } & y_t^* = w_t + \epsilon_t \end{aligned} \tag{2.6}$$

where v_t's are i.i.d. $N(0, \sigma_v^2)$ and ϵ_t's are i.i.d. centered $log(\chi_1^2)$.

Assumption 2.4 (Reduced-form noncancellation)**.** For $p = q \geq 1$, impose $\theta_p \sigma_v^2 + \phi_p \sigma_\epsilon^2 \neq 0$.

2.1 ARMA representation of SV(p, q)

We now establish that y_t^* admits a reduced-form ARMA representation whose moving-average order is at most $m = \max\{p, q\}$.

Theorem 2.1 (ARMA representation of SV(p, q))**.** *Under Assumptions 2.1–2.3, let* $m := \max\{p, q\}$*. Then, for some* $r \leq m$*, the transformed series* $\{y_t^*\}$ *satisfies the ARMA*(p, r) *representation*

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + \eta_t - \sum_{j=1}^{r} \theta_j^* \eta_{t-j}, \tag{2.7}$$

$$\eta_t - \sum_{j=1}^{r} \theta_j^* \eta_{t-j} = v_t - \sum_{j=1}^{q} \theta_j v_{t-j} + \epsilon_t - \sum_{j=1}^{p} \phi_j \epsilon_{t-j}, \tag{2.8}$$

where $\{\eta_t\}$ *is weak white noise with variance* $\sigma_\eta^2 > 0$*. When the reduced-form MA representation is taken in its standard invertible form, the coefficients* $(\theta_1^*, \ldots, \theta_r^*)$ *and* σ_η^2 *are uniquely determined. Moreover, the reduced-form AR and MA polynomials have no common roots, so the ARMA*(p, r) *representation is minimal.*

Theorem 2.1 shows that the log-squared transformation converts the latent SV(p, q) model into an observable ARMA model. This representation forms the basis for the regression-based estimation and order-selection procedures developed below. The term "weak" indicates that η_t is serially uncorrelated, but need not be independent. We next characterize its exact moving-average order; under the noncancellation condition maintained throughout the remainder of the paper, $r = m = \max\{p, q\}$.

The remaining question is whether the upper bound $r \leq m = \max\{p, q\}$ is attained. When $p \neq q$, the equality $r = m$ follows directly because only one of the two moving-average components contributes at the outermost lag. When $p = q$, the two outer-lag contributions may cancel. We exclude this special case through the following condition.

Proposition 2.2 (Exact reduced-form order)**.** *Under the assumptions of Theorem 2.1 and Assumption 2.4, $r = m = \max\{p, q\}$. Hence, the transformed process has a minimal ARMA(p, m) representation.*

Remark 2.1 (Special cases). The result nests several familiar specifications. When $q = 0$, the model reduces to the SV(p) specification considered by Ahsan and Dufour (2021). For $p > 0$, $\phi(B)y_t^* = v_t + \phi(B)\epsilon_t$, so the filtered disturbance is the sum of independent MA(0) and MA(p) processes and therefore has order p. Hence y_t^* has a weak ARMA(p, p) representation. Similarly, when $p = 0$ and $q > 0$, $y_t^* = \theta(B)v_t + \epsilon_t$ is the sum of independent MA(q) and MA(0) processes and therefore has order q, yielding a weak ARMA$(0, q)$ representation.

Proposition 2.3 (Autocovariances of the Observed Process)**.** *Under the assumptions of Theorem 2.1, let $m := \max\{p, q\}$ and write*

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + \left(v_t - \sum_{j=1}^{m} \theta_j v_{t-j} \right) + \left(\epsilon_t - \sum_{j=1}^{m} \phi_j \epsilon_{t-j} \right), \tag{2.9}$$

where $\theta_j = 0$ for $j > q$ and $\phi_j = 0$ for $j > p$. Let ψ_h denote the impulse-response coefficients determined recursively by the structural parameters:

$$\psi_0 = 1, \qquad \psi_h = \sum_{j=1}^{\min\{p,h\}} \phi_j \psi_{h-j} - \theta_h, \qquad h \geq 1. \tag{2.10}$$

Then the autocovariances $\gamma_{y^*}(k) := \mathrm{Cov}(y_t^*, y_{t-k}^*)$ satisfy

$$\gamma_{y^*}(k) = \begin{cases} \displaystyle\sum_{j=1}^{p} \phi_j \gamma_{y^*}(j) + \sigma_v^2 \left(1 - \sum_{j=1}^{m} \theta_j \psi_j \right) + \sigma_\epsilon^2, & k = 0, \\ \displaystyle\sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) - \sigma_v^2 \left(\theta_k + \sum_{l=1}^{m-k} \theta_{k+l} \psi_l \right) - \phi_k \sigma_\epsilon^2, & 1 \leq k \leq m, \\ \displaystyle\sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j), & k > m, \end{cases} \tag{2.11}$$

where empty sums are interpreted as zero and $\gamma_{y^*}(-k) = \gamma_{y^*}(k)$*. The coefficients* ψ_h *are functions of the structural AR and MA coefficients.*

These identities express the autocovariances in terms of the original structural parameters. Only $\psi_0, \ldots, \psi_q$ enter the non-autoregressive terms, so they can be evaluated by a finite recursion. Beyond lag m, the autocovariances satisfy the autoregressive recursion.

2.2 Parameter transformation

We maintain Assumption 2.4, so that Proposition 2.2 gives $m = \max\{p, q\}$ as the exact reduced-form MA order. Parameter conversion follows by matching the autocovariances of the two representations of the filtered disturbance,

$$u_t := \phi(B)y_t^* = \theta(B)v_t + \phi(B)\epsilon_t = \theta^*(B)\eta_t.$$

With $\theta_j = 0$ for $j > q$ and $\phi_j = 0$ for $j > p$, these identities imply

$$\begin{aligned}
\left(1 + \sum_{j=1}^{q} \theta_j^2\right)\sigma_v^2 + \left(1 + \sum_{j=1}^{p} \phi_j^2\right)\sigma_\epsilon^2 &= \left(1 + \sum_{j=1}^{m} (\theta_j^*)^2\right)\sigma_\eta^2, && k = 0, \\
\left(-\theta_k + \sum_{j=1}^{q-k} \theta_j\theta_{j+k}\right)\sigma_v^2 + \left(-\phi_k + \sum_{j=1}^{p-k} \phi_j\phi_{j+k}\right)\sigma_\epsilon^2 &= \left(-\theta_k^* + \sum_{j=1}^{m-k} \theta_j^*\theta_{j+k}^*\right)\sigma_\eta^2, && 1 \le k \le m.
\end{aligned} \tag{2.12}$$

Empty sums are interpreted as zero. When $m = 0$, only the variance equation applies. These equations concern the autocovariances of u_t, rather than those of y_t^* itself.

In the forward direction, the structural parameters determine the reduced-form MA coefficients and innovation variance. Conversely, for a fixed structural order q, any admissible reduced form generated by an SV(p, q) model maps back to a unique structural parameter vector. Thus, for fixed q, the covariance-matching equations establish a one-to-one relationship between admissible structural parameters and their image in the reduced-form parameter space. The following proposition formalizes this forward and backward parameter conversion.

Proposition 2.4 (Uniqueness of Parameter Conversion)**.** (i) *Under Assumptions 2.1–2.4, given the structural parameters and* σ_ϵ^2*, there is a unique normalized invertible reduced-form solution* $(\theta^*(B), \sigma_\eta^2)$ *satisfying the covariance-matching system* (2.12)*.*

(ii) *Conversely, fix a structural order* $q \in \{0, \ldots, m\}$*. Given* $(\phi(B), \theta^*(B), \sigma_\eta^2)$ *and known* σ_ϵ^2*, there exists at most one admissible pair* $(\theta(B), \sigma_v^2)$*, where* $\theta(B)$ *is a normalized invertible MA polynomial of exact order* q *and* $\sigma_v^2 > 0$*, satisfying* (2.12)*.*

Part (ii) makes clear the role of structural admissibility. An arbitrary reduced-form parameter vector may not correspond exactly to an SV(p, q) model. For a fixed q, existence

requires that, after removing the known covariance contribution of the measurement error, the remaining autocovariance structure be generated by an invertible MA(q) process with positive innovation variance. This condition is automatically satisfied at population parameter values generated by the maintained SV model. With estimated reduced-form parameters, however, the restrictions need not hold exactly, which motivates the constrained minimum-distance implementation used below.

In the forward direction, once the structural orders (p, q) are specified, the exact reduced-form order is determined as (p, m) with $m = \max\{p, q\}$. The backward direction is different. If the reduced-form ARMA model has $m > p$, then $m = \max\{p, q\}$ immediately implies $q = m$. If $m = p$, however, the order pair (p, m) alone is compatible with any $q \leq p$. For example, SV$(2, 0)$, SV$(2, 1)$, and SV$(2, 2)$ can all generate reduced-form ARMA$(2, 2)$ models for suitable parameter values. They do not, however, imply the same complete covariance restrictions. Under the maintained exact-order assumptions, the full population reduced-form parameters and known σ_ϵ^2 determine the structural MA order through these restrictions; once q is fixed, Proposition 2.4 implies that the admissible backward mapping is unique.

In finite samples, the population covariance restrictions will generally hold only approximately. We therefore first estimate the reduced-form ARMA(p, m) model. If $m > p$, we set $q = m$; if $m = p = 0$, we set $q = 0$; and if $m = p > 0$, we compare the candidate orders $q \in \{0, \ldots, m\}$ using the complete covariance-matching system and constrained profile minimum distance. The weighting and penalized comparison across candidate orders are described in Section 5.

The first-order case provides a simple explicit illustration of the backward conversion.

Corollary 2.5 (Explicit Conversion for ARMA$(1, 1)$)**.** *Let $(\phi_1, \theta_1^*, \sigma_\eta^2)$ specify a minimal, causal, and invertible weak ARMA$(1, 1)$ model, and suppose $\sigma_\epsilon^2 > 0$ is known. Define*

$$C := \left(1 + (\theta_1^*)^2\right)\sigma_\eta^2 - (1 + \phi_1^2)\sigma_\epsilon^2, \qquad D := \theta_1^*\sigma_\eta^2 - \phi_1\sigma_\epsilon^2.$$

Then an invertible structural representation with MA order at most one and $\sigma_v^2 > 0$ exists if and only if $C > 2|D|$. When this condition holds, the admissible structural parameters are uniquely given by

$$\sigma_v^2 = \frac{C + \sqrt{C^2 - 4D^2}}{2}, \qquad \theta_1 = \frac{D}{\sigma_v^2}. \tag{2.13}$$

If $D \neq 0$, the structural MA order is one, so the corresponding model is SV$(1, 1)$. If $D = 0$, then $\theta_1 = 0$ and $\sigma_v^2 = C$, corresponding to SV$(1, 0)$.

3 Estimation method

For fixed structural orders (p, q), this section presents three approaches to estimating the SV model without direct likelihood optimization over the latent volatility process. We first develop a simple estimator based on autocovariance restrictions and winsorized variants that combine information across lag blocks. We then present the three-step Hannan–Rissanen procedure, our main estimation method, which estimates the observable ARMA representation of centered log-squared returns and recovers the structural parameters through covariance matching. In each case, the scale parameter is obtained from the sample mean of log-squared returns.

3.1 Simple ARMA-based estimator

We first consider a simple ARMA-based estimator that exploits a finite set of autocovariances of $y_t^* = \log(y_t^2) - \mu$. This extends the autocovariance approach for SV(p) models in Ahsan and Dufour (2021) to the SV(p, q) framework and uses moments of centered log-squared returns. We denote this estimator by S-ARMA-based.

For fixed structural orders (p, q), let $m = \max\{p, q\}$. By Proposition 2.3, the autocovariances satisfy the AR recursion at lags greater than m. For $p \geq 1$ and an integer $j \geq m - p + 1$, define

$$\gamma(p, j) := \begin{bmatrix} \gamma_{y^*}(p+j) & \gamma_{y^*}(p+j+1) & \dots & \gamma_{y^*}(2p+j-1) \end{bmatrix}'$$
$$\Gamma(p, j) := \begin{bmatrix} \gamma_{y^*}(p+j-1) & \gamma_{y^*}(p+j-2) & \dots & \gamma_{y^*}(j) \\ \gamma_{y^*}(p+j) & \gamma_{y^*}(p+j-1) & \dots & \gamma_{y^*}(j+1) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{y^*}(2p+j-2) & \gamma_{y^*}(2p+j-3) & \dots & \gamma_{y^*}(p+j-1) \end{bmatrix}, \tag{3.1}$$

where $\gamma_{y^*}(k) := \mathrm{Cov}(y_t^*, y_{t-k}^*)$, with $y_t^* = [log(y_t^2) - \mu]$ and $\mu = E[log(y_t^2)]$.

Corollary 3.1 (Expression for SV(p, q) Parameters)**.** *Under the assumptions of Theorem 2.1 and the autocovariance structure in Proposition 2.3, let $m := \max\{p, q\}$. Then:*

(i) AR parameters. *If $p \geq 1$, the AR parameter vector $\Phi_p := (\phi_1, \dots, \phi_p)'$ satisfies the Yule–Walker equations at lags strictly larger than m. More precisely, for any integer j such that $p + j > m$, $\Gamma(p, j)\Phi_p = \gamma(p, j)$. Whenever $\Gamma(p, j)$ is nonsingular, we therefore have*

$$\Phi_p = \Gamma(p, j)^{-1} \gamma(p, j), \tag{3.2}$$

where $\gamma(p, j)$ and $\Gamma(p, j)$ are defined in (3.1). When $p = 0$ this step is vacuous.

(ii) MA and variance parameters. *Let* $\Theta := (\theta_1, \ldots, \theta_q, \sigma_v^2)'$ *and, for a fixed integer* $K \geq m$*, collect the autocovariances* $\gamma(0:K) := \big(\gamma_{y^*}(0), \gamma_{y^*}(1), \ldots, \gamma_{y^*}(K)\big)'$*. For given* $(\Phi_p, \sigma_\varepsilon^2)$*, denote by*

$$G_K(\Phi_p, \Theta) := \big(\gamma_{y^*}(0; \Phi_p, \Theta), \ldots, \gamma_{y^*}(K; \Phi_p, \Theta)\big)'$$

the vector of theoretical autocovariances implied by the SV(p, q) *model and Proposition 2.3. The true* Θ *is the unique admissible solution of*

$$F(\Theta) := G_K(\Phi_p, \Theta) - \gamma(0{:}K) = 0, \tag{3.3}$$

where admissibility requires positive innovation variance, invertibility, and no common roots with the AR polynomial.

(iii) Return scale. *The return-scale parameter in (2.1) satisfies*

$$\sigma_y^2 = \exp(\mu - c_z), \tag{3.4}$$

where $\mu = E[\log(y_t^2)]$ *and* $c_z = E[\log(z_t^2)]$*. Equivalently,* $\sigma_y = \exp\{(\mu - c_z)/2\}$*.*

For a stable estimate $\widehat{\Phi}_p$, we estimate the remaining parameters by constrained nonlinear least squares:

$$\widehat{\Theta} \in \arg \min_{\Theta \in \mathcal{A}_q(\widehat{\Phi}_p)} \left\| G_K(\widehat{\Phi}_p, \Theta) - \widehat{\gamma}(0{:}K) \right\|_2^2, \tag{3.5}$$

whenever an admissible minimum is attained. Here $\mathcal{A}_q(\widehat{\Phi}_p)$ imposes positive innovation variance, an invertible MA polynomial of degree at most q, and no common roots with the fitted AR polynomial.

When $q \leq p$ and $K = p$, the moment system has $p + 1$ restrictions and $q + 1$ free MA and variance parameters; zero padding does not add unknown parameters.

SV$(1,1)$ **illustration.** We can obtain the closed-form solution for SV$(1,1)$. By Theorem 2.1, y_t^* has an ARMA$(1,1)$ representation and the autocovariance equations for any lag k can be represented in terms of parameters $(\phi_1, \theta_1, \sigma_v^2, \sigma_\epsilon^2)$,

$$\begin{aligned} \gamma_{y^*}(0) &= \phi_1 \gamma_{y^*}(-1) + (1 - \theta_1(\phi_1 - \theta_1))\sigma_v^2 + \sigma_\epsilon^2, \\ \gamma_{y^*}(1) &= \phi_1 \gamma_{y^*}(0) - \theta_1 \sigma_v^2 - \phi_1 \sigma_\epsilon^2, \\ \gamma_{y^*}(k) &= \phi_1 \gamma_{y^*}(k-1) \qquad \text{for } k > 1. \end{aligned} \tag{3.6}$$

This gives

$$\phi_1 = \frac{\gamma_{y^*}(2)}{\gamma_{y^*}(1)}. \tag{3.7}$$

The denominator is nonzero under the maintained stable, invertible, and minimal exact SV$(1,1)$ specification. Define

$$\begin{aligned} a &:= -\gamma_{y^*}(1) + \phi_1\gamma_{y^*}(0) - \phi_1\sigma_\epsilon^2 = \theta_1\sigma_v^2, \\ b &:= \gamma_{y^*}(0) - \phi_1\gamma_{y^*}(1) - \sigma_\epsilon^2 = (1 - \phi_1\theta_1 + \theta_1^2)\sigma_v^2. \end{aligned} \tag{3.8}$$

Thus $b + \phi_1 a = (1 + \theta_1^2)\sigma_v^2$. Applying Corollary 2.5 with $C = b + \phi_1 a$ and $D = a$ yields

$$\sigma_v^2 = \frac{b + \phi_1 a + \sqrt{(b + \phi_1 a)^2 - 4a^2}}{2}, \qquad \theta_1 = \frac{a}{\sigma_v^2}, \tag{3.9}$$

provided $b + \phi_1 a > 2|a|$. The same expression covers the nested SV$(1,0)$ case: if $a = 0$ and $b > 0$, then $\theta_1 = 0$ and $\sigma_v^2 = b$.

In the sample, define

$$\widehat{\mu} := \frac{1}{T}\sum_{t=1}^{T} \log(y_t^2), \qquad \widehat{y}_t^* := \log(y_t^2) - \widehat{\mu},$$

and

$$\widehat{\gamma}_{y^*}(k) := \frac{1}{T-k}\sum_{t=1}^{T-k} \widehat{y}_t^*\widehat{y}_{t+k}^*, \qquad k = 0, \ldots, T-1. \tag{3.10}$$

The sample estimator replaces the population autocovariances by these quantities. Its denominator, stability, and admissibility conditions must be checked in finite samples.

When the order (p, q) is greater than 1, we need to numerically solve the nonlinear system of $m + 1$ equations to obtain the parameters. Invertibility identifies the population solution but does not guarantee a unique finite-sample optimizer.

3.2 ARMA-Winsorized estimator

We extend the simple ARMA-based estimator by combining information from J autocovariance lag blocks. The four ARMA-Winsorized estimators use an arithmetic mean, a linearly declining weighted mean, a median, or a least-squares fit. Following the construction for autoregressive SV models in Ahsan and Dufour (2021), these rules provide alternative ways to use the covariance restrictions. Such improvement can improve finite-sample stability and estimation accuracy.

From (3.2), it is clear that Φ_p can be expressed using multiple formulas when j is set to different values. In other words, we can provide a more general class of ARMA-based estimators that take the weighted average of the estimate of Φ_p. For $p \geq 1$, consider the valid block indices

$$j_r := m - p + r, \qquad r = 1, \ldots, J.$$

Then $p + j_r > m$ for every block. Choose J so that all required sample lags are available: $m + p + J - 1 \leq T - 1$. Whenever the corresponding population matrices are nonsingular, define the block-specific representation using block indices j_r:

$$B(p, j_r) := \Gamma(p, j_r)^{-1}\gamma(p, j_r).$$

Then

$$\Phi_p = \sum_{r=1}^{J} w_r B(p, j_r). \tag{3.11}$$

where the weights $w_r \geq 0$ satisfy $\sum_{r=1}^{J} w_r = 1$.

The sample weighted-average estimator is

$$\widetilde{\Phi}_p := \sum_{r=1}^{J} w_r \widehat{B}(p, j_r), \qquad \widehat{B}(p, j_r) := \widehat{\Gamma}(p, j_r)^{-1}\widehat{\gamma}(p, j_r), \tag{3.12}$$

provided the individual sample matrices are nonsingular.

We consider four combination rules. The mean and linearly declining mean use (3.12). The median combines the block estimates componentwise, and OLS fits the stacked covariance equations.

1. **ARMA-Winsorized-Mean**: the estimate $\hat{\Phi}_p^m$ is the arithmetic mean of $\widehat{B}(p, j_r)$ with equal weights:

$$w_r = \frac{1}{J}, \qquad r = 1, \ldots, J,$$

2. **ARMA-Winsorized-LDMean**: the estimate $\hat{\Phi}_p^{ld}$ is the mean of $\widehat{B}(p, j_r)$ with linearly declining (LD) weights:

$$w_r = \frac{2}{J}\left(1 - \frac{r}{J+1}\right), \qquad r = 1, \ldots, J.$$

3. **ARMA-Winsorized-Median**: the estimate is the median-based estimate obtained by taking the median of J estimates of each one of the p components of Φ_p:

$$\widehat{\phi}_i^{\text{med}} := \text{median}\{\widehat{B}_i(p, j_r) : r = 1, \ldots, J\}, \qquad i = 1, \ldots, p.$$

4. **ARMA-Winsorized-OLS**: the estimate is obtained by fitting the stacked covariance equations by least squares without an intercept. Let

$$A(p, J) := \begin{bmatrix} \sqrt{w_1}\,\widehat{\Gamma}(p, j_1) \\ \vdots \\ \sqrt{w_J}\,\widehat{\Gamma}(p, j_J) \end{bmatrix}, \qquad e(p, J) := \begin{bmatrix} \sqrt{w_1}\,\widehat{\gamma}(p, j_1) \\ \vdots \\ \sqrt{w_J}\,\widehat{\gamma}(p, j_J) \end{bmatrix}.$$

If $A(p, J)$ has full column rank,

$$\widehat{\Phi}_p^{\text{ols}} = [A(p, J)'A(p, J)]^{-1}A(p, J)'e(p, J).$$

Here, $A(p, J)$ is a $pJ \times p$ matrix and $e(p, J)$ is a $pJ \times 1$ matrix. This method exploits the fact that Φ_p captures the linear relationship between multiple autocovariance functions, and thus an OLS regression can be used to find the best fitting Φ_p for all empirically available autocovariance functions considered. Note the underlying identity is $\Gamma(p, j_r)\Phi_p = \gamma(p, j_r)$ for $r = 1, .., J$. Using sample versions and weights w_r, a weighted least-squares estimator solves

$$\hat{\Phi}_p^{ols} = \arg\min_{\Phi} \sum_{r=1}^{J} w_r ||\hat{\gamma}(p, j_r) - \hat{\Gamma}(p, j_r)\Phi||^2. \tag{3.13}$$

Equal weights are considered in our context, but different ARMA-Winsorized OLS estimators can be generated if the weights are assigned differently.

For SV(1, 1), $m = p = 1$ and $j_r = r$. With equal weights,

$$\widehat{\phi}_1^{\text{ols}} = \frac{\sum_{r=1}^{J} \widehat{\gamma}_{y^*}(r)\widehat{\gamma}_{y^*}(r+1)}{\sum_{r=1}^{J} \widehat{\gamma}_{y^*}(r)^2}, \tag{3.14}$$

provided the denominator is positive. When $J = 1$, this reduces to $\widehat{\gamma}_{y^*}(2)/\widehat{\gamma}_{y^*}(1)$.

3.3 Hannan–Rissanen method

Our estimation procedure exploits the observable reduced-form representation of the transformed return process. Under the maintained assumptions, including the reduced-form noncancellation condition, y_t^* follows a minimal ARMA(p, m) process. We first estimate this reduced form using a three-step Hannan–Rissanen-type procedure and then recover the structural SV(p, q) parameters through the covariance-matching equations in (2.12). Thus, estimation of the latent volatility model is reduced to estimating an observable ARMA model followed by parameter conversion.

The main computational advantage of this approach is that it avoids direct likelihood optimization over the latent volatility process. Instead, the observable ARMA representation allows the reduced-form parameters to be estimated in three simple steps: a high-order autoregressive approximation is first used to obtain proxies for the unobserved innovations, and these proxies are then refined through two regression-based estimation steps. This same procedure applies to general orders (p, q) without changing the basic estimation structure. Let $\gamma := (\phi_1, \ldots, \phi_p, \theta_1^*, \ldots, \theta_m^*)'$ denote the reduced-form coefficient vector. We estimate γ using the following three-step Hannan–Rissanen procedure; see Hannan and Rissanen (1982) for the original method and Dufour and Pelletier (2022) for its weak-ARMA extension. Empty

lag blocks are omitted, and the regression formulas below require nonsingular cross-product matrices.

In step 1, invertibility of the MA polynomial in $\phi(B)y_t^* = \theta^*(B)\eta_t$ gives the infinite autoregressive representation

$$y_t^* = \sum_{j=1}^{\infty} \Pi_j y_{t-j}^* + \eta_t.$$

We approximate this representation by an AR(n_T) model, estimating its coefficients, intercept, and innovation variance by Gaussian quasi-maximum likelihood. The residuals are

$$\widehat{U}_t := y_t^* - \widehat{c} - \sum_{j=1}^{n_T} \widehat{\Pi}_j^{(n_T)} y_{t-j}^*, \qquad t = n_T + 1, \ldots, T, \tag{3.15}$$

where $\widehat{c}$ is the fitted intercept of the preliminary autoregression. These residuals provide proxies for the reduced-form innovations.

Additionally, the order n_T is supplied to the estimation procedure. In the parameter-estimation simulations, we evaluate several prespecified values of n_T to examine sensitivity to the length of the preliminary autoregression. For each sample-size and preliminary-order configuration, n_T is held fixed across replications. The model-selection experiments use the rule specified in Section 5.

In step 2, we use $\widehat{U}_t$ as proxies for the lagged innovations and form the residual variance

$$\widehat{\sigma}_{\eta,1}^2 := \frac{1}{T} \sum_{t=n_T+1}^{T} \widehat{U}_t^2.$$

Define the regressor row

$$\widehat{Z}_{t-1} := [y_{t-1}^*, \ldots, y_{t-p}^*, -\widehat{U}_{t-1}, \ldots, -\widehat{U}_{t-m}],$$

which has dimension $1 \times (p+m)$. The second-step estimate is

$$\widetilde{\gamma} = \left[\sum_{t=\ell}^{T} \widehat{Z}_{t-1}' (\widehat{\sigma}_{\eta,1}^2)^{-1} \widehat{Z}_{t-1} \right]^{-1} \left[\sum_{t=\ell}^{T} \widehat{Z}_{t-1}' (\widehat{\sigma}_{\eta,1}^2)^{-1} y_t^* \right], \qquad \ell = n_T + m + 1. \tag{3.16}$$

This is the GLS estimator for the regression

$$y_t^* = \widehat{Z}_{t-1}\gamma + e_t.$$

Because the GLS weight is a common positive scalar, the coefficient estimate is identical to OLS. The entries of $\widetilde{\gamma}$ are denoted by $\widetilde{\phi}_i$ and $\widetilde{\theta}_j^*$.

In step 3, we use the second-step coefficients to reconstruct the innovations:

$$\widetilde{U}_t = y_t^* - \sum_{i=1}^{p} \widetilde{\phi}_i y_{t-i}^* + \sum_{j=1}^{m} \widetilde{\theta}_j^* \widetilde{U}_{t-j}, \qquad t = m+1, \ldots, T. \tag{3.17}$$

Using the same reduced-form MA filter, define

$$X_t = y_t^* + \sum_{j=1}^{m} \widetilde{\theta}_j^* X_{t-j}, \tag{3.18}$$

$$W_t = \widetilde{U}_t + \sum_{j=1}^{m} \widetilde{\theta}_j^* W_{t-j}, \tag{3.19}$$

for $t = m+1, \ldots, T$. Replace the first-step residuals in the regressor row by the reconstructed residuals:

$$\widetilde{Z}_{t-1} := [y_{t-1}^*, \ldots, y_{t-p}^*, -\widetilde{U}_{t-1}, \ldots, -\widetilde{U}_{t-m}].$$

The filtered regressor rows are computed as

$$\widetilde{V}_{t-1} = \widetilde{Z}_{t-1} + \sum_{j=1}^{m} \widetilde{\theta}_j^* \widetilde{V}_{t-j-1}, \qquad t = m+1, \ldots, T, \tag{3.20}$$

with $\widetilde{V}_s = 0$ for $s < m$. The residual and response filters are initialized by $\widetilde{U}_t = X_t = W_t = 0$ for $t \leq m$.

Let

$$\widetilde{\sigma}_\eta^2 := \frac{1}{T} \sum_{t=m+1}^{T} \widetilde{U}_t^2.$$

Regressing $\widetilde{U}_t + X_t - W_t$ on $\widetilde{V}_{t-1}$ gives

$$\widehat{\gamma} = \left[\sum_{t=\ell'}^{T} \widetilde{V}_{t-1}' (\widetilde{\sigma}_\eta^2)^{-1} \widetilde{V}_{t-1} \right]^{-1} \left[\sum_{t=\ell'}^{T} \widetilde{V}_{t-1}' (\widetilde{\sigma}_\eta^2)^{-1} (\widetilde{U}_t + X_t - W_t) \right], \qquad \ell' = m+1. \tag{3.21}$$

The entries of $\widehat{\gamma}$ give the final coefficient estimates $\widehat{\phi}_i$ and $\widehat{\theta}_j^*$. As in Step 2, the common scalar GLS weight cancels from the coefficient estimate.

Using the final coefficients, we reconstruct the innovations as

$$\widehat{\eta}_t = y_t^* - \sum_{i=1}^{\min\{p,t-1\}} \widehat{\phi}_i y_{t-i}^* + \sum_{j=1}^{\min\{m,t-1\}} \widehat{\theta}_j^* \widehat{\eta}_{t-j}, \qquad t = 1, \ldots, T,$$

with unavailable presample values set to zero. The final innovation variance is estimated by

$$\widehat{\sigma}_\eta^2 = \frac{1}{T-m} \sum_{t=m+1}^{T} \widehat{\eta}_t^2.$$

When $p = m = 0$, no AR or MA coefficients are estimated: $\widehat{\eta}_t = y_t^*$ and $\widehat{\sigma}_\eta^2 = T^{-1} \sum_{t=1}^{T} (y_t^*)^2$.

The final reduced-form estimates $(\widehat{\phi}_1, \ldots, \widehat{\phi}_p, \widehat{\theta}_1^*, \ldots, \widehat{\theta}_m^*, \widehat{\sigma}_\eta^2)$ are converted into the structural parameters $(\widehat{\theta}_1, \ldots, \widehat{\theta}_q, \widehat{\sigma}_v^2)$ using the covariance-matching relations in (2.12). The scale

parameter is recovered from

$$\widehat{\sigma}_y^2 = \exp(\widehat{\mu} - c_z), \qquad \widehat{\mu} = \frac{1}{T}\sum_{t=1}^{T} \log(y_t^2),$$

where $c_z = E[\log(z_t^2)]$ is defined in the framework.

4 Asymptotic distributional theory

We study the three estimation steps for the observable weak ARMA(p, m) representation, where $m = \max\{p, q\}$ under the maintained noncancellation assumption, with $\sigma_y > 0$ and $\sigma_v^2 > 0$. The results concern fixed, correctly specified orders and fixed parameter values. The first two steps provide sufficiently accurate starting estimates; the third step yields the distribution used for reduced-form coefficient inference. We retain the three-step organization of Dufour and Pelletier (2022) and verify the required bounds for the SV-induced process. Throughout this subsection, $U_t := \eta_t$ denotes the reduced-form innovation.

Assumption 4.1 (Reduced-form stability and invertibility)**.** The reduced-form ARMA(p, m) representation of y_t^* is minimal, causal, and invertible, with weak-white-noise innovations η_t and $\sigma_\eta^2 > 0$.

Assumption 4.2 (Stationarity and moments)**.** The joint process $(y_t^*, \eta_t)'$ is strictly stationary and ergodic, and $E|\eta_t|^{4+2\zeta} < \infty$ for some $\zeta > 0$.

Assumption 4.3 (Innovation dependence)**.** For the same $\zeta > 0$, the strong-mixing coefficients of $\{\eta_t\}$ satisfy

$$\sum_{h=1}^{\infty} \alpha_\eta(h)^{\zeta/(2+\zeta)} < \infty.$$

where $\alpha_\eta(h)$ is the $\alpha-$mixing coefficient of order h.

These reduced-form conditions are verified from the maintained SV framework in Appendix B.1. In particular,

$$\eta_t = \theta^*(B)^{-1}\{\theta(B)v_t + \phi(B)\epsilon_t\}.$$

Thus η_t is generally an infinite causal filter of the primitive innovations. Its coefficients decay geometrically, and the maintained Gaussian and log-χ_1^2 primitives imply finite moments of every positive order. The mixing condition is verified through a stable vector-autoregressive representation with absolutely continuous i.i.d. driving innovations. The proofs below also use this primitive structure to obtain bounds uniform in lag indices.

To state the first-step requirement precisely, let

$$Q_T(c,\pi) := \frac{1}{T-n_T}\sum_{t=n_T+1}^{T}\left(y_t^* - c - \sum_{j=1}^{n_T}\pi_j y_{t-j}^*\right)^2.$$

The first-step fit is required to satisfy

$$Q_T(\widehat{c},\widehat{\Pi}^{(n_T)}) - \inf_{c,\pi} Q_T(c,\pi) = O_p(n_T/T). \tag{4.1}$$

This condition holds automatically for an OLS fit. It also permits other first-step estimators whose least-squares discrepancy satisfies the stated bound. In particular, application of these results to the Gaussian quasi-likelihood implementation requires an additional accuracy condition. A conditional Gaussian criterion that profiles an unrestricted common variance and uses the same observations and unrestricted regression parameters has the same coefficient minimizers as OLS.

For $p+m>0$, define the population regressor rows

$$\begin{aligned} Z_{t-1} &= [y_{t-1}^*,\dots,y_{t-p}^*,-\eta_{t-1},\dots,-\eta_{t-m}],\\ V_{t-1} &= \theta^*(B)^{-1}Z_{t-1}. \end{aligned} \tag{4.2}$$

The inverse filter in the second line uses the true reduced-form MA coefficients and acts componentwise. Put

$$\begin{aligned} s_{2t} &= Z_{t-1}'(\sigma_\eta^2)^{-1}\eta_t, & J_{(2)} &= E[Z_{t-1}'(\sigma_\eta^2)^{-1}Z_{t-1}],\\ s_{3t} &= V_{t-1}'(\sigma_\eta^2)^{-1}\eta_t, & J_{(3)} &= E[V_{t-1}'(\sigma_\eta^2)^{-1}V_{t-1}],\\ I_{(k)} &= \sum_{j=-\infty}^{\infty} E[s_{kt}s_{k,t-j}'], & k &= 2,3. \end{aligned} \tag{4.3}$$

These scores have zero means. Minimality and $\sigma_\eta^2>0$ imply that $J_{(2)}$ and $J_{(3)}$ are positive definite; the long-run covariance series are absolutely convergent under the maintained SV framework. All inverse-variance weights are scalar, consistent with the estimation formulas in the preceding subsection.

Throughout, filtering recursions are initialized as specified in the estimation procedure. Stability and invertibility of the true reduced form, together with consistency of the preliminary estimators, ensure that the estimated filters are asymptotically well behaved. The corresponding local filter regularity and nonsingularity arguments are given in Appendix B.2.

Proposition 4.1 (First-step estimates)**.** *Under the maintained SV assumptions and* (4.1)*, suppose* $\frac{n_T}{\log T}\longrightarrow\infty$*,* $\frac{n_T^2}{T}\longrightarrow 0$ *as* $T\longrightarrow\infty$*. Let* $\Pi^{(n_T)}=(\Pi_1,\dots,\Pi_{n_T})'$ *contain the first* n_T

coefficients of the infinite autoregression. Then

$$\|\widehat{\Pi}^{(n_T)} - \Pi^{(n_T)}\|_2 = O_p\Big(\sqrt{n_T/T}\Big), \tag{4.4}$$

$$\sum_{j=1}^{n_T} |\widehat{\Pi}_j^{(n_T)} - \Pi_j| = O_p\Big(n_T/\sqrt{T}\Big). \tag{4.5}$$

Moreover, $\widehat{c} = O_p(\sqrt{n_T/T})$ *for population-centered data, and the first-step residuals satisfy*

$$\frac{1}{T}\sum_{t=n_T+1}^{T} (\widehat{U}_t - \eta_t)^2 = O_p(n_T/T). \tag{4.6}$$

Proposition 4.2 (Second-step estimates)**.** *Under the assumptions of Proposition 4.1, the second-step estimator in (3.16) satisfies*

$$\widetilde{\gamma} \xrightarrow{p} \gamma, \qquad \|\widetilde{\gamma} - \gamma\|_2 = O_p\Big(\sqrt{n_T/T}\Big) = o_p(T^{-1/4}). \tag{4.7}$$

Also, $\widehat{\sigma}^2_{\eta,1} \xrightarrow{p} \sigma^2_\eta$ *and*

$$\widehat{J}_{(2)} := \frac{1}{T}\sum_{t=\ell}^{T} \widehat{Z}'_{t-1}(\widehat{\sigma}^2_{\eta,1})^{-1}\widehat{Z}_{t-1} \xrightarrow{p} J_{(2)}, \qquad \ell = n_T + m + 1. \tag{4.8}$$

If, in addition, the first-step fit satisfies the score-equivalence condition

$$\mathcal{R}_{2T} := \frac{1}{\sqrt{T}}\sum_{t=\ell}^{T} \widehat{Z}'_{t-1}(\sigma^2_\eta)^{-1}(Z_{t-1} - \widehat{Z}_{t-1})\gamma = o_p(1), \tag{4.9}$$

then the oracle-form second-step distribution is

$$\sqrt{T}(\widetilde{\gamma} - \gamma) \xrightarrow{d} N\Big(0, J_{(2)}^{-1} I_{(2)} J_{(2)}^{-1}\Big). \tag{4.10}$$

For $\omega(j, m_T) = 1 - |j|/(m_T + 1)$*, define* $\widehat{s}_{2t} := \widehat{Z}'_{t-1}(\widehat{\sigma}^2_{\eta,1})^{-1}\widetilde{U}_t$*, where* $\widetilde{U}_t$ *is the recursively reconstructed innovation from (3.17). If* $m_T \to \infty$*,* $m_T^4/T \to 0$ *and* $m_T n_T/\sqrt{T} \to 0$*, then*

$$\widehat{I}_{(2)} := \frac{1}{T}\sum_{j=-m_T}^{m_T} \omega(j, m_T) \sum_{t=\ell+\max\{j,0\}}^{T+\min\{j,0\}} \widehat{s}_{2t}\widehat{s}'_{2,t-j} \xrightarrow{p} I_{(2)}. \tag{4.11}$$

Consequently, under (4.9), $\widehat{J}_{(2)}^{-1}\widehat{I}_{(2)}\widehat{J}_{(2)}^{-1}$ *consistently estimates the covariance matrix in (4.10).*

Condition (4.9) is an additional condition on the first-step implementation, not a consequence asserted here of the SV assumptions or of (4.1). It is needed only for the displayed oracle-form second-step distribution. Without this condition, the preliminary rate (4.7) still holds and is sufficient for the third-step result below.

Proposition 4.3 (Third-step estimates)**.** *Under the assumptions of Proposition 4.1, the*

third-step estimator in (3.21) *is consistent and*

$$\sqrt{T}(\widehat{\gamma}-\gamma) \xrightarrow{d} N\Big(0, J_{(3)}^{-1} I_{(3)} J_{(3)}^{-1}\Big). \tag{4.12}$$

Furthermore, $\widetilde{\sigma}_\eta^2 \xrightarrow{p} \sigma_\eta^2$ *and*

$$\widehat{J}_{(3)} := \frac{1}{T}\sum_{t=\ell'}^{T} \widetilde{V}_{t-1}'(\widetilde{\sigma}_\eta^2)^{-1}\widetilde{V}_{t-1} \xrightarrow{p} J_{(3)}, \qquad \ell' = m+1. \tag{4.13}$$

Let $\bar{U}_t := \widehat{\eta}_t$ *be the innovation reconstructed with the final coefficients, as defined in the estimation method, and put* $\widehat{s}_{3t} := \widetilde{V}_{t-1}'(\widetilde{\sigma}_\eta^2)^{-1}\bar{U}_t$. *If* $m_T \to \infty$ *and* $m_T^4/T \to 0$, *then*

$$\widehat{I}_{(3)} := \frac{1}{T}\sum_{j=-m_T}^{m_T} \omega(j, m_T) \sum_{t=\ell'+\max\{j,0\}}^{T+\min\{j,0\}} \widehat{s}_{3t}\widehat{s}_{3,t-j}' \xrightarrow{p} I_{(3)}, \tag{4.14}$$

where $\omega(j, m_T) = 1 - |j|/(m_T+1)$. *Hence* $\widehat{J}_{(3)}^{-1}\widehat{I}_{(3)}\widehat{J}_{(3)}^{-1}$ *consistently estimates the covariance matrix in* (4.12). *An estimator of the sampling covariance of* $\widehat{\gamma}$ *is this matrix divided by* T.

The proof preserves the three-step argument discussed in Dufour and Pelletier (2022). Step 1 shows that as the preliminary autoregressive order n_T increases at the stated rate, the approximation error from estimating a finite order AR when the truth is $AR(\infty)$ will shrink to zero with the sample size. Step 2 provides an $o_p(T^{-1/4})$ starting estimate. The regression coefficients are consistently estimated because the generated-regressor error from using residual proxies vanishes asymptotically. Replacing the innovation U_t with a residual induces an MA structure in the measurement error. Step 3 thus uses the MA filter to form a one-step least-squares correction: the first-order preliminary-estimation error cancels, leaving a quadratic remainder of order $O_p(n_T/T) = o_p(T^{-1/2})$. The conclusions for the coefficient vector are unchanged when the population mean of the log-squared observations is replaced by its sample mean, as verified in the appendix. When $p = m = 0$, the coefficient vector is empty and the variance-only procedure in the estimation method applies instead.

5 Model selection

This section develops a two-stage procedure for selecting the structural orders (p, q) and evaluates its finite-sample performance. We maintain Assumption 2.4. By Proposition 2.2, the centered log-squared process has a minimal ARMA(p, m) representation, $\phi(B)y_t^* = \theta^*(B)\eta_t, m = \max\{p, q\}$, where $\{\eta_t\}$ is weak white noise with variance $\sigma_\eta^2 > 0$. In implementation, the population mean μ is replaced by $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\log(y_t^2)$.

Order selection proceeds by first selecting the reduced-form orders (p, m) and then recovering the structural MA order q. If $m > p$, the order relation implies $q = m$. If $m = p$,

the reduced-form orders alone do not determine q, and candidate values $q \in \{0, \dots, p\}$ are compared using the structural autocovariance restrictions.

First, we select the reduced-form orders (p, m). Let (P, M) be fixed upper bounds and define the admissible set $\mathcal{S}_{P,M} := \{(p, m) : 0 \le p \le P,\ p \le m \le M\}$. For each $(p, m) \in \mathcal{S}_{P,M}$, we estimate an ARMA$(p, m)$ model for y_t^* using the three-step procedure in Hannan and Rissanen (1982). Let $\hat{\sigma}_\eta^2(p, m)$ denote the resulting reduced-form innovation-variance estimate. A candidate (p, m) is treated as admissible if: (i) $\hat{\sigma}_\eta^2(p, m) > 0$, (ii) the estimated AR polynomial is stable, and (iii) the estimated MA polynomial is invertible. Among admissible candidates, we select $(\hat{p}, \hat{m})$ by minimizing

$$IC_T(p, m) := \log\left(\hat{\sigma}_\eta^2(p, m)\right) + (p + m)\frac{(\log T)^{1+\delta_{\mathrm{IC}}}}{T} \qquad (p, m) \in \mathcal{S}_{P,M} \tag{5.1}$$

for some $\delta_{\mathrm{IC}} > 0$. We adopt the strengthened penalty rate proposed by Dufour and Pelletier (2022). The original Hannan–Rissanen criterion used a penalty proportional to $\frac{(\log T)^\delta}{T}$ for $\delta > 0$, which need not dominate the approximation error induced by replacing the innovations with residuals from the preliminary autoregression. The modified criterion therefore replaces $(\log T)^\delta$ by $(\log T)^{1+\delta_{\mathrm{IC}}}$. Our implementation evaluates candidate models using the final, third-step Hannan–Rissanen innovation-variance estimate, $\hat{\sigma}_\eta^2(p, m)$, together with this strengthened penalty. Its finite-sample performance is evaluated in the simulations below.

Consistent with the rate conditions, we choose the preliminary AR order as

$$n_T = \left\lceil c(\log T)^{1+\delta_{\mathrm{AR}}} \right\rceil, \qquad 0 < \delta_{\mathrm{AR}} < \delta_{\mathrm{IC}}, \tag{5.2}$$

for a fixed constant $c > 0$. This last condition ensures that n_T grows slowly with T while satisfying the rate restrictions used in the HR-based order-selection argument.

Second, conditional on the selected reduced-form order and estimates $(\hat{p}, \hat{m}, \hat{\phi}, \hat{\theta}^*, \hat{\sigma}_\eta^2)$, we select the structural MA order q. To simplify notation throughout this step, write $p = \hat{p}$ and $m = \hat{m}$. The relationship $m = \max(p, q)$ produces two cases:

- **Case 1:** $m > p$**.** A minimal reduced-form ARMA(p, m) with $m > p$ necessarily implies $q = m$. We therefore set $\hat{q} = m$. No comparison across structural orders is required, and the structural parameters can be estimated using the profile minimum-distance criterion described below.

- **Case 2:** $m = p$**.** When $m = p > 0$, the reduced-form order alone does not identify q, because every $q \in \{0, \dots, m\}$ is potentially compatible with an ARMA(p, p) reduced form. We therefore compare these candidate structural orders using the complete autocovariance-matching system. If $m = p = 0$, we set $\hat{q} = 0$ directly.

Let $(\hat{\phi}_1, \dots, \hat{\phi}_p)$, $(\hat{\theta}_1^*, \dots, \hat{\theta}_m^*)$, and $\hat{\sigma}_\eta^2$ denote the selected reduced-form estimates, and

treat σ_ϵ^2 as known. For $k = 0, 1, \ldots, m$, define the adjusted reduced-form moments

$$\hat{C}_k := \hat{\sigma}_\eta^2 \, \hat{\gamma}_{\theta^*}(k) - \sigma_\epsilon^2 \, \hat{\gamma}_\phi(k), \tag{5.3}$$

where $\hat{\gamma}_{\theta^*}(k)$ and $\hat{\gamma}_\phi(k)$ are the coefficients of $\hat{\theta}^*(B)\hat{\theta}^*(B^{-1})$ and $\hat{\phi}(B)\hat{\phi}(B^{-1})$, respectively.[1]

For a structural MA polynomial $\theta_q(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$, define

$$\gamma_\theta(0; \theta_q) = 1 + \sum_{j=1}^{q} \theta_j^2, \qquad \gamma_\theta(k; \theta_q) = -\theta_k + \sum_{j=1}^{q-k} \theta_j \theta_{j+k}, \quad 1 \le k \le q, \tag{5.4}$$

and set $\gamma_\theta(k; \theta_q) = 0$ for $k > q$. The structural autocovariance-matching equations in (2.12) imply

$$C_k = \sigma_v^2 \gamma_\theta(k; \theta_q), \qquad k = 0, 1, \ldots, m. \tag{5.5}$$

In particular, order q implies $C_{q+1} = \cdots = C_m = 0$. We therefore select q by measuring how far the estimated adjusted-moment vector $\hat{C} = (\hat{C}_0, \ldots, \hat{C}_m)'$ is from the moment vectors allowed by each candidate structural model.

Because the elements of $\hat{C}$ are nonlinear functions of estimated reduced-form parameters, they can have different sampling variances and can be correlated across lags. We approximate the covariance matrix of $\sqrt{T}(\hat{C} - C)$ using a residual circular block bootstrap, holding the selected reduced-form order (p, m) fixed. The bootstrap resamples short blocks of centered innovations reconstructed from the fitted ARMA model, generates a new series through the fitted ARMA recursion, and re-estimates the same reduced-form specification in each bootstrap sample. Block resampling retains within-block dependence in the reconstructed innovations, while re-estimation propagates reduced-form parameter uncertainty into the adjusted moments. Holding (p, m) fixed means that this covariance estimate does not incorporate uncertainty from reduced-form order selection. The use of circular blocks avoids treating residuals near the sample endpoints differently; see Künsch (1989) and Politis and Romano (1992).

Let $\hat{C}^{*(b)}$ denote the adjusted-moment vector from valid bootstrap replication b, and let

$$\bar{C}^* = \frac{1}{B} \sum_{b=1}^{B} \hat{C}^{*(b)}.$$

[1] Equivalently, for $k \ge 1$,

$$\hat{\gamma}_{\theta^*}(k) = -\hat{\theta}_k^* + \sum_{j=1}^{m-k} \hat{\theta}_j^* \hat{\theta}_{j+k}^*, \qquad \hat{\gamma}_\phi(k) = -\hat{\phi}_k + \sum_{j=1}^{p-k} \hat{\phi}_j \hat{\phi}_{j+k},$$

where coefficients beyond the corresponding polynomial order are defined as zero. For $k = 0$, $\hat{\gamma}_{\theta^*}(0) = 1 + \sum_{j=1}^{m} \hat{\theta}_j^{*2}$, $\hat{\gamma}_\phi(0) = 1 + \sum_{j=1}^{p} \hat{\phi}_j^2$.

We estimate the sampling covariance of $\sqrt{T}\hat{C}$

$$\hat{\Omega}_C = \frac{T}{B-1}\sum_{b=1}^{B}\left(\hat{C}^{*(b)} - \bar{C}^*\right)\left(\hat{C}^{*(b)} - \bar{C}^*\right)'. \tag{5.6}$$

Here, $\hat{\Omega}_C$ captures both the sampling variability of the adjusted moments and the dependence among them. This weighting is important because the moments at different lags need not be equally precise and can be strongly correlated. In implementation, we use a symmetric, numerically stabilized inverse of $\hat{\Omega}_C$ when its smallest eigenvalues are close to zero.

For candidate q, let $\beta_q = (\theta_1, \ldots, \theta_q, \sigma_v^2)'$ collect the structural MA coefficients and innovation variance, and define its model-implied moment vector by

$$c_q(\beta_q) := \sigma_v^2\left(\gamma_\theta(0;\theta_q) \quad \gamma_\theta(1;\theta_q) \quad \cdots \quad \gamma_\theta(m;\theta_q)\right)'. \tag{5.7}$$

For structural-order comparison, candidate q allows an MA polynomial of degree at most q, so trailing zero coefficients are permitted. Let $\mathcal{B}_q$ denote the parameter space satisfying $\sigma_v^2 > 0$, invertibility of $\theta_q(B)$, and no common roots with the selected AR polynomial. The true structural order is the smallest admissible degree. We profile out the structural parameters for each candidate order:

$$\hat{\beta}(q) \in \arg\min_{\beta_q \in \mathcal{B}_q}\left\{\left[\hat{C} - c_q(\beta_q)\right]'\hat{\Omega}_C^{-1}\left[\hat{C} - c_q(\beta_q)\right]\right\}. \tag{5.8}$$

The associated lack-of-fit statistic is

$$J_T(q) = T\left[\hat{C} - c_q\big(\hat{\beta}(q)\big)\right]'\hat{\Omega}_C^{-1}\left[\hat{C} - c_q\big(\hat{\beta}(q)\big)\right], \qquad q = 0, \ldots, m. \tag{5.9}$$

The statistic $J_T(q)$ is the covariance-standardized distance between the estimated adjusted-moment vector and the closest moment vector that can be generated by an admissible SV(p, q) model. A smaller q imposes more structure on the complete autocovariance pattern and may therefore produce a larger discrepancy. A larger q provides greater flexibility, but is penalized for the additional structural parameters. Accordingly, in Case 2 we select

$$\hat{q} = \arg\min_{0 \le q \le m}\left\{J_T(q) + (q+1)\kappa(\log T)^{1+\Delta_{str}}\right\}. \tag{5.10}$$

The term $q + 1$ counts the structural MA coefficients and the innovation variance.

The tuning constants $\kappa > 0$ and $\Delta_{str} \ge 0$ are fixed and common across candidate orders. Our baseline uses $\kappa = 1$ and $\Delta_{str} = 0$, giving the penalty $(q+1)\log T$. Since all candidates use the same $m + 1$ adjusted moments, this penalty is equivalent, up to a common additive constant, to rewarding the number of overidentifying restrictions as in Andrews and Lu (2001). When $q = m$, the candidate model has $m+1$ structural parameters and is exactly identified by the $m + 1$ adjusted moments. If the estimated moments admit an interior admissible MA(m)

Table 1: Model-selection frequencies for structural SV(p, q) models.

Scenario	True model	Parameters	Selected structural order (proportions over 500 replications)								
			$(0,0)$	$(0,1)$	$(0,2)$	$(1,0)$	$(1,1)$	$(1,2)$	$(2,0)$	$(2,1)$	$(2,2)$
S1	SV$(0,1)$	$\theta_1 = 0.6$	0.000	**0.988**	0.012	0.000	0.000	0.000	0.000	0.000	0.000
S2	SV$(0,1)$	$\theta_1 = -0.9$	0.000	**0.988**	0.012	0.000	0.000	0.000	0.000	0.000	0.000
S3	SV$(0,2)$	$(\theta_1, \theta_2) = (0.5, -0.25)$	0.000	0.000	**0.944**	0.048	0.008	0.000	0.000	0.000	0.000
S4	SV$(1,0)$	$\phi_1 = 0.7$	0.000	0.000	0.000	**0.992**	0.002	0.006	0.000	0.000	0.000
S5	SV$(1,0)$	$\phi_1 = 0.95$	0.000	0.000	0.000	**0.974**	0.010	0.016	0.000	0.000	0.000
S6	SV$(1,1)$	$(\phi_1, \theta_1) = (0.7, 0.3)$	0.000	0.000	0.000	0.018	**0.982**	0.000	0.000	0.000	0.000
S7	SV$(1,2)$	$(\phi_1, \theta_1, \theta_2) = (-0.6, 1.3, -0.5)$	0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
S8	SV$(2,0)$	$(\phi_1, \phi_2) = (1.3, -0.6)$	0.000	0.000	0.000	0.000	0.000	0.000	**0.996**	0.004	0.000
S9	SV$(2,1)$	$(\phi_1, \phi_2, \theta_1) = (0.4, -0.6, 0.9)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.998**	0.002
S10	SV$(2,2)$	$(\phi_1, \phi_2, \theta_1, \theta_2) = (0.8, -0.6, 1.3, -0.5)$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**

Notes: $T = 10{,}000$ and 500 Monte Carlo replications per scenario. Entries are selection proportions; bold entries indicate the true structural order. For the reported correct-selection frequencies, the largest plug-in Monte Carlo standard error is approximately 0.010, or one percentage point. Simulations are conducted directly for the centered log-squared process y_t^*, with $\sigma_v^2 = 2.5$, $z_t \sim N(0,1)$, and $\sigma_\epsilon^2 = \pi^2/2$. The return-scale parameter σ_y drops out after centering. A burn-in of 1,000 observations is used.

The reduced-form order is selected over $0 \le p \le 2$ and $p \le m \le 2$ using the HR-type criterion with $c = 1$, $\delta_{\mathrm{IC}} = 0.1$, $\delta_{\mathrm{AR}} = 0.05$, $n_T = \lceil c(\log T)^{1+\delta_{\mathrm{AR}}} \rceil$, and penalty $(\log T)^{1+\delta_{\mathrm{IC}}}/T$.

When the selected reduced-form order satisfies $\hat{m} > \hat{p}$, the structural order is set to $\hat{q} = \hat{m}$. When $\hat{m} = \hat{p}$, candidate structural orders are compared using the bootstrap-weighted profile criterion in (5.10), with penalty $(q+1)\log T$. The residual circular block bootstrap uses $B = 199$ valid replications and block length $L = 5$.

All reduced-form and structural orders are resolved, and all selected parameter profiles converge. Complete reduced-form and structural selection-frequency matrices are reported in Appendix Table D.1.

representation, the profiled criterion therefore attains $J_T(m) = 0$ up to numerical tolerance. Otherwise, the constraints defining $\mathcal{B}_m$ can result in a positive constrained minimum.

Table 1 reports the estimated selection probabilities for the structural SV(p, q) order. At $T = 10{,}000$, the empirical frequency of selecting the correct structural order ranges from 94.4% to 100% of the replications across the ten designs, with an average correct-selection rate of 98.6%. The reduced-form stage is similarly accurate: the correct (p, m) order is selected in 94.4% to 100% of the replications, with an average correct-selection rate of 99.0%.

When the selected reduced-form order satisfies $\hat{m} > \hat{p}$, the structural order follows directly from $\hat{q} = \hat{m}$. When $\hat{m} = \hat{p}$, several structural MA orders are compatible with the same reduced-form specification, and the bootstrap-weighted profile criterion is required to distinguish among them. The correct structural specification is selected in 97.4% to 100% of the replications in these designs.

All 5,000 replications produce resolved reduced-form and structural orders, and the selected parameter profiles converge in every case. The complete reduced-form and structural selection-frequency matrices are reported in Appendix Table D.1.

6 Simulation

This section evaluates the finite-sample performance of the estimators using Monte Carlo simulations implemented in MATLAB. For each design and sample size, we generate $N = 2000$ independent datasets from the same data-generating process. The structural orders are fixed at their data-generating values. A replication is successful when the estimator returns finite real parameter estimates satisfying the required stationarity, invertibility, and positive-variance restrictions. Let $\mathcal{S}$ be the set of successful replications for a given estimator, design, sample size, and tuning choice, and let $N_s = |\mathcal{S}|$. For each scalar structural parameter $\vartheta \in \{\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma_v^2, \sigma_y\}$, we report

$$\widehat{\text{Bias}} = \frac{1}{N_s} \sum_{r \in \mathcal{S}} \left(\widehat{\vartheta}^{(r)} - \vartheta \right), \qquad \widehat{\text{RMSE}} = \left\{ \frac{1}{N_s} \sum_{r \in \mathcal{S}} \left(\widehat{\vartheta}^{(r)} - \vartheta \right)^2 \right\}^{1/2}. \tag{6.1}$$

Here $\widehat{\vartheta}^{(r)}$ is an estimate from replication r, and RMSE denotes root mean squared error. Both measures are conditional on successful estimation. Success counts are reported in the appendix.

6.1 Main results for SV$(1, 1)$

We compare the simple ARMA-based estimator, denoted S-ARMA-based, four ARMA-Winsorized estimators, and the Hannan–Rissanen estimator, denoted HR. The winsorized estimators combine the autocovariance-based estimates using the arithmetic mean, linearly declining weights, the componentwise median, or ordinary least squares. We denote these procedures S-ARMA-W-Mean, S-ARMA-W-LDMean, S-ARMA-W-Median, and S-ARMA-W-OLS, respectively. The main tables report the four winsorized estimators with $J = 5$ lag blocks and the HR estimator with preliminary autoregressive order $n_T = 8$. Appendix Tables E.1 and E.2 report the complete tuning grids, together with the restricted S-ARMA estimator, the simple moment estimator S-Moment, and success counts.

We consider two SV$(1, 1)$ designs. Design 1 sets $(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.7, 0.1, 2.5, 2.5)$, while Design 2 sets $(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.9, -0.3, 2.5, 2.5)$. We report results for $T \in \{1000, 5000\}$ to assess how finite-sample accuracy changes as more observations become available. The log-squared estimators recover the return-scale parameter using

$$\widehat{\sigma}_y = \exp\left(\frac{\widehat{\mu} - c_z}{2} \right), \qquad \widehat{\mu} = \frac{1}{T} \sum_{t=1}^{T} \log(y_t^2),$$

where $c_z = E[\log(z_t^2)] \simeq -1.2704$ under Gaussian return innovations. In finite samples, bias or RMSE for σ_y can be sensitive to admissibility screening and winsorization, especially at

smaller T; this sensitivity diminishes as T increases.

Table 2 reports the results for Design 1. RMSE decreases for every displayed estimator and parameter when sample size T increases from 1000 to 5000. HR and S-ARMA-W-OLS provide the most accurate estimates of the dynamic coefficients. At $T = 1000$, they have the same reported RMSE of 0.0690 for ϕ_1, while HR has the lowest RMSE for θ_1 and σ_v^2. At $T = 5000$, HR has the lowest RMSE for all three structural dynamic parameters. Overall, the gains of HR in this design arise primarily through variance reduction rather than systematic bias corrections.

Table 2: Monte Carlo performance in $SV(1,1)$ (Design 1)

DGP: $(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.7, 0.1, 2.5, 2.5)$; $N = 2000$ simulated datasets								
	$T = 1000$				$T = 5000$			
	ϕ_1	θ_1	σ_v^2	σ_y	ϕ_1	θ_1	σ_v^2	σ_y
Panel A. Bias								
S-ARMA-based	-0.003	-0.026	-0.055	0.016	0.000	-0.007	-0.020	0.001
S-ARMA-W-Mean ($J = 5$)	0.015	0.010	-0.066	0.018	0.000	-0.006	-0.021	0.001
S-ARMA-W-LDMean ($J = 5$)	0.007	-0.012	-0.051	0.018	0.000	-0.006	-0.016	0.001
S-ARMA-W-Median ($J = 5$)	-0.012	-0.039	-0.044	0.017	-0.002	-0.010	-0.017	0.001
S-ARMA-W-OLS ($J = 5$)	-0.017	-0.052	-0.030	0.017	-0.003	-0.012	-0.015	0.001
HR ($n_T = 8$)	-0.018	-0.050	-0.029	0.005	-0.002	-0.009	-0.014	0.002
Panel B. RMSE								
S-ARMA-based	0.0988	0.2407	0.6144	0.2119	0.0430	0.1021	0.2827	**0.0909**
S-ARMA-W-Mean ($J = 5$)	0.1204	0.2539	0.6169	0.2109	0.0498	0.1057	0.2864	**0.0909**
S-ARMA-W-LDMean ($J = 5$)	0.0848	0.2146	0.6184	0.2127	0.0324	0.0820	0.2833	**0.0909**
S-ARMA-W-Median ($J = 5$)	0.0999	0.2288	0.6214	0.2120	0.0413	0.0930	0.2838	**0.0909**
S-ARMA-W-OLS ($J = 5$)	**0.0690**	0.1979	0.6218	0.2119	0.0288	0.0797	0.2834	**0.0909**
HR ($n_T = 8$)	**0.0690**	**0.1924**	**0.5987**	**0.2081**	**0.0283**	**0.0758**	**0.2787**	0.0927

Notes. Bias and RMSE are computed conditional on successful estimation. For each sample size, $N = 2000$ datasets are simulated; the number of successful replications may differ across estimators. Bold entries indicate the smallest RMSE among the displayed estimators at the reported precision, including ties. Success counts and the complete tuning grid are reported in Appendix Table E.1.

Table 3 reports Design 2. Estimation is more difficult than in Design 1, particularly for σ_v^2 and σ_y, and the appendix shows a lower frequency of admissible estimates at $T = 1000$. Nevertheless, RMSE decreases substantially as the sample size increases. HR has the lowest RMSE for ϕ_1 and σ_v^2 at both sample sizes and for θ_1 at $T = 1000$. At $T = 5000$, S-ARMA-W-OLS and HR give nearly identical RMSE for θ_1. S-ARMA-W-Median has the lowest reported RMSE for σ_y at $T = 5000$. Thus HR remains particularly effective for estimating the latent volatility dynamics, although it does not dominate every parameter in every design.

Appendix Tables E.1 and E.2 report the complete tuning grids and success counts. The HR

Table 3: Monte Carlo performance in $SV(1,1)$ (Design 2)

DGP: $(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.9, -0.3, 2.5, 2.5)$; $N = 2000$ simulated datasets								
	$T = 1000$				$T = 5000$			
	ϕ_1	θ_1	σ_v^2	σ_y	ϕ_1	θ_1	σ_v^2	σ_y
Panel A. Bias								
S-ARMA-based	-0.0016	0.0516	0.3131	0.1437	-0.0005	-0.0220	-0.0083	0.0195
S-ARMA-W-Mean ($J = 5$)	-0.0030	0.0442	0.3211	0.1602	-0.0009	-0.0274	-0.0171	0.0194
S-ARMA-W-LDMean ($J = 5$)	-0.0033	0.0459	0.3303	0.1536	-0.0009	-0.0271	-0.0165	0.0199
S-ARMA-W-Median ($J = 5$)	-0.0029	0.0480	0.3272	0.1577	-0.0009	-0.0273	-0.0173	0.0195
S-ARMA-W-OLS ($J = 5$)	-0.0035	0.0423	0.3234	0.1562	-0.0009	-0.0261	-0.0139	0.0191
HR ($n_T = 8$)	-0.0039	0.0352	0.2978	0.1119	-0.0004	-0.0217	-0.0049	0.0270
Panel B. RMSE								
S-ARMA-based	0.0202	0.2264	0.8614	0.8763	0.0088	0.1606	0.4968	0.3669
S-ARMA-W-Mean ($J = 5$)	0.0193	0.2153	0.8665	0.8842	0.0084	0.1558	0.4964	0.3665
S-ARMA-W-LDMean ($J = 5$)	0.0185	0.2089	0.8605	0.8789	0.0079	0.1564	0.4980	0.3665
S-ARMA-W-Median ($J = 5$)	0.0197	0.2143	0.8630	0.8780	0.0086	0.1565	0.4958	**0.3658**
S-ARMA-W-OLS ($J = 5$)	0.0188	0.2125	0.8637	0.8788	0.0080	**0.1535**	0.4941	0.3661
HR ($n_T = 8$)	**0.0181**	**0.2046**	**0.8406**	**0.8762**	**0.0077**	0.1539	**0.4906**	0.3710

Notes. Bias and RMSE are computed conditional on successful estimation. For each sample size, $N = 2000$ datasets are simulated; the number of successful replications may differ across estimators. Bold entries indicate the smallest RMSE among the displayed estimators at the reported precision, including ties. Success counts and the complete tuning grid are reported in Appendix Table E.2.

results change little across $n_T \in \{3, 8, 10\}$ in these designs, whereas the effect of increasing J varies across the winsorized estimators. The appendix also reports S-Moment, which produces fewer admissible estimates and substantially larger scale errors in Design 2. Finally, Table E.3 evaluates the estimators under standardized Student-t return innovations using the corresponding distribution-specific constants.

6.2 Main results for SV(p, q)

Table 4 evaluates HR's performance in a purely autoregressive model, a purely moving-average model, and a higher-order SV$(2, 2)$ model. The specifications are

$$\begin{aligned}
&\text{M(1): SV}(1,0), \quad \phi(B) = 1 - 0.8B, && \theta(B) = 1,\\
&\text{M(2): SV}(0,1), \quad \phi(B) = 1, && \theta(B) = 1 - 0.5B,\\
&\text{M(3): SV}(2,2), \quad \phi(B) = 1 - 0.7B + 0.3B^2, && \theta(B) = 1 - 0.4B - 0.2B^2.
\end{aligned}$$

Each design uses $\sigma_v^2 = \sigma_y = 2.5$, $N = 2000$ simulated datasets, and $n_T = 8$. Panel A reports bias and Panel B reports RMSE for $T = 1000$ and $T = 5000$. Both panels use only successful

replications. In SV$(1,0)$, the RMSE of $\widehat{\phi}_1$ falls from 0.0374 to 0.0153 as T increases. In SV$(0,1)$, the RMSE of $\widehat{\theta}_1$ falls from 0.1554 to 0.0876, while the success rate rises from 86.80% to 99.45%. The corresponding bias is small at both sample sizes. In the higher-order SV$(2,2)$ design, the RMSE of each dynamic coefficient falls by more than half between the two sample sizes, and the success rate rises from 74.75% to 99.35%.

Appendix Table E.4 compares $n_T = 3, 8, 10$ for SV$(1,0)$ and SV$(0,1)$ at $T = 5000$, and $n_T = 8, 10$ for SV$(2,2)$. The reported results change little over these choices. Table E.5 varies θ_1 in SV$(0,1)$ and extends the sample-size comparison to $T = 10000$. Smaller MA coefficients yield higher success rates in these designs, and conditional RMSE decreases with sample size.

Table 4: HR performance across selected SV(p,q) models (fixed $n_T = 8$)

	$T = 1000$			$T = 5000$		
	M(1)	M(2)	M(3)	M(1)	M(2)	M(3)
Panel A. Bias						
ϕ_1	-0.0081	–	-0.0554	-0.0022	–	-0.0113
ϕ_2	–	–	-0.0354	–	–	-0.0014
θ_1	–	0.0008	-0.0794	–	0.0128	-0.0174
θ_2	–	–	-0.0391	–	–	0.0081
σ_v^2	-0.0250	0.0500	-0.0274	-0.0084	-0.0207	-0.0385
σ_y	0.0195	-0.0108	-0.0105	0.0000	-0.0004	0.0004
Panel B. RMSE						
ϕ_1	0.0374	–	0.1985	0.0153	–	0.0773
ϕ_2	–	–	0.1642	–	–	0.0773
θ_1	–	0.1554	0.2907	–	0.0876	0.1243
θ_2	–	–	0.2559	–	–	0.1235
σ_v^2	0.7250	0.5564	0.5499	0.3284	0.2884	0.2732
σ_y	0.3353	0.0930	0.0982	0.1394	0.0423	0.0441
Success	*1999*	*1736*	*1495*	*2000*	*1989*	*1987*

Notes. For each model and sample size, $N = 2000$ datasets are simulated. HR uses preliminary autoregressive order $n_T = 8$. M(1) is SV$(1,0)$ with $\phi_1 = 0.8$; M(2) is SV$(0,1)$ with $\theta_1 = 0.5$; and M(3) is SV$(2,2)$ with $(\phi_1, \phi_2) = (0.7, -0.3)$ and $(\theta_1, \theta_2) = (0.4, 0.2)$. All designs use $\sigma_v^2 = \sigma_y = 2.5$. Bias and RMSE are computed conditional on successful estimation. A dash indicates that the parameter is absent from the model.

7 Empirical application to the S&P 500 index

We illustrate the proposed methodology using the daily closing level of the Standard & Poor's (S&P 500) Composite Price Index obtained from Wharton Research Data Services

(WRDS). The sample spans July 2, 1962 to December 31, 2024, yielding 15,732 daily index observations, from which we construct $T = 15{,}731$ adjacent daily log returns. The sample covers several episodes of elevated volatility, including the 1973 oil crisis, Black Monday in 1987, the 1997 Asian financial crisis, the dot-com bust, the 2008 global financial crisis, the 2014 Russian financial crisis, and the COVID-19 market disruption in 2020. Returns are constructed as $r_t = 100\Big(\log p_t - \log p_{t-1}\Big)$, and we remove the sample mean to obtain residual returns $y_t := r_t - \bar{r}$, where $\bar{r}$ is the sample average of returns. The estimation method is applied to the centered log-squared series $y_t^* = \log(y_t^2) - \frac{1}{T}\sum_{s=1}^{T}\log(y_s^2)$. The log transformation is useful empirically because squared daily returns are extremely right-skewed, whereas $\log(y_t^2)$ is substantially less skewed and heavy-tailed while retaining the serial dependence associated with volatility dynamics. Table 5 summarizes these transformations.

Table 5: Summary statistics for S&P 500 daily returns and volatility transformations

Series	Mean	Std. dev.	Skewness	Kurtosis	Min	Max
y_t	0.00	1.04	-0.99	28.17	-22.93	10.93
y_t^2	1.08	5.61	57.25	4963.10	0.00	525.75
$\log(y_t^2)$	-1.87	2.50	-1.15	5.71	-20.38	6.26
y_t^*	0.00	2.50	-1.15	5.71	-18.51	8.13

Notes: The sample contains $T = 15{,}731$ daily returns constructed from 15,732 index observations. Returns are expressed in percentage points. $r_t = 100(\log p_t - \log p_{t-1})$, $y_t = r_t - \bar{r}$, and $y_t^* = \log(y_t^2) - T^{-1}\sum_{s=1}^{T}\log(y_s^2)$.

7.1 Model selection and the role of moving-average dynamics

We first select the observable reduced-form order. For each admissible ARMA(p, m) specification satisfying $0 \le p \le 5$ and $p \le m \le 5$, we estimate the model by the three-step Hannan–Rissanen procedure and minimize the information criterion in Section 5. This first stage selects $(\hat{p}, \hat{m}) = (1, 1)$. Because $\hat{m} = \hat{p} = 1$, the reduced-form order alone does not distinguish between the structural specifications SV$(1, 0)$ and SV$(1, 1)$: both can generate an ARMA$(1, 1)$ representation for y_t^*. We therefore use the second-stage bootstrap-weighted profile criterion developed in Section 5 to distinguish between $q = 0$ and $q = 1$ using their structural covariance restrictions. We estimate the moment covariance matrix using $B = 199$ residual circular block-bootstrap replications. To assess sensitivity to the treatment of short-range residual dependence, we repeat the calculation using block lengths $L = 1, 5$, and 10. The selected structural order is $\hat{q} = 1$ in all three cases. Bootstrap success rates range from 94.8% to 99.0%, all selected profiles converge, and neither the boundary nor the weak-identification diagnostic is triggered. We therefore use SV$(1, 1)$ as the preferred

specification for the remainder of the empirical analysis. Table 6 summarizes the two-stage selection, while the complete reduced-form criterion grid and structural-selection diagnostics are reported in Appendix F.2.

Table 6: Two-stage model selection for S&P 500 centered log-squared returns

Selection stage	Candidate specifications	Criterion	Empirical result
Reduced-form order	ARMA(p, m), $0 \le p \le 5,\ p \le m \le 5$	HR–IC	$(\hat{p}, \hat{m}) = (1, 1)$
Structural MA order	SV$(1, 0)$ vs. SV$(1, 1)$	Bootstrap profile IC	$\hat{q} = 1$ for $L = 1, 5, 10$

Notes: The first stage selects the reduced-form ARMA order using the Hannan–Rissanen information criterion. Conditional on the selected ARMA$(1, 1)$ reduced form, the second stage compares $q = 0$ and $q = 1$ using the bootstrap-weighted structural profile criterion. The covariance weighting matrix is estimated using residual circular-block bootstrap replications with $B = 199$ valid draws targeted for each block length. Block lengths $L = 1, 5$, and 10 all select SV$(1, 1)$. Detailed criterion values and bootstrap diagnostics are reported in the Appendix.

After selecting $q = 1$, we recover the structural SV$(1, 1)$ parameters from the estimated ARMA$(1, 1)$ reduced form using the first-order backward mapping. For economic comparison, we also fit AR-only SV$(1, 0)$ and SV$(2, 0)$ benchmarks on the same sample.

The economic role of the moving-average term is most transparent in the propagation of a volatility innovation. Let ψ_h denote the response of latent log variance w_{t+h} to an innovation v_t, normalized so that $\psi_0 = 1$. Under SV$(1, 0)$, $\psi_h = \phi_1^h$, so the autoregressive coefficient governs both the initial propagation and the subsequent persistence of the shock. Under SV$(2, 0)$, it is again $\psi_1 = \phi_1$, while subsequent responses satisfy $\psi_h = \phi_1\psi_{h-1} + \phi_2\psi_{h-2}$ for $h \ge 2$. The additional autoregressive lag allows a richer subsequent propagation pattern. Under SV$(1, 1)$, by contrast,

$$\psi_1 = \phi_1 - \theta_1, \qquad \psi_h = (\phi_1 - \theta_1)\phi_1^{h-1}, \quad h \ge 1.$$

The MA coefficient therefore allows the initial adjustment to a volatility innovation to differ from the persistence of the remaining effect. With $\hat{\phi}_1 = 0.9914$ and $\hat{\theta}_1 = 0.8417$, the selected SV$(1, 1)$ specification implies $\hat{\psi}_1 = 0.1498$, so only about 15% of the initial log-variance effect remains one trading day after the shock.

To express these differences in economically interpretable units, we use a common-impact normalization. Specifically, we choose the innovation magnitude

$$\delta = 2\log(1.10),$$

which raises conditional volatility by 10% on impact in every model. Since a shock of size δ changes log variance at horizon h by $\delta\psi_h$, the corresponding percentage increase in conditional volatility is

$$\Delta\sigma_h(\%) = 100\left[\exp\left(\frac{\delta\psi_h}{2}\right) - 1\right] = 100\left(1.10^{\psi_h} - 1\right).$$

Thus the normalization holds the impact magnitude fixed across models, so differences at subsequent horizons reflect only differences in shock propagation. Table 7 reports the resulting paths at selected horizons.

Table 7: Structural estimates and propagation of a common-impact volatility shock

	Structural estimates				Conditional-vol increase after the shock (%)				
Model	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\sigma}_v^2$	Day 1	Day 5	Day 10	Day 20	Day 30
SV(1,0)	0.9914	—	—	0.0181	9.91	9.56	9.14	8.35	7.64
SV(2,0)	0.7490	0.2382	—	0.0312	7.40	7.55	7.16	6.43	5.78
SV(1,1)	0.9914	—	0.8417	0.5756	1.44	1.39	1.33	1.22	1.12

Notes: The first four columns report the estimated structural parameters. The common return-scale estimate is $\hat{\sigma}_y = 0.7412$. The last five columns report the percentage increase in conditional volatility remaining at the indicated horizon after a common-impact shock. The innovation magnitude is normalized to $\delta = 2\log(1.10)$, so that conditional volatility increases by exactly 10% on impact ($h = 0$) in every model. If ψ_h denotes the normalized log-variance response, the reported entries give the model-implied percentage increase in conditional volatility relative to its no-shock path at horizons $h = 1, 5, 10, 20, 30$, calculated as $100(1.10^{\psi_h} - 1)$. SV(1,0) and SV(2,0) are AR-only comparison specifications, whereas SV(1,1) is selected by the two-stage order-selection procedure.

Figure 1 gives two complementary views of this difference. Panel (a) compares the empirical autocorrelation function of the *observed transformed series* y_t^* with the autocorrelation functions implied by the three fitted structural models. Because $y_t^* = w_t + \epsilon_t$, with independent measurement noise of variance $\sigma_\epsilon^2 = \pi^2/2$, and the measurement disturbance is independent of latent volatility,

$$\gamma_{y^*}(0) = \gamma_w(0) + \sigma_\epsilon^2, \qquad \gamma_{y^*}(k) = \gamma_w(k), \quad k \geq 1.$$

Hence the model-implied autocorrelation at lag $k \geq 1$ is

$$\rho_{y^*}(k) = \frac{\gamma_w(k)}{\gamma_w(0) + \sigma_\epsilon^2}, \qquad k \geq 1,$$

evaluated at the corresponding full-sample parameter estimates. The selected SV$(1, 1)$ specification tracks the empirical autocorrelation pattern substantially more closely over the displayed lags than the two AR-only comparison models.

Panel (b) plots the full 30-day propagation paths underlying Table 7. Each model is subjected to the same common-impact normalization, so conditional volatility rises by 10% at $h = 0$. The subsequent paths differ sharply. Under SV$(1, 0)$ almost the entire increase remains on the following day and then declines gradually. SV$(2, 0)$ permits a richer AR propagation pattern, including a modest short-run rebound generated by the second lag. In contrast, the selected SV$(1, 1)$ specification reduces the increase to about 1.4% after one

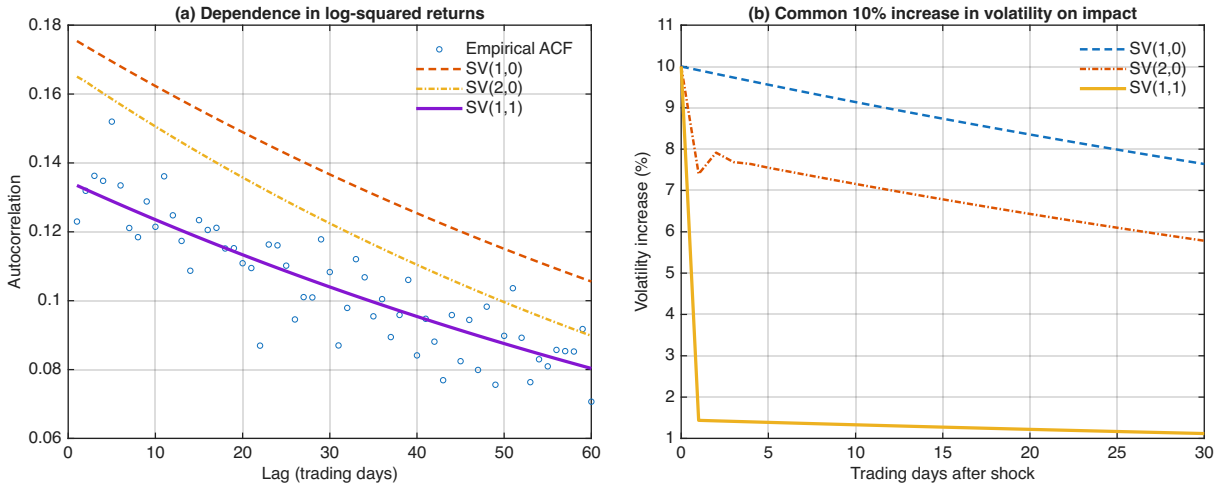


Figure 1: **Observed dependence and model-implied shock propagation.** Panel (a) reports the sample autocorrelation function of centered log-squared returns y_t^* together with the autocorrelation functions implied by the fitted SV(1,0), SV(2,0), and SV(1,1) models. Panel (b) reports the model-implied conditional-volatility response after setting the latent log-variance innovation to $\delta = 2\log(1.10)$, which generates a 10% increase in conditional volatility on impact in every model. Holding the impact magnitude fixed isolates differences in the propagation of the volatility innovation across specifications.

trading day, after which the smaller remaining effect decays slowly. The MA term therefore separates rapid initial adjustment from subsequent persistence, a feature that cannot be generated by a first-order AR volatility law alone.

Since the measurement disturbance is independent and mean zero,

$$\mathbb{E}\Big[y_{t+h}^* \mid v_t = \delta, \mathcal{F}_{t-1}\Big] - \mathbb{E}\Big[y_{t+h}^* \mid v_t = 0, \mathcal{F}_{t-1}\Big] = \delta\,\psi_h.$$

Hence the conditional-mean response of y_t^* has exactly the same shape as the log-variance response; realized y_t^* additionally contains the high-variance log-χ_1^2 measurement disturbance.

7.2 Pseudo out-of-sample volatility forecasting

We next evaluate the recursive forecasting performance of the proposed estimation and filtering pipeline. Adapting the expanding-window design of Awartani and Corradi (2005), we use the same S&P 500 return series as in the full-sample application, reserve $P = 330$ forecast origins, and consider horizons $h \in \{1, 5, 10, 20, 30\}$. The candidate model orders are held fixed throughout the exercise, while all parameters are re-estimated at every forecast origin. Thus the exercise evaluates the pseudo out-of-sample performance of the prespecified specifications rather than an out-of-sample order-selection rule.

Let $H_{\max} = 30$. With $T = 15{,}731$, the first forecasting window contains 15,372 returns,

and each subsequent window adds one observation. The last forecast origin is $t = 15{,}701$, leaving all targets available through the 30-day horizon. At origin t, we estimate the return mean from the available observations,

$$\bar{r}_t = \frac{1}{t}\sum_{s=1}^{t} r_s, \qquad y_s^{(t)} = r_s - \bar{r}_t, \quad s \le t,$$

and construct

$$y_s^{*,(t)} = \log\Big([y_s^{(t)}]^2\Big) - \frac{1}{t}\sum_{j=1}^{t}\log\Big([y_j^{(t)}]^2\Big).$$

We evaluate each h-step variance forecast using the squared mean-adjusted return $[r_{t+h} - \bar{r}_t]^2$, as a proxy for conditional variance.

Motivated by the full-sample selection result, SV$(1,1)$ is the main SV–ARMA specification. At each origin, its ARMA$(1,1)$ reduced-form parameters are re-estimated by the HR procedure and mapped to (θ_1, σ_v^2). We also re-estimate the AR-only SV$(1,0)$ and SV$(2,0)$ comparison models under their structural restrictions; these models use ARMA$(1,1)$ and ARMA$(2,2)$ reduced forms, respectively. Given the resulting structural parameters, the transformed training sample is filtered using the Gaussian state-space approximation described in Appendix G, and the final filtered state is projected forward. The return-variance forecast is

$$\widehat{h}_{t+h|t} = \widehat{\sigma}_{y,t}^2 \exp\left(\widehat{m}_{w,t+h|t} + \frac{1}{2}\widehat{s}_{w,t+h|t}^2\right), \tag{7.1}$$

where $\widehat{m}_{w,t+h|t}$ and $\widehat{s}_{w,t+h|t}^2$ are the predicted conditional mean and variance of latent log variance. We compare the SV forecasts with GARCH(1,1), EGARCH(1,1), and GJR–GARCH(1,1), each estimated by Gaussian quasi-maximum likelihood on the same recursively demeaned training returns.

Forecast accuracy at horizon h is measured by

$$\mathrm{MSE}(h) = \frac{1}{P_h}\sum_{t\in\mathcal{T}_h}\Big([r_{t+h} - \bar{r}_t]^2 - \widehat{h}_{t+h|t}\Big)^2,$$

where $\mathcal{T}_h$ is the common set of origins with valid forecasts from all six models and $P_h = |\mathcal{T}_h|$. In the present exercise, $P_h = 330$ at every horizon, so no forecast replacement or imputation is required.

The results in Table 8 are best viewed as evidence on operational forecasting performance rather than model dominance. EGARCH(1,1) has the lowest one-day MSE, while the AR-only SV models have the lowest reported losses at the longer horizons. The gain from SV$(2,0)$ over SV$(1,0)$, however, is small. The selected SV$(1,1)$ specification does not minimize MSE in this evaluation period, but its performance remains close to the established GARCH-family benchmarks. At $h = 10$, for example, its MSE is 0.1139, compared with 0.1137

Table 8: Pseudo out-of-sample variance-forecast accuracy

Model	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 30$
GARCH(1,1)	0.1071	0.1101	0.1169	0.1193	0.1444
EGARCH(1,1)	**0.1020**	0.1080	0.1137	0.1169	0.1403
GJR–GARCH(1,1)	0.1033	0.1088	0.1158	0.1177	0.1430
SV(1,0)	0.1063	0.1068	0.1109	**0.1151**	0.1394
SV(2,0)	0.1061	**0.1065**	**0.1108**	**0.1151**	**0.1393**
SV(1,1)	0.1077	0.1090	0.1139	0.1198	0.1437

Notes: Entries are MSE multiplied by 10^7, expressed in unscaled log-return units. Estimation is implemented with returns multiplied by 100, and the reported losses are rescaled accordingly. The exercise uses $P = 330$ expanding-window origins and horizons $h \in \{1, 5, 10, 20, 30\}$. Model orders are fixed, while the return mean and all model parameters are re-estimated at every origin using only observations available through that date. SV forecasts use HR reduced-form estimation, structural recovery, and Kalman-filter state projection. All six models produce valid forecasts at all 330 origins for every horizon; no imputation is used. Bold entries denote the smallest MSE at the displayed precision.

for EGARCH(1,1), 0.1158 for GJR–GARCH(1,1), and 0.1169 for GARCH(1,1). The full-sample structural criterion and the out-of-sample MSE target different empirical objects, so selection of an MA component does not imply uniform forecast-loss dominance. Rather, the forecasting exercise shows that the proposed HR–structural-recovery–filtering pipeline can be implemented recursively without numerical failures and delivers competitive multi-step forecasts for a richer stochastic-volatility specification.

8 Conclusion

We studied higher-order stochastic volatility models in which latent log-volatility follows an ARMA(p, q) process. Allowing a moving-average component in volatility is motivated by discrete sampling of continuous-time systems and provides additional flexibility for short-run volatility responses to shocks. Using the log-squared transformation, we established that the centered log-squared returns admit an observable weak ARMA(p, m) representation, with $m = \max\{p, q\}$ under the maintained noncancellation condition. Under Gaussian return innovations, the transformation also produces an additive measurement disturbance with known variance. The latent-volatility problem can therefore be recast as an observable time-series problem without losing the structural distinction between autoregressive and moving-average volatility dynamics.

This representation yields a unified estimation and selection procedure. We estimate the

reduced-form ARMA model using a three-step Hannan–Rissanen method based on linear regressions and recover the structural SV parameters through covariance matching. For fixed, correctly specified orders, the final reduced-form coefficient estimator is consistent and asymptotically normal under the stated regularity and preliminary-estimation conditions. Order selection proceeds in two stages: the observable orders (p, m) are selected first, after which the structural MA order is recovered. When $m > p$, the order relation identifies q directly; when $m = p$, candidate structural orders are distinguished through their complete covariance restrictions using a bootstrap-weighted profile criterion. The same framework thus handles estimation, structural recovery, and order determination.

In the reported Monte Carlo designs, the conditional RMSE of HR generally decreases as the sample size increases, and the method remains applicable to higher-order SV specifications. The success counts show how often an admissible structural estimate is obtained. The two-stage selection procedure also recovers the correct structural order at high rates across the reported designs. These results support estimation and specification of richer SV dynamics without direct likelihood optimization over the latent volatility path.

The empirical application shows why this flexibility can matter. For daily S&P 500 returns from 1962 to 2024, the reduced-form stage selects ARMA$(1, 1)$, and the structural criterion selects SV$(1, 1)$ for every bootstrap block length considered. The estimated MA coefficient separates the initial adjustment to a volatility innovation from the persistence of the remaining effect: only about 15 percent of the initial log-variance response remains after one trading day, while the smaller residual component decays slowly thereafter. The fitted SV$(1, 1)$ model also tracks the dependence pattern in centered log-squared returns more closely than the AR-only comparison specifications. In recursive forecasting, SV$(1, 1)$ does not uniformly minimize squared-return loss, but it produces stable multi-step forecasts whose accuracy is comparable to standard GARCH-family benchmarks. Structural model selection and forecast-loss rankings should therefore be viewed as complementary rather than equivalent criteria.

The main practical implication is that moving-average volatility dynamics need not make SV models difficult to use. Researchers can estimate the observable reduced form, recover and select the latent volatility structure, and then use a Gaussian state-space approximation for filtering and forecasting. This provides a tractable way to distinguish short-run volatility adjustments from persistent effects. Useful extensions include relaxing the Gaussian return-innovation assumption, developing inference for nonlinear structural functions near admissibility boundaries, and extending the framework to multivariate volatility and temporal aggregation. A further direction is to adapt the framework to macroeconomic uncertainty measurement, where distinguishing short-lived changes in innovation variances

from persistent uncertainty dynamics may be particularly valuable.

References

Ahsan, M. N. and J.-M. Dufour (2019): "A Simple Efficient Moment-based Estimator for the Stochastic Volatility Model," in *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A*, ed. by I. Jeliazkov and J. L. Tobias, Emerald Publishing Limited, vol. 40A of *Advances in Econometrics*, 157–201.

——— (2021): "Simple Estimators and Inference for Higher-Order Stochastic Volatility Models," *Journal of Econometrics*, 224, 181–197.

Ahsan, M. N., J.-M. Dufour, and G. Rodriguez-Rondon (2025a): "Estimation and Inference for Higher-Order Stochastic Volatility Models With Leverage," *Journal of Time Series Analysis*, 46, 1064–1084.

——— (2025b): "Estimation pratique de modèles de volatilité stochastique à haute dimension avec des applications à l'incertitude macroéconomique au Québec et au Canada," Project Report 2025RP-19, CIRANO.

Andrews, D. W. and B. Lu (2001): "Consistent Model and Moment Selection Procedures for GMM Estimation With Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101, 123–164.

Asai, M. (2008): "Autoregressive Stochastic Volatility Models With Heavy-Tailed Distributions: A Comparison With Multifactor Volatility Models," *Journal of Empirical Finance*, 15, 332–341.

Awartani, B. M. A. and V. Corradi (2005): "Predicting the Volatility of the S&P-500 Stock Index via GARCH Models: The Role of Asymmetries," *International Journal of Forecasting*, 21, 167–183.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–685.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Broto, C. and E. Ruiz (2004): "Estimation Methods for Stochastic Volatility Models: A Survey," *Journal of Economic Surveys*, 18, 613–649.

Chambers, M. J. (1999): "Discrete Time Representation of Stationary and Non-Stationary Continuous Time Systems," *Journal of Economic Dynamics and Control*, 23, 619–639.

Chambers, M. J. and M. A. Thornton (2012): "Discrete Time Representation of Continuous Time ARMA Processes," *Econometric Theory*, 28, 219–238.

Drost, F. C. and T. E. Nijman (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909–927.

Dufour, J.-M. and D. Pelletier (2022): "Practical Methods for Modeling Weak VARMA Processes: Identification, Estimation and Specification With a Macroeconomic Application," *Journal of Business & Economic Statistics*, 40, 1140–1152.

Engle, R. F. and A. J. Patton (2001): "What Good Is a Volatility Model?" *Quantitative Finance*, 1, 237–245.

Francq, C. and J.-M. Zakoïan (2006): "Linear-representation Based Estimation of Stochastic Volatility Models," *Scandinavian Journal of Statistics*, 33, 785–806.

Hannan, E. J. and J. Rissanen (1982): "Recursive Estimation of Mixed Autoregressive-Moving Average Order," *Biometrika*, 69, 81–94.

Harvey, A., E. Ruiz, and N. Shephard (1994): "Multivariate Stochastic Variance Models," *The Review of Economic Studies*, 61, 247–264.

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

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison With ARCH Models," *The Review of Economic Studies*, 65, 361–393.

Künsch, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *The Annals of Statistics*, 17, 1217–1241.

Lindner, A. M. (2009): "Continuous Time Approximations to GARCH and Stochastic Volatility Models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, Berlin: Springer, 481–496.

López-Pérez, A., M. Febrero-Bande, and W. González-Manteiga (2025): "Estimation and Specification Test for Diffusion Models With Stochastic Volatility," *Statistical Papers*, 66, 40.

Meddahi, N. and E. Renault (2004): "Temporal Aggregation of Volatility Models," *Journal of Econometrics*, 119, 355–379.

Politis, D. N. and J. P. Romano (1992): "A Circular Block-Resampling Procedure for Stationary Data," in *Exploring the Limits of Bootstrap*, ed. by R. LePage and L. Billard, New York: John Wiley & Sons, 263–270.

Shephard, N. and T. G. Andersen (2009): "Stochastic Volatility: Origins and Overview," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiß, and T. Mikosch, Berlin: Springer, 233–254.

Stock, J. H. and M. W. Watson (2007): "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 39, 3–33.

Taylor, S. J. (1982): "Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices, 1961–79," in *Time Series Analysis: Theory and Practice*, ed. by O. D. Anderson, Amsterdam: North-Holland, vol. 1, 203–226.

Urteaga, I., M. F. Bugallo, and P. M. Djurić (2017): "Sequential Monte Carlo for Inference of Latent ARMA Time-Series with Innovations Correlated in Time," *EURASIP Journal on Advances in Signal Processing*, 2017, article 84.