Appendix for

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

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

October 2026

A Additional estimation procedures

This appendix documents two supplementary procedures used in the simulation comparisons. We first derive a moment-based estimator for SV(1,1) models that recovers structural parameters from moments and cross-moments of observed returns. We then describe a root adjustment that enforces stability and invertibility of the estimated reduced-form ARMA model before structural parameter recovery.

A.1 Moment-based estimation for SV(1,1)

This subsection derives the S-Moment benchmark reported in Appendix Tables E.1–E.3. It extends the return-moment approaches of Dufour and Valéry (2006, 2009) and Ahsan and Dufour (2021) to the SV(1,1) specification. It is based on the following identity:

$$y_t := \sigma_y exp((\sum_{j=1}^{p} \phi_j w_{t-j} + v_t - \sum_{j=1}^{q} \theta_j v_{t-j})/2) z_t, \forall t. \tag{A.1}$$

The moments and cross-moments of y_t are given in the following lemma.

Lemma A.1 (Moments and cross-moments of observed returns)**.** *Under Assumptions 2.1–2.3, consider the stationary solution, with i.i.d. standard normal return innovations $\{z_t\}$ independent of the latent volatility process $\{w_t\}$. Let $\bar{\theta}$ denote the model parameter vector, and define*

$$\gamma_h := \mathrm{Cov}(w_t, w_{t+h}), \qquad \rho_h := \frac{\gamma_h}{\gamma_0}, \qquad h \geq 0.$$

For every even nonnegative integer a, let $\lambda(a) := \frac{a!}{2^{a/2}(a/2)!}$. Then, for nonnegative integers k and l and every integer $h \geq 1$,

$$\mu_k(\bar{\theta}) := E[y_t^k] = \begin{cases} \sigma_y^k \lambda(k) \exp\left(\dfrac{k^2 \gamma_0}{8}\right), & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd,} \end{cases} \tag{A.2}$$

$$\mu_{k,l}(h \mid \bar{\theta}) := E[y_t^k y_{t+h}^l] = \begin{cases} \sigma_y^{k+l} \lambda(k)\lambda(l) \exp\left(\dfrac{\gamma_0}{8}[k^2 + l^2 + 2kl\rho_h]\right), & \text{if } k \text{ and } l \text{ are even,} \\ 0, & \text{otherwise.} \end{cases} \tag{A.3}$$

In addition, defining $\gamma_{wv}(-j) := \mathrm{Cov}(w_t, v_{t-j})$, the latent variance satisfies

$$\gamma_0 \left(1 - \sum_{j=1}^{p} \phi_j \rho_j\right) = \sigma_v^2 - \sum_{j=1}^{q} \theta_j \gamma_{wv}(-j).$$

Proof. For $U \sim N(0,1)$ and every nonnegative integer s,

$$E[U^{2s+1}] = 0, \qquad E[U^{2s}] = \frac{(2s)!}{2^s s!}.$$

Also, for any zero-mean Gaussian random variable X,

$$E[\exp(X)] = \exp\left(\frac{1}{2}\operatorname{Var}(X)\right).$$

First, suppose that k is even. Using $y_t = \sigma_y z_t \exp(w_t/2)$ and independence of z_t and w_t gives

$$\begin{aligned}\mu_k(\bar{\theta}) = E(y_t^k) &= \sigma_y^k E[z_t^k] E\left[\exp\left(\frac{kw_t}{2}\right)\right] \\ &= \sigma_y^k \lambda(k) \exp\left(\frac{1}{2}\operatorname{Var}\left(\frac{kw_t}{2}\right)\right) \\ &= \sigma_y^k \lambda(k) \exp\left(\frac{k^2\gamma_0}{8}\right).\end{aligned}$$

If k is odd, the same factorization gives $\mu_k(\bar{\theta}) = 0$ because $E[z_t^k] = 0$.

Next, let $h \geq 1$. Since the latent volatility process is zero-mean Gaussian, (w_t, w_{t+h}) is jointly Gaussian, and

$$\begin{aligned}\operatorname{Var}\left(\frac{kw_t + lw_{t+h}}{2}\right) &= \frac{k^2}{4}\operatorname{Var}(w_t) + \frac{l^2}{4}\operatorname{Var}(w_{t+h}) + \frac{kl}{2}\operatorname{Cov}(w_t, w_{t+h}) \\ &= \frac{k^2 + l^2}{4}\gamma_0 + \frac{kl}{2}\gamma_h \\ &= \frac{\gamma_0}{4}\left(k^2 + l^2 + 2kl\rho_h\right).\end{aligned}$$

Because $\{z_t\}$ is i.i.d. and independent of the entire latent volatility process, z_t and z_{t+h} are independent of each other and of (w_t, w_{t+h}). Thus, when k and l are even,

$$\begin{aligned}\mu_{k,l}(h \mid \bar{\theta}) = E(y_t^k y_{t+h}^l) &= \sigma_y^{k+l} E[z_t^k] E[z_{t+h}^l] E\left[\exp\left(\frac{kw_t + lw_{t+h}}{2}\right)\right] \\ &= \sigma_y^{k+l}\lambda(k)\lambda(l)\exp\left(\frac{1}{2}\operatorname{Var}\left(\frac{kw_t + lw_{t+h}}{2}\right)\right) \\ &= \sigma_y^{k+l}\lambda(k)\lambda(l)\exp\left(\frac{\gamma_0}{8}\left[k^2 + l^2 + 2kl\rho_h\right]\right).\end{aligned}$$

If either k or l is odd, the corresponding Gaussian innovation moment is zero, so $\mu_{k,l}(h \mid \bar{\theta}) = 0$.

Finally, causality of the latent ARMA process implies $\operatorname{Cov}(w_t, v_t) = \sigma_v^2$. Taking covariances of the latent ARMA equation with w_t therefore gives

$$\gamma_0 = \sum_{j=1}^{p}\phi_j\gamma_j + \sigma_v^2 - \sum_{j=1}^{q}\theta_j\gamma_{wv}(-j).$$

Rearranging and using $\gamma_j = \rho_j\gamma_0$ yields the stated variance identity. □

Lemma A.2 (Closed-form moment estimation for SV(1, 1))**.** *Under the assumptions and notation of Lemma A.1, consider the specification* $w_t = \phi_1 w_{t-1} + v_t - \theta_1 v_{t-1}$, *where* $|\phi_1| < 1$, $|\theta_1| < 1$, $\sigma_v^2 > 0$, *and* $\sigma_y > 0$. *Suppose that* $\gamma_1 \neq 0$. *Suppressing parameter arguments in the*

moment notation, the latent autocovariances are recovered as

$$\gamma_0 = \log\left(\frac{\mu_4}{3\mu_2^2}\right), \qquad \gamma_1 = \log\left(\frac{\mu_{2,2}(1)}{\mu_2^2}\right), \qquad \gamma_2 = \log\left(\frac{\mu_{2,2}(2)}{\mu_2^2}\right). \tag{A.4}$$

Define

$$A := \gamma_1^3 - \gamma_0\gamma_1\gamma_2, \qquad B := \gamma_0\gamma_2^2 - 2\gamma_1^2\gamma_2 + \gamma_0\gamma_1^2.$$

Then the structural parameters of SV(1,1) are uniquely recovered as

$$\phi_1 = \frac{\gamma_2}{\gamma_1}, \quad \theta_1 = \begin{cases} \dfrac{-B + \sqrt{B^2 - 4A^2}}{2A}, & A \neq 0, \\ 0, & A = 0, \end{cases} \quad \sigma_v^2 = \frac{\gamma_0 - \phi_1\gamma_1}{1 - \phi_1\theta_1 + \theta_1^2}, \quad \sigma_y = \frac{3^{1/4}\mu_2}{\mu_4^{1/4}}. \tag{A.5}$$

For $A \neq 0$, the displayed quadratic root is the unique root satisfying $|\theta_1| < 1$.

Proof. Applying Lemma A.1 with $k = 2$, $k = 4$, and $k = l = 2$ gives

$$\mu_2 := E(y_t^2) = \sigma_y^2 \frac{2!}{2^1(1!)} exp(\frac{2^2}{8}\gamma_0) = \sigma_y^2 \exp\left(\frac{\gamma_0}{2}\right), \tag{A.6}$$

$$\mu_4 := E(y_t^4) = 3\sigma_y^4 \exp(2\gamma_0), \tag{A.7}$$

$$\mu_{2,2}(1) := E(y_t^2 y_{t-1}^2) = \sigma_y^4 \exp(\gamma_0 + \gamma_1), \tag{A.8}$$

$$\mu_{2,2}(2) := E(y_t^2 y_{t-2}^2) = \sigma_y^4 \exp(\gamma_0 + \gamma_2). \tag{A.9}$$

Consequently,

$$\frac{\mu_4}{3\mu_2^2} = \exp(\gamma_0), \qquad \frac{\mu_{2,2}(h)}{\mu_2^2} = \exp(\gamma_h), \qquad h = 1, 2.$$

Taking logarithms yields the stated expressions for γ_0, γ_1, and γ_2. For example, the autocovariance γ_0 can be recovered by

$$\gamma_0 = \log(\frac{\mu_4}{3\mu_2^2}) = \log(\frac{E(y_t^4)}{3E(y_t^2)^2}) \tag{A.10}$$

Moreover,

$$\sigma_y = \left[\mu_2 \exp\left(-\frac{\gamma_0}{2}\right)\right]^{1/2} = \frac{3^{1/4}\mu_2}{\mu_4^{1/4}}. \tag{A.11}$$

For the latent ARMA(1,1) process, the first moving-average coefficient is $\psi_1 = \phi_1 - \theta_1$. The autocovariance identities therefore become

$$\gamma_0 = \phi_1\gamma_1 + (1 - \phi_1\theta_1 + \theta_1^2)\sigma_v^2, \qquad \gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_v^2, \qquad \gamma_2 = \phi_1\gamma_1. \tag{A.12}$$

Since $\gamma_1 \neq 0$, the last two equations imply

$$\phi_1 = \frac{\gamma_2}{\gamma_1}, \qquad \theta_1\sigma_v^2 = \phi_1\gamma_0 - \gamma_1 = \frac{\gamma_0\gamma_2 - \gamma_1^2}{\gamma_1}. \tag{A.13}$$

Multiplying the first equation in (A.12) by θ_1 and substituting $\theta_1\sigma_v^2 = \phi_1\gamma_0 - \gamma_1$ gives

$$\theta_1(\gamma_0 - \phi_1\gamma_1) = (1 - \phi_1\theta_1 + \theta_1^2)(\phi_1\gamma_0 - \gamma_1).$$

Substituting $\phi_1 = \gamma_2/\gamma_1$ and collecting terms yields

$$(\gamma_1^3 - \gamma_0\gamma_1\gamma_2)\theta_1^2 + (\gamma_0\gamma_2^2 - 2\gamma_1^2\gamma_2 + \gamma_0\gamma_1^2)\theta_1 + (\gamma_1^3 - \gamma_0\gamma_1\gamma_2) = 0.$$

Equivalently, $A\theta_1^2 + B\theta_1 + A = 0$.

The autocovariance identities also imply

$$A = -\gamma_1^2\theta_1\sigma_v^2, \qquad B = \gamma_1^2(1+\theta_1^2)\sigma_v^2.$$

Hence $B - 2|A| = \gamma_1^2\sigma_v^2(1-|\theta_1|)^2 > 0$. If $A \neq 0$, the quadratic has two distinct real roots whose product is one. Since $B > 2|A|$, the unique root inside the unit circle is

$$\theta_1 = \frac{-B + \sqrt{B^2 - 4A^2}}{2A}.$$

If $A = 0$, the identity $A = -\gamma_1^2\theta_1\sigma_v^2$ implies $\theta_1 = 0$.

Finally, rearranging the first equation in (A.12) gives $\sigma_v^2 = \frac{\gamma_0 - \phi_1\gamma_1}{1-\phi_1\theta_1+\theta_1^2}$. Its denominator is strictly positive because

$$1 - \phi_1\theta_1 + \theta_1^2 = \left(\theta_1 - \frac{\phi_1}{2}\right)^2 + 1 - \frac{\phi_1^2}{4} > 0.$$

For $\theta_1 \neq 0$, this expression is equivalent to

$$\sigma_v^2 = \frac{\gamma_0\gamma_2 - \gamma_1^2}{\theta_1\gamma_1}.$$

The variance formula in the lemma also covers $\theta_1 = 0$. □

For higher lags, the latent autocovariances satisfy $\gamma_{h+1} = \phi_1\gamma_h$ for every $h \geq 1$. Thus, whenever $\gamma_h \neq 0$, $\phi_1 = \frac{\gamma_{h+1}}{\gamma_h}$. These ratios coincide at the population level, so the first two autocovariances suffice for the recovery above.

Given demeaned returns $y_1, \ldots, y_T$, define empirical moments

$$\widehat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2, \qquad \widehat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T} y_t^4, \qquad \widehat{\mu}_{2,2}(h) = \frac{1}{T-h}\sum_{t=h+1}^{T} y_t^2 y_{t-h}^2, \qquad h = 1, 2.$$

An admissible estimate requires $\widehat{\gamma}_1 \neq 0$, $|\widehat{\phi}_1| < 1$, and $\widehat{B} > 2|\widehat{A}|$, where $\widehat{A}$ and $\widehat{B}$ are obtained by replacing the population autocovariances in Lemma A.2 with their sample estimates.

A.2 Restricted estimation

In finite samples, preliminary or final estimates of the ARMA(p, m) representation for y_t^* may occasionally fall outside the admissible (causal and/or invertible) region, especially for small T or in the presence of atypical realizations. We therefore consider a simple post-estimation adjustment that projects the estimated parameters back into the admissible region by modifying the roots of the associated lag polynomials. This device is used as a practical fallback and may introduce additional bias relative to the unrestricted estimator. Let $\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$ denote the estimated AR coefficients and define $C(z) = z^p - \hat{\phi}_1 z^{p-1} - \ldots - \hat{\phi}_p$.

Write $\{\hat{z}_j\}_{j=1}^{p}$ for the roots of $C(z) = 0$. Under this normalization, a sufficient and necessary condition for a causal (hence stationary) AR component is $|\hat{z}_j| < 1$ for all j. For a fixed $0 < \Delta < 1$, define the restricted roots

$$\tilde{z}_j = \begin{cases} \hat{z}_j, & |\hat{z}_j| < 1, \\ (1-\Delta)\,\hat{z}_j/|\hat{z}_j|, & |\hat{z}_j| \geq 1, \end{cases} \tag{A.14}$$

so that all unstable roots are mapped radially to modulus $1-\Delta$. The restricted AR coefficients $\tilde{\phi}$ are then obtained by expanding $\tilde{C}(z) = \prod_{j=1}^{p}(z - \tilde{z}_j)$, which can be implemented using `poly`.

Let $\hat{\theta}^* = (\hat{\theta}_1^*, \ldots, \hat{\theta}_m^*)$ denote the estimated MA coefficients and define $D(z) = z^m - \hat{\theta}_1^* z^{m-1} - \ldots - \hat{\theta}_m^*$. Let $\{\hat{z}_j'\}_{j=1}^{m}$ be the roots of $D(z) = 0$. We apply the same radial projection to any root with $|\hat{z}_j'| \geq 1$ to obtain restricted roots $\{\tilde{z}_j'\}$ and thus restricted MA coefficients $\tilde{\theta}^*$ via `poly`. For ARMA models, stationarity depends on the AR polynomial and invertibility on the MA polynomial.

Lastly, given the restricted parameters, we recompute innovations $\tilde{\eta}_t$ from the ARMA recursion

$$\tilde{\eta}_t = y_t^* - \sum_{i=1}^{p} \tilde{\phi}_i y_{t-i}^* + \sum_{j=1}^{m} \tilde{\theta}_j^* \tilde{\eta}_{t-j}$$

initialized with $\tilde{\eta}_t = 0$ for $t \leq L$, where $L = \max\{p, m\}$, and estimate

$$\tilde{\sigma}^2_{\eta,\text{restr}} = \frac{1}{T-L} \sum_{t=L+1}^{T} \tilde{\eta}_t^2. \tag{A.15}$$

Either polynomial may have complex conjugate roots when its degree is at least two. $z \mapsto (1-\Delta)z/|z|$ preserves the root angle and therefore preserves conjugate pairing, which ensures the expanded restricted coefficients remain real up to numerical error. In the reported simulations we set $\Delta \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$; results are qualitatively similar across these values. In MATLAB, roots computes the characteristic roots and poly reconstructs the monic polynomial. If

$$\text{poly}(\widetilde{z}_1, \ldots, \widetilde{z}_p) = (1, c_1, \ldots, c_p),$$

then the restricted AR coefficients are $\widetilde{\phi}_i = -c_i$. The same sign conversion applies to the restricted MA coefficients.

The adjustment enforces reduced-form stability and invertibility. Structural recovery must still satisfy the covariance-matching admissibility conditions in Section 2.2.

B Stationarity, ergodicity and mixing properties

This appendix verifies the stationarity, moment, and dependence properties used in the asymptotic analysis. We consider fixed orders $p, q \geq 0$ and fixed model parameters. The following assumption restates the Gaussian innovation and AR stability conditions maintained in the main text.

Assumption B.1 (Stability and Gaussian innovations)**.**

(i) The polynomial $\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ has no zeros in $|z| \leq 1$.

(ii) The sequences $\{v_t\}$ and $\{z_t\}$ are mutually independent and i.i.d., with $v_t \sim N(0, \sigma_v^2)$, $\sigma_v^2 > 0$, and $z_t \sim N(0,1)$. The return scale satisfies $\sigma_y > 0$.

Recall the latent recursion

$$w_t = \sum_{i=1}^{p} \phi_i w_{t-i} + v_t - \sum_{j=1}^{q} \theta_j v_{t-j}, \qquad t \in \mathbb{Z}, \tag{B.1}$$

with $\theta(z) = 1 - \sum_{j=1}^{q} \theta_j z^j$. Empty sums are zero, and an order-zero lag polynomial equals one. Stability ensures existence of the causal stationary solution; structural MA invertibility is not needed for this conclusion.

We use "geometric β-mixing" to mean that the absolute-regularity coefficients satisfy $\beta(h) \leq K r^h$ for constants $K < \infty$ and $0 < r < 1$. These constants may depend on the fixed model parameters.

To obtain a first-order Markov representation for all orders, let $\bar{p} := \max\{p, 1\}$ and define

$$X_t := (w_t, \ldots, w_{t-\bar{p}+1}, v_t, \ldots, v_{t-q+1})',$$

where the innovation block is omitted when $q = 0$. The state has dimension $\bar{p} + q$ and satisfies

$$X_{t+1} = AX_t + bv_{t+1}. \tag{B.2}$$

Here A is the companion-type matrix induced by (B.1), and b injects the new innovation v_{t+1} into the w- and v-blocks. The AR block is stable and the innovation-shift block is nilpotent, so $\rho(A) < 1$.

Proposition B.1 (Geometric ergodicity and mixing of the latent state)**.** *Under Assumption B.1, the Markov chain in (B.2) has a unique invariant distribution π. Its stationary version is Gaussian, strictly stationary, ergodic, and geometrically β-mixing. Writing P for its transition kernel, it is geometrically ergodic in the following sense: there exist $0 < r_0 < 1$ and a finite $M(x)$ for π-almost every x such that*

$$\|P^n(x, \cdot) - \pi\|_{\mathrm{TV}} \leq M(x) r_0^n, \qquad n \geq 1.$$

In particular, the causal solution w_t is a zero-mean stationary Gaussian process with geometric β-mixing and finite absolute moments of every positive order.

Proof sketch. Since $\rho(A) < 1$ and the innovations have finite variance, the series $X_t = \sum_{j=0}^{\infty} A^j b v_{t-j}$ converges in mean square and defines a stationary Gaussian solution. Its distribution is invariant. Iterating the state recursion shows that the contribution of any initial state is $A^n X_0 \to 0$, whereas the accumulated innovation term converges in distribution to the series above. Thus the invariant distribution is unique.

For mixing, let S_t be the causal stationary solution of $\phi(B)S_t = v_t$. If $p = 0$, $S_t = v_t$ is i.i.d. Otherwise, S_t is a stable AR process driven by nondegenerate Gaussian innovations. The stable-ARMA mixing result of Mokkadem (1988) gives geometric β-mixing of S_t. Since $w_t = \theta(B)S_t, v_t = \phi(B)S_t$, X_t is a function of a fixed finite block of S_t. If d is the largest lag in this block, then $\beta_X(h) \le \beta_S(h-d)$ for $h > d$. Consequently X_t, and its coordinate w_t, are geometrically β-mixing and hence ergodic. Gaussianity gives the stated moments.

For a stationary Markov chain, geometric absolute regularity implies the displayed geometric-ergodicity bound for π-almost every initial state; see Bradley (2005, Theorem 3.7). □

The observed returns satisfy

$$y_t = \sigma_y \exp(w_t/2) z_t, \qquad \epsilon_t = \log(z_t^2) - c_z, \qquad c_z = E[\log(z_t^2)].$$

Then ϵ_t is an i.i.d. centered log-χ_1^2 sequence, independent of the entire latent process, and

$$y_t^* = \log(y_t^2) - E[\log(y_t^2)] = w_t + \epsilon_t.$$

Corollary B.2 (Stationarity, ergodicity, mixing, and moments of returns)**.** *Under Assumption B.1, take the stationary latent solution from Proposition B.1. Then the joint process $(w_t, y_t, y_t^*)'$ is strictly stationary, ergodic, and geometrically β-mixing. Moreover, for every $s > 0$,*

$$E|w_t|^s + E|y_t|^s + E|y_t^*|^s + E|\epsilon_t|^s < \infty.$$

Proof sketch. The sequence $\{z_t\}$ is i.i.d. and independent of $\{X_t\}$. The product process (X_t, z_t) therefore has geometric β-mixing; adjoining an independent i.i.d. sequence does not increase these coefficients. Measurable coordinate transformations preserve this property (Bradley, 2005). Since $(w_t, y_t, y_t^*)'$ is such a transformation, the stationarity, mixing, and ergodicity conclusions follow.

Write $\tau_w^2 := \mathrm{Var}(w_t)$. Gaussianity and independence give, for every $s > 0$,

$$E|y_t|^s = \sigma_y^s E|z_t|^s \, E[\exp(s w_t/2)] = \sigma_y^s E|z_t|^s \exp(s^2 \tau_w^2/8) < \infty.$$

Also, w_t has finite absolute moments of every positive order. The density of ϵ_t is

$$f_\epsilon(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ \frac{x + c_z}{2} - \frac{e^{x+c_z}}{2} \right\}, \qquad x \in \mathbb{R}.$$

Its left tail decays exponentially and its right tail faster than exponentially, so $E|\epsilon_t|^s < \infty$ for every $s > 0$. The corresponding moment claim for $y_t^* = w_t + \epsilon_t$ follows from these bounds. □

B.1 Reduced-form regularity

Take the normalized invertible reduced form established in Theorem 2.1. Its innovations satisfy

$$\eta_t = \frac{\theta(B)}{\theta^*(B)} v_t + \frac{\phi(B)}{\theta^*(B)} \epsilon_t. \tag{B.3}$$

This is generally an infinite causal filter, even though $u_t := \phi(B) y_t^* = \theta(B) v_t + \phi(B)\epsilon_t$ depends on only finitely many primitive innovations.

Corollary B.3 (Regularity of the reduced-form process and innovations)**.** *Under the maintained framework assumptions, the population process $(y_t^*, \eta_t)'$ is strictly stationary, ergodic, and geometrically β-mixing. For every $s > 0$, $E|y_t^*|^s + E|\eta_t|^s < \infty$. In particular, for every $a > 0$, the strong-mixing coefficients of η_t satisfy $\sum_{h=1}^{\infty} \alpha_\eta(h)^a < \infty$.*

Proof. The representations

$$y_t^* = \frac{\theta(B)}{\phi(B)} v_t + \epsilon_t, \qquad \eta_t = \frac{\theta(B)}{\theta^*(B)} v_t + \frac{\phi(B)}{\theta^*(B)} \epsilon_t$$

have stable rational filters with geometrically decaying causal coefficients. Since v_t and ϵ_t have finite absolute moments of every positive order, Minkowski's inequality gives the same property for the filtered processes for $s \geq 1$. Orders $0 < s < 1$ follow from the first-moment bound.

To prove mixing, put $\xi_t = (v_t, \epsilon_t)'$ and $a(z) = \phi(z)\theta^*(z)$. The vectors ξ_t are i.i.d., have finite moments, and have a strictly positive density on $\mathbb{R}^2$. Let $H_t = (H_{1,t}, H_{2,t})'$ be the causal stationary solution of

$$a(B) H_t = \xi_t, \tag{B.4}$$

where the same scalar polynomial acts on each coordinate. The polynomial $a(z)$ has no zeros in $|z| \leq 1$. Hence (B.4) is a stable vector autoregression driven by i.i.d. absolutely continuous innovations. By Mokkadem (1988), H_t is geometrically β-mixing. If $a(z) = 1$, this conclusion follows directly from $H_t = \xi_t$ being i.i.d.

The identities $v_t = a(B) H_{1,t}$ and $\epsilon_t = a(B) H_{2,t}$ imply

$$y_t^* = \theta(B)\theta^*(B) H_{1,t} + a(B) H_{2,t},$$
$$\eta_t = \phi(B)\theta(B) H_{1,t} + \phi(B)^2 H_{2,t}.$$

Thus $(y_t^*, \eta_t)'$ is a function of a fixed finite block of H_t. If d is the largest lag in these finite filters, then $\beta_{(y^*,\eta)}(h) \leq \beta_H(h-d)$, for $h > d$. This proves geometric β-mixing, along with strict stationarity and ergodicity. Since strong-mixing coefficients are bounded by the corresponding absolute-regularity coefficients, the stated positive-power summability follows. □

The mixing argument uses stability of ϕ and invertibility of θ^*, not the equality of its MA order with $\max\{p, q\}$. Thus noncancellation is needed for the exact-order statements elsewhere, but not for this dependence conclusion.

The same construction applies jointly to any fixed finite collection of stable causal rational filters of ξ_t: use the product of their stable denominators in (B.4). Finite-lag products of these filtered variables also inherit geometric mixing and finite moments. This covers the population regressor, score, and fixed-parameter derivative processes used in the HR proofs. These results verify population regularity. The separate arguments for growing autoregressive orders, replacement by estimated filters, and HAC bandwidths remain necessary.

B.2 Regularity for the Hannan–Rissanen proofs

This subsection records three properties used in the estimator proofs: regularity of the population regressors and scores, identification of the regression coefficients, and stability of the estimated filters. Throughout, the model orders and true parameters are fixed, and the maintained SV assumptions apply.

First, the limit theorems used in Steps 2 and 3 concern products of lagged observations and innovations. We therefore need the regularity of $(y_t^*, \eta_t)'$ to extend to these derived processes.

Corollary B.4 (Regularity of fixed-filter HR processes)**.** *The stationary outputs of any fixed finite collection of stable causal rational filters of $\xi_t = (v_t, \epsilon_t)'$, together with finite products of their lags, are jointly strictly stationary, ergodic, and geometrically β-mixing. They have finite absolute moments of every positive order.*

In particular, this applies to the population regressors Z_{t-1} and V_{t-1}, the scores s_{2t} and s_{3t}, and any filter derivatives evaluated at fixed parameter values used in the proofs.

Proof. Let $d(z)$ be the product of the denominator polynomials of the filters under consideration. As in Corollary B.3, the auxiliary process $d(B)K_t = \xi_t$ is a stable vector autoregression with absolutely continuous i.i.d. innovations. Its stationary solution is geometrically β-mixing by Mokkadem (1988). Each filter output, and each product under consideration, is a function of a fixed finite block of K_t, so it inherits stationarity and mixing. Absolute summability of the filter coefficients and Hölder's inequality give the moment conclusion. □

This result supplies the population moment and dependence conditions for the score central limit theorems and population-score HAC estimators. It does not assert stationarity of processes constructed using full-sample parameter estimates.

Second, the population Gram matrices must be nonsingular so that the regressions distinguish the different AR and MA coefficients.

Lemma B.5 (Identification of the population HR regressors)**.** *Let Z_{t-1} and V_{t-1} be defined in (4.2), with $p+m>0$. Minimality of the reduced-form ARMA(p,m) representation and $\sigma_\eta^2>0$ imply that $E[Z_{t-1}'Z_{t-1}]$ and $E[V_{t-1}'V_{t-1}]$ are positive definite. Consequently, $J_{(2)}$ and $J_{(3)}$ are positive definite.*

Proof. Suppose a linear combination of Z_{t-1} vanishes almost surely. It can be written as $a(B)y_t^* - b(B)\eta_t = 0$, where $a(0) = b(0) = 0$, $\deg a \le p$, and $\deg b \le m$. Using $\phi(B)y_t^* = \theta^*(B)\eta_t$ gives $\{a(B)\theta^*(B) - b(B)\phi(B)\}\eta_t = 0$. Since a nonzero finite filter of weak white noise with positive variance cannot vanish almost surely, $a(z)\theta^*(z) = b(z)\phi(z)$. Coprimeness implies that ϕ divides a; the degree restriction and $a(0)=0$ then give $a=0$ and hence $b=0$. Thus the regressors in Z_{t-1} are linearly independent.

A zero linear combination of V_{t-1} would, after application of $\theta^*(B)$, give a zero linear combination of Z_{t-1}. This proves the second claim. Multiplication by $(\sigma_\eta^2)^{-1}>0$ preserves positive definiteness. □

Once the sample Gram matrices converge to these positive-definite limits, their inverses are bounded in probability. Estimators may be assigned an arbitrary value when a sample Gram matrix is singular, because the probability of that event tends to zero.

Third, Step 3 evaluates an inverse MA filter at the preliminary estimates. The relevant requirement is local: estimates close to the true parameter should produce filters close to the true filter. We also need the effect of the specified initial values to disappear as filtering proceeds.

Lemma B.6 (Local filter regularity and initialization)**.** *Let γ_0 be the true reduced-form parameter. There is a neighborhood $\mathcal{N}$ of γ_0 on which all the relevant AR and MA polynomials remain stable and invertible. For each of the fixed-order rational filters used in the HR proofs, write $F(z;g) = \sum_{j=0}^{\infty} F_j(g)z^j$. There are constants $C<\infty$ and $0<\rho<1$, common over $g\in\mathcal{N}$, such that*

$$\|F_j(g)\| \le C\rho^j, \qquad \|F_j(g) - F_j(\gamma_0)\| \le C\|g-\gamma_0\|_2\rho^j. \tag{B.5}$$

The first two parameter derivatives also have uniformly geometrically decaying coefficients when they are needed.

For the stationary SV input, the difference between a filter initialized at a fixed date and its stationary infinite-past version decreases geometrically. This decay holds uniformly over $\mathcal{N}$, with a random multiplier having finite moments of every positive order.

Proof. The true AR and MA roots lie strictly outside the unit circle. A sufficiently small closed ball around γ_0 therefore keeps their roots outside a slightly larger circle as well. The usual power-series coefficient bound gives uniform geometric decay for the rational filters and their parameter derivatives. Applying the mean-value theorem to the coefficients gives the second inequality in (B.5).

These bounds also control the filter outputs uniformly in the parameter: for an input y_t^*, a common bound is proportional to $\sum_{j\geq 0} \rho^j |y_{t-j}^*|$, which has finite moments under the SV assumptions. Finally, differences due to initialization satisfy a stable homogeneous recursion. Their geometric decay follows from stability, and the finite-moment multiplier comes from the stationary initial state. □

The first inequality in (B.5) says that distant observations receive little weight. The second says that a small parameter error produces a proportionally small change in those weights. Together with the finite-moment bounds, these properties justify the filter and product approximations used in Step 3 and in feasible covariance estimation.

Since $\widetilde{\gamma} \to_p \gamma_0$, it belongs to $\mathcal{N}$ with probability tending to one. The uniform bounds can therefore be applied to the estimated filters; invertibility of every finite-sample preliminary estimate is not required.

These results concern fixed-dimensional population processes and local filtering. The growing-order first-step bounds, the second-step generated-regressor condition, and the HAC bandwidth arguments remain in the estimator proofs. Reduced-form noncancellation fixes the exact MA order used there, but is not needed for the mixing argument itself.

C Proofs

Theorem 2.1

Proof of Theorem 2.1. From the measurement equation (2.5) and the state equation (2.2),

$$y_t^* = w_t + \epsilon_t, \qquad \phi(B) w_t = \theta(B) v_t.$$

Applying $\phi(B)$ to y_t^* gives

$$\phi(B) y_t^* = \theta(B) v_t + \phi(B)\epsilon_t =: u_t. \tag{C.1}$$

Because $\{v_t\}$ and $\{\epsilon_t\}$ are mutually independent i.i.d. sequences with zero means and finite variances, $\theta(B)v_t$ is an MA(q) process and $\phi(B)\epsilon_t$ is an independent MA(p) process.

By the moving-average aggregation result of Granger and Morris (1976), their sum admits an MA(r) representation for some $r \leq m := \max\{p, q\}$. Hence, for some coefficients $\theta_1^*, \ldots, \theta_r^*$ and some weak-white-noise process $\{\eta_t\}$ with variance $\sigma_\eta^2 > 0$,

$$u_t = \eta_t - \sum_{j=1}^{r} \theta_j^* \eta_{t-j}. \tag{C.2}$$

Substituting (C.2) into (C.1) yields

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

and equating the two expressions for u_t gives

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

This establishes (2.7) and (2.8).

We next establish uniqueness of the conventional invertible reduced-form MA representation. Since $\{v_t\}$ and $\{\epsilon_t\}$ are mutually independent, the spectral density of u_t is

$$f_u(\lambda) = \frac{1}{2\pi} \left[\sigma_v^2 \left| \theta(e^{-i\lambda}) \right|^2 + \sigma_\epsilon^2 \left| \phi(e^{-i\lambda}) \right|^2 \right]. \tag{C.3}$$

The cross-spectrum is zero because the two component processes are independent. Moreover, stability of the structural AR polynomial implies $\phi(e^{-i\lambda}) \neq 0$ for every λ, and $\sigma_\epsilon^2 > 0$. Therefore, $f_u(\lambda) > 0$ for every λ.

Because $f_u(\lambda)$ is a strictly positive trigonometric polynomial of degree at most m, the Fejér–Riesz spectral-factorization theorem yields a normalized minimum-phase factor. Equivalently, there is a unique invertible polynomial

$$\theta_r^*(z) = 1 - \sum_{j=1}^{r} \theta_j^* z^j$$

and a unique variance $\sigma_\eta^2 > 0$ such that

$$\sigma_\eta^2 \theta_r^*(z) \theta_r^*(z^{-1}) = \sigma_v^2 \theta(z) \theta(z^{-1}) + \sigma_\epsilon^2 \phi(z) \phi(z^{-1}). \tag{C.4}$$

Thus, after imposing the standard invertibility normalization, the reduced-form MA coefficients and innovation variance are uniquely determined.

It remains to establish minimality. Suppose, to the contrary, that $\phi(z)$ and $\theta_r^*(z)$ have a common zero α. Since the structural AR polynomial is stable, $|\alpha| > 1$. Evaluating (C.4) at $z = \alpha$ gives

$$0 = \sigma_v^2 \theta(\alpha) \theta(\alpha^{-1}),$$

because both $\phi(\alpha)$ and $\theta_r^*(\alpha)$ are zero. By Assumption 2.3, $\phi(z)$ and $\theta(z)$ have no common roots, so $\theta(\alpha) \neq 0$. Furthermore, since $|\alpha^{-1}| < 1$, structural invertibility implies $\theta(\alpha^{-1}) \neq 0$. Hence the right-hand side is nonzero, which is a contradiction. Therefore $\phi(z)$ and $\theta_r^*(z)$ have no common roots, and the weak ARMA(p, r) representation is minimal. □

Proposition 2.2

Proof. When $p \neq q$, the equality $r = m$ follows because the component of order m is the only one contributing to the autocovariance of the filtered disturbance at lag m, and its outer coefficient is nonzero by exact structural order. If $p = q = 0$, then $y_t^* = v_t + \epsilon_t$ is white noise, so $r = m = 0$. When $p = q = m \geq 1$, the lag-m autocovariance of the filtered disturbance is $-\theta_m \sigma_v^2 - \phi_m \sigma_\epsilon^2$. Assumption 2.4 makes this quantity nonzero, so the reduced-form disturbance cannot have MA order below m. Theorem 2.1 already gives $r \leq m$, hence $r = m$. □

Proposition 2.3

Proof of Proposition 2.3. By stability of the autoregressive polynomial, the latent volatility process has the causal representation

$$w_t = \frac{\theta(B)}{\phi(B)} v_t = \sum_{h=0}^{\infty} \psi_h v_{t-h}, \tag{C.5}$$

where the coefficients $\{\psi_h\}_{h \geq 0}$ are absolutely summable. Equating coefficients in $\phi(B) \sum_{h=0}^{\infty} \psi_h B^h = \theta(B)$ gives

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

with $\theta_h = 0$ for $h > q$. An immediate consequence of (C.5) is

$$\mathrm{Cov}(v_{t-j}, w_{t-k}) = \begin{cases} \sigma_v^2 \psi_{j-k}, & j \geq k, \\ 0, & j < k. \end{cases} \tag{C.6}$$

Indeed, when $j \geq k$, the innovation v_{t-j} appears in w_{t-k} with coefficient ψ_{j-k}; when $j < k$, it occurs after time $t - k$ and is independent of w_{t-k}.

From (2.9), write

$$y_t^* = \sum_{j=1}^{p} \phi_j y_{t-j}^* + A_t + B_t, \qquad A_t := v_t - \sum_{j=1}^{m} \theta_j v_{t-j}, \qquad B_t := \epsilon_t - \sum_{j=1}^{m} \phi_j \epsilon_{t-j},$$

where $\theta_j = 0$ for $j > q$ and $\phi_j = 0$ for $j > p$. Taking covariance with y_{t-k}^* yields

$$\gamma_{y^*}(k) = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) + \mathrm{Cov}(A_t, y_{t-k}^*) + \mathrm{Cov}(B_t, y_{t-k}^*). \tag{C.7}$$

As $\{\epsilon_t\}$ is independent of the entire innovation sequence $\{v_t\}$, it is independent of $\{w_t\}$,

$$\mathrm{Cov}(A_t, y_{t-k}^*) = \mathrm{Cov}(A_t, w_{t-k}), \qquad \mathrm{Cov}(B_t, y_{t-k}^*) = \mathrm{Cov}(B_t, \epsilon_{t-k}).$$

For $k = 0$, equation (C.6) gives

$$\begin{aligned}\operatorname{Cov}(A_t, w_t) &= \operatorname{Cov}\left(v_t - \sum_{j=1}^{m} \theta_j v_{t-j}, w_t\right) \\ &= \sigma_v^2 - \sigma_v^2 \sum_{j=1}^{m} \theta_j \psi_j = \sigma_v^2 \left(1 - \sum_{j=1}^{m} \theta_j \psi_j\right).\end{aligned}$$

Moreover, since $\{\epsilon_t\}$ is i.i.d., $\operatorname{Cov}(B_t, \epsilon_t) = \sigma_\epsilon^2$. Substituting these expressions into (C.7) and using $\gamma_{y^*}(-j) = \gamma_{y^*}(j)$ gives

$$\gamma_{y^*}(0) = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(j) + \sigma_v^2 \left(1 - \sum_{j=1}^{m} \theta_j \psi_j\right) + \sigma_\epsilon^2.$$

Now consider $1 \le k \le m$. Since v_t is independent of w_{t-k}, equation (C.6) implies

$$\begin{aligned}\operatorname{Cov}(A_t, w_{t-k}) &= -\sum_{j=1}^{m} \theta_j \operatorname{Cov}(v_{t-j}, w_{t-k}) \\ &= -\sigma_v^2 \sum_{j=k}^{m} \theta_j \psi_{j-k} = -\sigma_v^2 \left(\theta_k + \sum_{l=1}^{m-k} \theta_{k+l} \psi_l\right),\end{aligned}$$

where the last equality follows from $\psi_0 = 1$ and the change of index $l = j - k$. For the measurement-error component, independence across time gives

$$\operatorname{Cov}(B_t, \epsilon_{t-k}) = -\phi_k \sigma_\epsilon^2, \qquad 1 \le k \le m,$$

where $\phi_k = 0$ for $k > p$. Substituting the preceding two expressions into (C.7) yields

$$\gamma_{y^*}(k) = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j) - \sigma_v^2 \left(\theta_k + \sum_{l=1}^{m-k} \theta_{k+l} \psi_l\right) - \phi_k \sigma_\epsilon^2, \qquad 1 \le k \le m.$$

Finally, let $k > m$. Since $m = \max\{p, q\}$, we have $k > p$ and $k > q$. Consequently,

$$\operatorname{Cov}(A_t, w_{t-k}) = 0 \qquad \text{and} \qquad \operatorname{Cov}(B_t, \epsilon_{t-k}) = 0.$$

Equation (C.7) therefore reduces to

$$\gamma_{y^*}(k) = \sum_{j=1}^{p} \phi_j \gamma_{y^*}(k-j), \qquad k > m.$$

This establishes all three cases in (2.11). □

Proposition 2.4

Proof of Proposition 2.4. We first prove part (i). Define $u_t := \phi(B) y_t^* = \theta(B) v_t + \phi(B) \epsilon_t$. The structural parameters determine the spectral density

$$f_u(\lambda) = \frac{1}{2\pi} \left[\sigma_v^2 \left|\theta(e^{-i\lambda})\right|^2 + \sigma_\epsilon^2 \left|\phi(e^{-i\lambda})\right|^2\right].$$

By Theorem 2.1, the normalized invertible spectral factor of this density is unique. Hence the reduced-form MA polynomial $\theta^*(B)$ and its innovation variance σ_η^2 are uniquely determined. Under Assumption 2.4, Proposition 2.2 further gives $m = \max\{p, q\}$ as the exact reduced-form

MA order. This proves part (i).

For part (ii), fix the structural order q. Let $\theta := (\theta_1, \ldots, \theta_q)'$ and $\widetilde{\theta} := (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_q)'$ denote two candidate structural MA coefficient vectors, with corresponding polynomials $\theta(B)$ and $\widetilde{\theta}(B)$. Similarly, write $\phi := (\phi_1, \ldots, \phi_p)'$ and $\theta^* := (\theta_1^*, \ldots, \theta_m^*)'$ for the structural AR and reduced-form MA coefficient vectors. Suppose that the two admissible pairs (θ, σ_v^2) and $(\widetilde{\theta}, \widetilde{\sigma}_v^2)$ both satisfy the covariance-matching system (2.12) for the same $(\phi, \theta^*, \sigma_\eta^2, \sigma_\epsilon^2)$.

The reduced-form parameters determine the autocovariances

$$\gamma_u(k) := \mathrm{Cov}(u_t, u_{t-k}), \qquad k = 0, \ldots, m,$$

where

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

As the two innovation sequences are independent, its autocovariances can be written as

$$\gamma_u(k) = A_k(\theta)\sigma_v^2 + B_k(\phi)\sigma_\epsilon^2,$$

where $A_k(\theta)$ and $B_k(\phi)$ are the MA autocovariance coefficients appearing explicitly in (2.12).

Since $(\phi, \sigma_\epsilon^2)$ is known, define the adjusted autocovariances

$$\widetilde{\gamma}(k) := \gamma_u(k) - B_k(\phi)\sigma_\epsilon^2.$$

For any admissible structural representation of order q, $\widetilde{\gamma}(k) = A_k(\theta)\sigma_v^2$, for $k = 0, \ldots, m$. Thus the two candidate pairs imply $A_k(\theta)\sigma_v^2 = A_k(\widetilde{\theta})\widetilde{\sigma}_v^2$, for $k = 0, \ldots, m$.

Because both structural MA polynomials have degree q, their autocovariances are zero for lags greater than q. Hence the preceding equalities imply that the MA(q) processes $\theta(B)v_t$ and $\widetilde{\theta}(B)\widetilde{v}_t$ have the same complete autocovariance function. Both MA polynomials are normalized and invertible. Therefore they must be the same normalized invertible spectral factor of this autocovariance structure. It follows that $\theta(B) = \widetilde{\theta}(B)$ and $\sigma_v^2 = \widetilde{\sigma}_v^2$. Hence, for the specified structural order q, there is at most one admissible structural pair satisfying (2.12). □

Corollary 2.5

Proof of Corollary 2.5. For an ARMA(1, 1) reduced form, the covariance-matching equations (2.12) reduce to

$$\left(1 + (\theta_1^*)^2\right)\sigma_\eta^2 = (1 + \theta_1^2)\sigma_v^2 + (1 + \phi_1^2)\sigma_\epsilon^2, \tag{C.8}$$

$$\theta_1^*\sigma_\eta^2 = \theta_1\sigma_v^2 + \phi_1\sigma_\epsilon^2. \tag{C.9}$$

Using the definitions of C and D in the corollary, these two equations are equivalently

$$C = (1 + \theta_1^2)\sigma_v^2, \qquad D = \theta_1\sigma_v^2. \tag{C.10}$$

Thus the problem reduces to solving these two equations for (θ_1, σ_v^2) subject to $\sigma_v^2 > 0$ and

$|\theta_1| < 1$.

First consider $D = 0$. Since $\sigma_v^2 > 0$, the second equation in (C.10) implies $\theta_1 = 0$, and the first then gives $\sigma_v^2 = C$. Hence an admissible solution exists if and only if $C > 0$, which in this case is exactly the condition $C > 2|D|$. The solution is unique.

Now suppose $D \neq 0$. From the second equation in (C.10), $\sigma_v^2 = \frac{D}{\theta_1}$. Substituting this into the first equation gives $D\theta_1^2 - C\theta_1 + D = 0$. The two candidate solutions are therefore

$$\theta_1 = \frac{C \pm \sqrt{C^2 - 4D^2}}{2D}.$$

We next determine when one of these solutions is admissible. If an admissible solution exists, then (C.10) implies

$$C - 2|D| = \sigma_v^2\Big(1 + \theta_1^2 - 2|\theta_1|\Big) = \sigma_v^2(1 - |\theta_1|)^2 > 0,$$

so $C > 2|D|$ is necessary. Conversely, suppose $C > 2|D|$. Then $C^2 - 4D^2 > 0$, and the root

$$\theta_1 = \frac{C - \sqrt{C^2 - 4D^2}}{2D} = \frac{2D}{C + \sqrt{C^2 - 4D^2}}$$

satisfies $|\theta_1| = \frac{2|D|}{C+\sqrt{C^2-4D^2}} < 1$. Using $\sigma_v^2 = D/\theta_1$ then gives $\sigma_v^2 = \frac{C+\sqrt{C^2-4D^2}}{2} > 0$. Thus $C > 2|D|$ is also sufficient.

Finally, when $D \neq 0$, the two roots of $D\theta_1^2 - C\theta_1 + D = 0$ have product one. Hence, whenever $C > 2|D|$, one root lies strictly inside the unit circle and the other lies strictly outside it. The invertibility restriction therefore selects a unique structural solution. Combining the cases $D = 0$ and $D \neq 0$, the unique admissible parameters can be written uniformly as

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

If $D \neq 0$, then $\theta_1 \neq 0$ and the structural MA order is one, so the model is SV$(1, 1)$. If $D = 0$, then $\theta_1 = 0$ and the model reduces to SV$(1, 0)$. □

Corollary 3.1

Proof. For $p \geq 1$, Proposition 2.3 implies

$$\gamma_{y^*}(k) = \sum_{i=1}^{p} \phi_i \gamma_{y^*}(k - i), \qquad k > m.$$

Fix any j with $p + j > m$ and consider the p equations for $k = p + j, \ldots, 2p + j - 1$. With the definitions in (3.1), they give

$$\Gamma(p, j)\Phi_p = \gamma(p, j),$$

where $\Gamma(p, j)$ is a $p \times p$ Toeplitz matrix and $\gamma(p, j)$ is a vector as defined before. If $\Gamma(p, j)$ is nonsingular, then

$$\Phi_p = \Gamma(p, j)^{-1}\gamma(p, j), \tag{C.11}$$

which proves part (i). For $p = 0$, no AR coefficients need to be recovered.

For part (ii), the true vector $\Theta = (\theta_1, \ldots, \theta_q, \sigma_v^2)'$ satisfies $G_K(\Phi_p, \Theta) = \gamma(0 : K)$ by Proposition 2.3. The short-lag covariance terms use the impulse-response coefficients ψ_h defined in (2.10). To show uniqueness, suppose two admissible structural vectors have the same Φ_p and the same observed autocovariances through lag $K \geq m$. At every lag greater than m, both autocovariance sequences satisfy the same AR recursion. Since $p \leq m$, equality through lag m determines equality at all subsequent lags. Symmetry then gives the same complete observed autocovariance sequence.

Subtracting σ_ϵ^2 at lag zero recovers the same latent autocovariance sequence for both vectors. Applying the common AR filter $\phi(B)$ therefore gives the same complete autocovariance sequence for $\theta(B)v_t$. The normalized invertible MA factor and its positive innovation variance are unique by the spectral-factor argument in Proposition 2.4. Thus both structural vectors coincide. Existence follows from the maintained data-generating model. In particular, when $q \leq p$, $K = p$ gives $p+1$ covariance restrictions for $q+1$ unknown MA and innovation-variance parameters; zero padding adds no unknowns.

Finally, taking expectations in (2.3) and using $E[w_t] = 0$ gives

$$\mu = E[\log(y_t^2)] = \log(\sigma_y^2) + c_z. \tag{C.12}$$

Consequently,

$$\sigma_y^2 = \exp(\mu - c_z), \tag{C.13}$$

which proves part (iii). □

In order to prove Proposition 4.1, consider below three lemmas first. They retain the geometric-tail and uniform-moment organization of the first-step argument. For a random sequence A_t, write $\|A\|_{T,r} := (T^{-1} \sum_t \|A_t\|^r)^{1/r}$, with the summation range depending on context. Stationarity allows the uniform bounds below to be applied to any consecutive sample block of the stated length.

The first lemma is a variant of **Lemma D.3** in Dufour and Pelletier (2022).

Lemma C.1 (Infinite AR convergence)**.** *If an ARMA model with AR polynomial $\phi(z)$ and MA polynomial $\theta^*(z)$ is invertible, write*

$$\frac{\phi(z)}{\theta^*(z)} = 1 - \sum_{j=1}^{\infty} \Pi_j z^j.$$

If $n_T / \log T \to \infty$, then

$$\sum_{j > n_T} |\Pi_j| = o(T^{-1}).$$

Proof. Choose $R > 1$ strictly smaller than the modulus of every zero of θ^*. The ratio $\phi(z)/\theta^*(z)$ is analytic on a neighborhood of $|z| \leq R$. Cauchy's coefficient bound gives

$|\Pi_j| \le KR^{-j}$ for a finite constant K, so

$$\sum_{j>n_T} |\Pi_j| \le \frac{KR^{-n_T}}{R-1}.$$

Since $a \log T - n_T \log R \to -\infty$ for every fixed $a > 0$, the last bound is $o(T^{-a})$; take $a = 1$. If θ^* is constant, the autoregressive expansion is finite and the result is immediate. □

The following lemma is a variant of **Lemma D.4** in Dufour and Pelletier (2022). It bounds sample second-moment errors uniformly over lag indices. This uniformity is useful when the preliminary autoregressive order increases with the sample size.

Lemma C.2 (Covariance estimation)**.** *Under the maintained SV assumptions, fix the model parameters and consider the stationary process $\{y_t^*\}$. There is a finite constant K, independent of T, r, s, such that*

$$E\left|\frac{1}{T}\sum_{t=1}^{T}\left\{y_{t-r}^* y_{t-s}^* - E[y_{t-r}^* y_{t-s}^*]\right\}\right|^2 \le \frac{K}{T}$$

for every positive integer T and integer lags r, s. Thus the bound also applies to deterministic lags that vary with T.

Proof. We first establish a uniform bound for products of the primitive innovations, then transfer it to y_t^* through its causal representation.

Primitive innovation products. Let $\{e_t\}$ and $\{f_t\}$ each denote either $\{v_t\}$ or $\{\epsilon_t\}$, allowing the two choices to coincide. For integer lags a, b, define

$$Q_T^{e,f}(a,b) := \frac{1}{T}\sum_{t=1}^{T}\left\{e_{t-a}f_{t-b} - E[e_{t-a}f_{t-b}]\right\}.$$

The primitive innovations have zero means and finite fourth moments. Set

$$M_4 := \max\{E|v_0|^4, E|\epsilon_0|^4\} < \infty.$$

For fixed a, b, the centered products in this average are uncorrelated across distinct dates. If $a = b$, this follows directly from independence across innovation dates. If $a \ne b$, each product has mean zero. For $t \ne t'$, the four factors in $e_{t-a}f_{t-b}e_{t'-a}f_{t'-b}$ contain at least one innovation date appearing only once. The corresponding factor is centered and independent of the remaining factors, so the expectation of this product is zero. Consequently,

$$\mathrm{Cov}(e_{t-a}f_{t-b}, e_{t'-a}f_{t'-b}) = 0, \qquad t \ne t'.$$

Expanding the variance therefore gives

$$\begin{aligned} E|Q_T^{e,f}(a,b)|^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{t'=1}^{T}\mathrm{Cov}(e_{t-a}f_{t-b}, e_{t'-a}f_{t'-b}) \\ &= \frac{1}{T}\mathrm{Var}(e_{-a}f_{-b}) \le \frac{1}{T}\left(E|e_{-a}|^4 E|f_{-b}|^4\right)^{1/2} \le \frac{M_4}{T}. \end{aligned} \tag{C.14}$$

This bound is independent of a, b and of the choices of e, f.

Transfer to the transformed process. Stability of the latent ARMA model gives

$$y_t^* = \sum_{j=0}^{\infty} \psi_j v_{t-j} + \epsilon_t, \qquad P := \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Let $\Delta_T(r,s)$ denote the centered sample average in the statement of the lemma. Expanding the two factors yields

$$\Delta_T(r,s) = \sum_{j=0}^{\infty}\sum_{h=0}^{\infty} \psi_j\psi_h Q_T^{v,v}(r+j,s+h) + \sum_{j=0}^{\infty} \psi_j Q_T^{v,\epsilon}(r+j,s) + \sum_{h=0}^{\infty} \psi_h Q_T^{\epsilon,v}(r,s+h) + Q_T^{\epsilon,\epsilon}(r,s).$$

These expansions can first be made with finite filters. Absolute summability and finite fourth moments ensure convergence of the filter series in L^4, and hence convergence of the products in L^2. Write $\|X\|_{L^2} := (E|X|^2)^{1/2}$. Applying the triangle inequality and (C.14) gives

$$\|\Delta_T(r,s)\|_{L^2} \le \sqrt{\frac{M_4}{T}}\left(P^2 + 2P + 1\right) = \sqrt{\frac{M_4}{T}}(P+1)^2.$$

Squaring both sides proves the result with $K = M_4(P+1)^4$, which is independent of T, r, s. □

Lemma C.3 (Moment estimation)**.** *Under the maintained SV assumptions, there exists a finite constant $K_{y\eta}$, independent of T, r, s, such that*

$$E\left|\frac{1}{T}\sum_{t=1}^{T}\left\{y_{t-r}^*\eta_{t-s} - E[y_{t-r}^*\eta_{t-s}]\right\}\right|^2 \le \frac{K_{y\eta}}{T}$$

for every positive integer T and integer lags r, s. Thus the bound also applies to deterministic lags that vary with T.

Proof. The argument follows the same structure as Lemma C.2. Recall that

$$y_t^* = \sum_{j=0}^{\infty} \psi_j v_{t-j} + \epsilon_t, \qquad P := \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

Moreover, since $\theta^*(B)\eta_t = \theta(B)v_t + \phi(B)\epsilon_t$ and $\theta^*(B)$ is invertible,

$$\eta_t = \sum_{h=0}^{\infty} a_h v_{t-h} + \sum_{h=0}^{\infty} b_h \epsilon_{t-h},$$

where $A := \sum_{h=0}^{\infty}|a_h| < \infty$ and $B := \sum_{h=0}^{\infty}|b_h| < \infty$.

Let

$$\Delta_T^{y\eta}(r,s) := \frac{1}{T}\sum_{t=1}^{T}\left\{y_{t-r}^*\eta_{t-s} - E[y_{t-r}^*\eta_{t-s}]\right\}.$$

Using the notation $Q_T^{e,f}(a,b)$ from the proof of Lemma C.2, expansion gives

$$\Delta_T^{y\eta}(r,s) = \sum_{j,h\ge 0} \psi_j a_h Q_T^{v,v}(r+j,s+h) + \sum_{j,h\ge 0} \psi_j b_h Q_T^{v,\epsilon}(r+j,s+h) + \sum_{h\ge 0} a_h Q_T^{\epsilon,v}(r,s+h) + \sum_{h\ge 0} b_h Q_T^{\epsilon,\epsilon}(r,s+h).$$

By Lemma C.2, $\|Q_T^{e,f}(a,b)\|_{L^2} \le \sqrt{\frac{M_4}{T}}$ uniformly over the primitive innovation types and all lags. Therefore, by the triangle inequality,

$$\|\Delta_T^{y\eta}(r,s)\|_{L^2} \le \sqrt{\frac{M_4}{T}}(P+1)(A+B).$$

Squaring both sides gives the stated result, with $K_{y\eta} = M_4(P+1)^2(A+B)^2$. □

The fixed-dimensional population regularity, identification, and local-filter results invoked below are established in subsection B.2.

Proposition 4.1 (First-step Estimates)

Proof of Proposition 4.1. Let $n := n_T$, $N := T - n$ and define $x_t := (1, y^*_{t-1}, \dots, y^*_{t-n})'$ where $\beta_n := (0, \Pi_1, \dots, \Pi_n)'$. The first coordinate of β_n is zero because y_t^* is population centered. By invertibility of the reduced-form MA polynomial, y_t^* admits the infinite autoregressive representation

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

Hence the exact finite-order regression is

$$y_t^* = x_t'\beta_n + \eta_t + r_{t,n}, \qquad r_{t,n} := \sum_{j>n} \Pi_j y^*_{t-j}. \tag{C.15}$$

Let $\delta_n := \sum_{j>n} |\Pi_j|$. By Lemma C.1 and $n_T/\log T \to \infty$, $\delta_n = o(T^{-1})$.

In the first step, we use OLS to estimate the first n_T $AR(\infty)$ coefficients. Let

$$B_n := E[x_t x_t'], \qquad \widehat{B}_n := \frac{1}{N}\sum_{t=n+1}^{T} x_t x_t',$$

and let $\widehat{\beta}_n^{\mathrm{LS}}$ minimize Q_T. On the event that $\widehat{B}_n$ is nonsingular, and the OLS normal equations and (C.15) give

$$\widehat{\beta}_n^{\mathrm{LS}} - \beta_n = \widehat{B}_n^{-1}\left(\frac{1}{N}\sum_{t=n+1}^{T} x_t r_{t,n} + \frac{1}{N}\sum_{t=n+1}^{T} x_t \eta_t\right). \tag{C.16}$$

Let

$$V_{1T} := \frac{1}{N}\sum_{t=n+1}^{T} x_t r_{t,n}, \qquad V_{2T} := \frac{1}{N}\sum_{t=n+1}^{T} x_t \eta_t \tag{C.17}$$

Thus V_{1T} is the omitted-tail term and V_{2T} is the regressor–innovation term.

Consider V_{2T}. Causality of the reduced-form ARMA representation and weak whiteness of η_t imply $E[y^*_{t-j}\eta_t] = 0$ for $j \ge 1$ and $E[\eta_t] = 0$. Consequently, every coordinate of V_{2T} has mean zero. The intercept coordinate is $(V_{2T})_0 = \frac{1}{N}\sum_{t=n+1}^{T} \eta_t$. Since η_t is weak white noise, $E|(V_{2T})_0|^2 = O(N^{-1})$. For the remaining coordinates, Lemma C.3, applied uniformly over

$j = 1, \ldots, n$, gives

$$E\left|\frac{1}{N}\sum_{t=n+1}^{T} y_{t-j}^{*}\eta_t\right|^2 \leq \frac{C}{N}$$

for a constant C independent of j, n, and T. It follows that

$$E\|V_{2T}\|_2^2 \leq C\frac{n+1}{N} = O(n/T), \tag{C.18}$$

because $n^2/T \to 0$ implies $N/T \to 1$. Therefore, $\|V_{2T}\|_2 = O_p\left(\sqrt{n/T}\right)$.

Similarly, consider V_{1T}. For $k = 1, \ldots, n$, the kth lag coordinate is

$$(V_{1T})_k = \sum_{j>n}\Pi_j\left\{\frac{1}{N}\sum_{t=n+1}^{T} y_{t-k}^{*}y_{t-j}^{*}\right\}.$$

Lemma C.2, together with $|E[y_{t-k}^{*}y_{t-j}^{*}]| \leq E[(y_t^{*})^2]$, implies that, for a finite constant C independent of k, j, n, and T, $\sup_{k,j}\left\|\frac{1}{N}\sum_{t=n+1}^{T} y_{t-k}^{*}y_{t-j}^{*}\right\|_{L^2} \leq C$. Writing $\delta_n := \sum_{j>n}|\Pi_j|$, Minkowski's inequality therefore gives $\|(V_{1T})_k\|_{L^2} \leq C\delta_n$. The first component of V_{1T}, corresponding to the constant regressor, is $(V_{1T})_0 = \sum_{j>n}\Pi_j\left\{\frac{1}{N}\sum_{t=n+1}^{T} y_{t-j}^{*}\right\}$. The corresponding uniform sample-mean bound gives $\|(V_{1T})_0\|_{L^2} \leq \frac{C\delta_n}{\sqrt{N}}$. Consequently, $E\|V_{1T}\|_2^2 \leq C\left(n + \frac{1}{N}\right)\delta_n^2$. By Lemma C.1, $\delta_n = o(T^{-1})$, and hence

$$E\|V_{1T}\|_2^2 = o\left(\frac{n}{T^2}\right), \qquad \|V_{1T}\|_2 = o_p\left(\frac{\sqrt{n}}{T}\right). \tag{C.19}$$

In particular, $\|V_{1T}\|_2 = o_p\left(\sqrt{\frac{n}{T}}\right)$, so the omitted-tail term is asymptotically negligible relative to V_{2T}, which is $O_p(\sqrt{n/T})$.

The last building block is the behavior of the sample Gram matrix $\widehat{B}_n$. Because y_t^{*} is a stable ARMA process with spectral density bounded above and bounded away from zero, the eigenvalues of the population matrix $B_n := E[x_t x_t']$ are bounded above and away from zero uniformly in n. Lemma C.2, together with the corresponding sample-mean bound for the intercept term, gives $E\|\widehat{B}_n - B_n\|_F^2 = O(n^2/T)$. Since $n^2/T \to 0$, $\|\widehat{B}_n - B_n\|_2 = o_p(1)$. Therefore, with probability tending to one, $\widehat{B}_n$ is nonsingular

$$\|\widehat{B}_n^{-1}\|_2 = O_p(1), \qquad \|\widehat{B}_n\|_2 = O_p(1). \tag{C.20}$$

Putting the preceding results together,

$$\widehat{\beta}_n^{\mathrm{LS}} - \beta_n = \widehat{B}_n^{-1}(V_{1T} + V_{2T}),$$

and hence

$$\|\widehat{\beta}_n^{\mathrm{LS}} - \beta_n\|_2 \leq \|\widehat{B}_n^{-1}\|_2\left(\|V_{1T}\|_2 + \|V_{2T}\|_2\right) = O_p\left(\sqrt{n/T}\right),$$

because

$$\|V_{1T}\|_2 = o_p\left(\sqrt{n/T}\right), \qquad \|V_{2T}\|_2 = O_p\left(\sqrt{n/T}\right).$$

Thus the sampling term V_{2T} determines the first-step rate, whereas the error from truncating the infinite autoregression is asymptotically negligible.

It remains to extend this result to the implemented first-step fit. Let $\widehat{\beta}_n := (\widehat{c}, \widehat{\Pi}_1^{(n)}, \dots, \widehat{\Pi}_n^{(n)})'$. Since $\widehat{\beta}_n^{\mathrm{LS}}$ minimizes the least-squares criterion, the exact quadratic identity gives

$$Q_T(\widehat{\beta}_n) - Q_T(\widehat{\beta}_n^{\mathrm{LS}}) = (\widehat{\beta}_n - \widehat{\beta}_n^{\mathrm{LS}})' \widehat{B}_n (\widehat{\beta}_n - \widehat{\beta}_n^{\mathrm{LS}}).$$

By condition (4.1), the left-hand side is $O_p(n/T)$. Together with (C.20), this implies $\|\widehat{\beta}_n - \widehat{\beta}_n^{\mathrm{LS}}\|_2 = O_p\Big(\sqrt{n/T}\Big)$. Combining this with the OLS rate yields

$$\|\widehat{\beta}_n - \beta_n\|_2 = O_p\Big(\sqrt{n/T}\Big). \tag{C.21}$$

Since $\beta_n = (0, \Pi_1, \dots, \Pi_n)'$, (C.21) implies $\widehat{c} = O_p\Big(\sqrt{n/T}\Big)$ and $\|\widehat{\Pi}^{(n)} - \Pi^{(n)}\|_2 = O_p\Big(\sqrt{n/T}\Big)$. By Cauchy–Schwarz,

$$\sum_{j=1}^{n} |\widehat{\Pi}_j^{(n)} - \Pi_j| \le \sqrt{n}\, \|\widehat{\Pi}^{(n)} - \Pi^{(n)}\|_2 = O_p(n/\sqrt{T}).$$

Finally, with $y_t^* = x_t' \beta_n + \eta_t + r_{t,n}$, the first-step residual error satisfies

$$\widehat{U}_t - \eta_t = -x_t'(\widehat{\beta}_n - \beta_n) + r_{t,n}.$$

Therefore,

$$\frac{1}{T} \sum_{t=n+1}^{T} (\widehat{U}_t - \eta_t)^2 \le \frac{2N}{T} (\widehat{\beta}_n - \beta_n)' \widehat{B}_n (\widehat{\beta}_n - \beta_n) + \frac{2}{T} \sum_{t=n+1}^{T} r_{t,n}^2.$$

The first term is $O_p(n/T)$ by (C.21) and (C.20). Moreover, $r_{t,n} = \sum_{j>n} \Pi_j y_{t-j}^*$ and Lemma C.1 implies $\sum_{j>n} |\Pi_j| = o(T^{-1})$, so the second term is negligible. Consequently,

$$\frac{1}{T} \sum_{t=n+1}^{T} (\widehat{U}_t - \eta_t)^2 = O_p(n/T),$$

which proves (4.6).

The argument can equivalently be written for sample-centered observations. With an intercept included in the first-step regression, subtracting a common sample mean changes only the intercept parameterization and leaves the slope estimates, residuals, and criterion differences unchanged. □

Proposition 4.2 (Second-step Estimates)

Proof of Proposition 4.2. Let $a_T := \sqrt{n_T/T}$, $\Delta_t := \widehat{U}_t - \eta_t$, and recall that $Z_{t-1} = [y_{t-1}^*, \dots, y_{t-p}^*, -\eta_{t-1}, \dots, -\eta_{t-m}]$ and $\widehat{Z}_{t-1} = [y_{t-1}^*, \dots, y_{t-p}^*, -\widehat{U}_{t-1}, \dots, -\widehat{U}_{t-m}]$. We can write

$$D_{t-1} := \widehat{Z}_{t-1} - Z_{t-1} = [0, \dots, 0, -\Delta_{t-1}, \dots, -\Delta_{t-m}].$$

Since $y_t^* = Z_{t-1}\gamma + \eta_t$, the regression evaluated at the estimated regressors satisfies

$$y_t^* - \widehat{Z}_{t-1}\gamma = \eta_t + (Z_{t-1} - \widehat{Z}_{t-1})\gamma = \eta_t + \sum_{j=1}^{m} \theta_j^* \Delta_{t-j}. \tag{C.22}$$

The second term in (C.22) is the generated-regressor error caused by replacing the unobserved

innovations with first-step residuals.

First, consider the first-step residual variance. Proposition 4.1 gives

$$\frac{1}{T}\sum_{t=n_T+1}^{T}\Delta_t^2 = O_p(a_T^2) = o_p(1).$$

Writing $\widehat{U}_t^2 - \eta_t^2 = (\eta_t + \Delta_t)^2 - \eta_t^2 = 2\eta_t\Delta_t + \Delta_t^2$, then by the Cauchy–Schwarz inequality,

$$\left|\frac{1}{T}\sum_{t=n_T+1}^{T}\left(\widehat{U}_t^2 - \eta_t^2\right)\right| \le 2\left(\frac{1}{T}\sum_{t=n_T+1}^{T}\eta_t^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=n_T+1}^{T}\Delta_t^2\right)^{1/2} + \frac{1}{T}\sum_{t=n_T+1}^{T}\Delta_t^2 = o_p(1).$$

The first term is $o_p(1)$ as for $n_T/T \to 0$, stationarity and ergodicity imply $\frac{1}{T}\sum_{t=n_T+1}^{T}\eta_t^2 \xrightarrow{p} \sigma_\eta^2$. Hence $(\frac{1}{T}\sum_{t=n_T+1}^{T}\eta_t^2)^{1/2} = O_p(1)$, and we can conclude that

$$\widehat{\sigma}_{\eta,1}^2 \xrightarrow{p} \sigma_\eta^2. \tag{C.23}$$

Second, we need to show that $\hat{J}_{(2)}$ converges to $J_{(2)}$ in probability. Consider the Step-2 Gram matrix. Since m is fixed and the nonzero entries of D_{t-1} are lagged values of Δ_t, $\frac{1}{T}\sum_{t=\ell}^{T}\|D_{t-1}\|_2^2 = O_p(a_T^2) = o_p(1)$. Using $\widehat{Z}_{t-1} = Z_{t-1} + D_{t-1}$, simply expand to get

$$\widehat{Z}_{t-1}'\widehat{Z}_{t-1} - Z_{t-1}'Z_{t-1} = Z_{t-1}'D_{t-1} + D_{t-1}'Z_{t-1} + D_{t-1}'D_{t-1},$$

Applying Cauchy–Schwarz gives $\left\|\frac{1}{T}\sum_{t=\ell}^{T}\left(\widehat{Z}_{t-1}'\widehat{Z}_{t-1} - Z_{t-1}'Z_{t-1}\right)\right\|_2 = o_p(1)$. Moreover, as $\ell/T \to 0$, stationarity and ergodicity imply $\frac{1}{T}\sum_{t=\ell}^{T}Z_{t-1}'Z_{t-1} \xrightarrow{p} E[Z_{t-1}'Z_{t-1}]$. Combining this result with (C.23) yields $\widehat{J}_{(2)} = \frac{1}{T}\sum_{t=\ell}^{T}\widehat{Z}_{t-1}'(\widehat{\sigma}_{\eta,1}^2)^{-1}\widehat{Z}_{t-1} \xrightarrow{p} J_{(2)}$. The identification argument in Appendix B.2 shows that $J_{(2)}$ is positive definite. Therefore,

$$\|\widehat{J}_{(2)}^{-1}\|_2 = O_p(1). \tag{C.24}$$

Third, we establish the rate of the second-step estimator. Substituting (C.22) into the Step-2 normal equations gives

$$\widetilde{\gamma} - \gamma = \widehat{J}_{(2)}^{-1}\left(A_{2T} + G_{2T}\right), \tag{C.25}$$

where

$$A_{2T} := \frac{1}{T}\sum_{t=\ell}^{T}\widehat{Z}_{t-1}'(\widehat{\sigma}_{\eta,1}^2)^{-1}\eta_t, \qquad G_{2T} := \frac{1}{T}\sum_{t=\ell}^{T}\widehat{Z}_{t-1}'(\widehat{\sigma}_{\eta,1}^2)^{-1}(Z_{t-1} - \widehat{Z}_{t-1})\gamma.$$

Here A_{2T} is the innovation-score term and G_{2T} is the generated-regressor term.

For the first term, if the true regressor Z_{t-1} is available, then $\frac{1}{T}\sum_{t=\ell}^{T}Z_{t-1}'(\sigma_\eta^2)^{-1}\eta_t = O_p(T^{-1/2})$, as the population score has zero mean and summable autocovariances. With $\widehat{Z}_{t-1}$,

$$\left\|\frac{1}{T}\sum_{t=\ell}^{T}D_{t-1}'\eta_t\right\|_2 \le \left(\frac{1}{T}\sum_{t=\ell}^{T}\|D_{t-1}\|_2^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=\ell}^{T}\eta_t^2\right)^{1/2} = O_p(a_T)O_p(1) = O_p(a_T).$$

Together with variance consistency, these bounds imply $A_{2T} = O_p(a_T)$. Similarly,

$$\|G_{2T}\|_2 \le (\widehat{\sigma}_{\eta,1}^2)^{-1}\left(\frac{1}{T}\sum_{t=\ell}^{T}\|\widehat{Z}_{t-1}\|_2^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=\ell}^{T}\|(Z_{t-1} - \widehat{Z}_{t-1})\gamma\|_2^2\right)^{1/2} = O_p(a_T).$$

Combining these results with (C.25) and (C.24) gives $\|\widetilde{\gamma}-\gamma\|_2 = O_p(a_T) = O_p\Big(\sqrt{n_T/T}\Big)$. In particular, $\widetilde{\gamma} \xrightarrow{p} \gamma$. Moreover, $\frac{a_T}{T^{-1/4}} = \left(\frac{n_T^2}{T}\right)^{1/4} \longrightarrow 0$, so $\|\widetilde{\gamma}-\gamma\|_2 = o_p(T^{-1/4})$, proves (4.7).

Having established the consistency and convergence rate of $\widetilde{\gamma}$, we next derive its root-T asymptotic distribution. The objective is to show that, after multiplication by $\sqrt{T}$, the feasible Step-2 normal equation can be replaced by its population counterpart, $J_{(2)}^{-1}\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T} Z_{t-1}'(\sigma_\eta^2)^{-1}\eta_t$, up to an $o_p(1)$ remainder. This requires controlling separately the error from replacing Z_{t-1} by $\widehat{Z}_{t-1}$ in the innovation score and the generated-regressor term arising because $\widehat{Z}_{t-1}$ itself contains estimated first-step residuals.

We first consider the replacement of $\widehat{Z}_{t-1}$ by Z_{t-1} in the innovation score. For $h = 1, \dots, m$,

$$\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T}\Delta_{t-h}\eta_t = -(\widehat{\beta}_n-\beta_n)'\left\{\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T}x_{t-h}\eta_t\right\} + \frac{1}{\sqrt{T}}\sum_{t=\ell}^{T} r_{t-h,n}\eta_t.$$

By Lemma C.3, $\left\|\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T}x_{t-h}\eta_t\right\|_2 = O_p(\sqrt{n_T})$. Since $\|\widehat{\beta}_n-\beta_n\|_2 = O_p\Big(\sqrt{n_T/T}\Big)$, the first term is $O_p(n_T/\sqrt{T}) = o_p(1)$. The second term is negligible because $\sum_{j>n_T}|\Pi_j| = o(T^{-1})$ and the product-moment bound is uniform over lag indices. Therefore,

$$\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T}(\widehat{Z}_{t-1}-Z_{t-1})'(\sigma_\eta^2)^{-1}\eta_t = o_p(1). \tag{C.26}$$

The generated-regressor term is different. At the root-T scale,

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

Thus condition (4.9), together with (C.23), makes this term $o_p(1)$. Using (C.26), variance consistency, and Gram convergence, the exact estimator decomposition becomes

$$\sqrt{T}(\widetilde{\gamma}-\gamma) = J_{(2)}^{-1}\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T}s_{2t} + o_p(1). \tag{C.27}$$

It remains to establish the limit distribution of the population score. Since Z_{t-1} contains only past lags of y_t^* and η_t, it is a stable, strictly past linear filter of η_t. Consequently, it can be written as

$$Z_{t-1} = \sum_{h=1}^{\infty} c_h\eta_{t-h}, \qquad \|c_h\|_2 \le C\rho^h, \qquad 0<\rho<1.$$

The reduced-form regularity results in Appendix B.1 imply that $s_{2t} = Z_{t-1}'(\sigma_\eta^2)^{-1}\eta_t$ is strictly stationary, geometrically mixing, and has a finite $(2+\zeta)$th moment.

The strong-mixing central limit theorem, as used in Lemma D.2 of Dufour and Pelletier (2022), therefore gives

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s_{2t} \xrightarrow{d} N(0, I_{(2)}), \qquad I_{(2)} = \sum_{j=-\infty}^{\infty}\mathrm{E}[s_{2t}s_{2,t-j}']$$

The removal of the first $\ell-1$ observations does not change the limit because $\ell/T \to 0$ and the score autocovariances are absolutely summable. Hence $\frac{1}{\sqrt{T}}\sum_{t=\ell}^{T}s_{2t} \xrightarrow{d} N(0, I_{(2)})$. Combining

this result with (C.27) proves (4.10).

Finally, consider estimation of $I_{(2)}$. Let $\widehat{I}^0_{(2)}$ denote the Bartlett HAC expression in (4.11) with the feasible scores $\widehat{s}_{2t}$ replaced by the population scores s_{2t}. The mixing and moment conditions established above, together with $m_T \longrightarrow \infty$ and $m_T^4/T \longrightarrow 0$ imply $\widehat{I}^0_{(2)} \xrightarrow{p} I_{(2)}$.

It remains to show that replacing the population scores by their feasible counterparts does not affect this limit. Recall $a_T = \sqrt{n_T/T}$. The first-step residual rate gives $\left\{\frac{1}{T}\sum_{t=\ell}^{T} \|\widehat{Z}_{t-1} - Z_{t-1}\|_2^2\right\}^{1/2} = O_p(a_T)$. The second-step starting rate, together with the local-filter regularity in Appendix B.2, similarly gives $\left\{\frac{1}{T}\sum_{t=\ell}^{T}(\widetilde{U}_t - \eta_t)^2\right\}^{1/2} = O_p(a_T)$. Moreover, the variance-consistency argument above gives $\widehat{\sigma}^2_{\eta,1} - \sigma^2_\eta = O_p(a_T)$. Using these bounds and the finite moments of the population regressors and innovations, $\left\{\frac{1}{T}\sum_{t=\ell}^{T} \|\widehat{s}_{2t} - s_{2t}\|_2^2\right\}^{1/2} = O_p(a_T)$.

Applying Cauchy–Schwarz to the difference between each feasible and population sample autocovariance, and then summing over the $2m_T + 1$ included lags, gives $\|\widehat{I}_{(2)} - \widehat{I}^0_{(2)}\|_2 = O_p(m_T a_T)$. The maintained bandwidth condition $m_T n_T/\sqrt{T} \to 0$ implies $m_T a_T = m_T\sqrt{n_T/T} \to 0$. Therefore, $\widehat{I}_{(2)} \xrightarrow{p} I_{(2)}$. Together with $\widehat{J}_{(2)} \xrightarrow{p} J_{(2)}$, this proves that, under (4.9), $\widehat{J}^{-1}_{(2)}\widehat{I}_{(2)}\widehat{J}^{-1}_{(2)} \xrightarrow{p} J^{-1}_{(2)} I_{(2)} J^{-1}_{(2)}$. The score-equivalence condition is required for interpreting this sandwich matrix as the asymptotic covariance of $\sqrt{T}(\widetilde{\gamma} - \gamma)$. It is not required for the consistency of $\widetilde{\gamma}$, its starting rate, $\widehat{J}_{(2)} \to_p J_{(2)}$, or $\widehat{I}_{(2)} \to_p I_{(2)}$. □

Proposition 4.3 (Third-step Estimates)

Proof of Proposition 4.3. Let $d_T := \widetilde{\gamma} - \gamma$ and $a_T := \sqrt{n_T/T}$. The unconditional second-step result gives $\|d_T\|_2 = O_p(a_T)$. By Appendix B.2, $\widetilde{\gamma}$ belongs, with probability tending to one, to a neighborhood on which the relevant inverse filters are uniformly stable.

First, we rewrite the filtering equations. Denote by $\mathcal{F}_0$ and $\widetilde{\mathcal{F}}_0$ the inverse filters $\theta^*(B)^{-1}$ and $\widetilde{\theta}^*(B)^{-1}$, respectively, with zero initialization at $t \le m$. Thus

$$X_t = \widetilde{\mathcal{F}}_0 y_t^*, \qquad W_t = \widetilde{\mathcal{F}}_0 \widetilde{U}_t, \qquad \widetilde{V}_{t-1} = \widetilde{\mathcal{F}}_0 \widetilde{Z}_{t-1}.$$

Since $y_t^* - \widetilde{U}_t = \widetilde{Z}_{t-1}\widetilde{\gamma}$, the common initialization gives $X_t - W_t = \widetilde{V}_{t-1}\widetilde{\gamma}$.

To identify the remaining regression error, put $e_t := \widetilde{U}_t - \eta_t$. Subtracting the true and estimated innovation recursions gives $e_t = -\widetilde{Z}_{t-1}d_T + \sum_{j=1}^{m}\theta_j^* e_{t-j}$ for $t \ge \ell'$. Consequently, $e_t = -\mathcal{F}_0(\widetilde{Z}_{t-1}d_T) + b_t$, where b_t is the initialization error. It satisfies the homogeneous recursion with coefficients θ_j^* and initial values $b_t = -\eta_t$ for $t = 1, \ldots, m$. Invertibility therefore makes b_t decay geometrically.

Writing $Y_t^{(3)} := \widetilde{U}_t + X_t - W_t$, we obtain

$$Y_t^{(3)} = \widetilde{U}_t + \widetilde{V}_{t-1}(\gamma + d_T) = \widetilde{V}_{t-1}\gamma + \eta_t + b_t + R_{tT}, \tag{C.28}$$

where $R_{tT} := (\widetilde{\mathcal{F}}_0 - \mathcal{F}_0)(\widetilde{Z}_{t-1} d_T)$. This expression shows the first-order cancellation: the remaining error is the product of an inverse-filter error and a parameter error.

Indeed, the inverse-filter identity gives $\widetilde{\mathcal{F}}_0 - \mathcal{F}_0 = \widetilde{\mathcal{F}}_0 \left\{ \sum_{j=1}^m (\widetilde{\theta}_j^* - \theta_j^*) B^j \right\} \mathcal{F}_0$. The two inverse filters are uniformly bounded, and the filtered regressors have bounded sample second moments. Hence

$$\left\{ \frac{1}{T} \sum_{t=\ell'}^{T} R_{tT}^2 \right\}^{1/2} = O_p(\|d_T\|_2^2) = O_p(n_T/T). \tag{C.29}$$

Next, local stable-filter continuity and $\widetilde{\gamma} \to_p \gamma$ imply that $\widetilde{U}_t - \eta_t$ and $\widetilde{V}_{t-1} - V_{t-1}$ vanish in sample mean square. Using the same Cauchy–Schwarz and ergodic arguments as in the second-step proof, $\widetilde{\sigma}_\eta^2 \xrightarrow{p} \sigma_\eta^2$ and $\widehat{J}_{(3)} \xrightarrow{p} J_{(3)}$. Identification gives $J_{(3)} > 0$, so $\|\widehat{J}_{(3)}^{-1}\|_2 = O_p(1)$.

We now substitute (C.28) into the third-step normal equations. The remainder is negligible at the root-T scale because

$$\left\| \frac{1}{\sqrt{T}} \sum_{t=\ell'}^{T} \widetilde{V}_{t-1}' R_{tT} \right\|_2 \le \left\{ \frac{1}{T} \sum_{t=\ell'}^{T} \|\widetilde{V}_{t-1}\|_2^2 \right\}^{1/2} \left\{ \sum_{t=\ell'}^{T} R_{tT}^2 \right\}^{1/2} = O_p(n_T/\sqrt{T}) = o_p(1).$$

Geometric decay and the finite moments similarly give $\sum_{t=\ell'}^{T} \|\widetilde{V}_{t-1}\|_2 |b_t| = O_p(1)$, so the initialization contribution is also negligible. Therefore,

$$\sqrt{T}(\widehat{\gamma} - \gamma) = \widehat{J}_{(3)}^{-1} \frac{1}{\sqrt{T}} \sum_{t=\ell'}^{T} \widetilde{V}_{t-1}' (\widetilde{\sigma}_\eta^2)^{-1} \eta_t + o_p(1).$$

It remains to replace $\widetilde{V}_{t-1}$ by V_{t-1} in this score. The difference between their stationary filter versions is a strictly past linear filter of y_t^* whose lag-h coefficient is bounded by $C\|d_T\|_2 \rho^h$, for some $0 < \rho < 1$. For $S_{hT} := \frac{1}{\sqrt{T}} \sum_{t=\ell'}^{T} y_{t-h}^* \eta_t$ where $h \ge 1$, Lemma C.3 gives $\sup_{h,T} E|S_{hT}|^2 < \infty$. Thus $\sum_{h \ge 1} \rho^h |S_{hT}| = O_p(1)$. After accounting for the geometrically decaying initialization terms, it follows that $\frac{1}{\sqrt{T}} \sum_{t=\ell'}^{T} (\widetilde{V}_{t-1} - V_{t-1})' \eta_t = o_p(1)$. Thus estimation of the filter has no first-order effect on the innovation score.

Combining this replacement with variance and Gram convergence yields

$$\sqrt{T}(\widehat{\gamma} - \gamma) = J_{(3)}^{-1} \frac{1}{\sqrt{T}} \sum_{t=\ell'}^{T} s_{3t} + o_p(1). \tag{C.30}$$

The regularity results in Appendix B.1 apply to the population score s_{3t}, which has zero mean, geometric mixing, and the required moments. The same strong-mixing central limit theorem used for the second-step score therefore gives $\frac{1}{\sqrt{T}} \sum_{t=\ell'}^{T} s_{3t} \xrightarrow{d} N(0, I_{(3)})$. Equation (4.12) follows by Slutsky's theorem. In particular, $\widehat{\gamma} \to_p \gamma$. Only the unconditional second-step starting rate has been used; condition (4.9) is not required.

Finally, consider estimation of $I_{(3)}$. Let $\widehat{I}_{(3)}^0$ denote the Bartlett estimator in (4.14) formed from the population scores s_{3t}. The population HAC argument used in the second-step proof gives $\widehat{I}_{(3)}^0 \xrightarrow{p} I_{(3)}$ under $m_T \to \infty$ and $m_T^4/T \to 0$.

For the feasible-score replacement, the fixed-dimensional local filter bounds have finite

fourth-moment envelopes under the maintained SV assumptions. Applying these bounds to products, using $\|d_T\|_2 = O_p(a_T)$ and $\|\widehat{\gamma} - \gamma\|_2 = O_p(T^{-1/2})$, gives

$$\left\{ \frac{1}{T} \sum_{t=\ell'}^{T} \left\| \widetilde{V}'_{t-1} \bar{U}_t - V'_{t-1} \eta_t \right\|_2^2 \right\}^{1/2} = O_p(a_T).$$

Here the initialization terms are handled separately using their geometric decay.

The variance weight is a common scalar and can be factored out of the HAC sum. Cauchy–Schwarz applied to each lag covariance therefore gives $\left\| \widehat{I}_{(3)} - \left(\frac{\sigma_\eta^2}{\widetilde{\sigma}_\eta^2} \right)^2 \widehat{I}^0_{(3)} \right\|_2 = O_p(m_T a_T)$. The maintained restrictions imply

$$m_T a_T = m_T \sqrt{n_T/T} = \left(\frac{m_T^4}{T} \right)^{1/4} \left(\frac{n_T^2}{T} \right)^{1/4} \longrightarrow 0.$$

Since $\widetilde{\sigma}_\eta^2 \to_p \sigma_\eta^2 > 0$, we conclude that $\widehat{I}_{(3)} \to_p I_{(3)}$. Together with Gram convergence, this proves consistency of $\widehat{J}_{(3)}^{-1} \widehat{I}_{(3)} \widehat{J}_{(3)}^{-1}$ for the covariance matrix in (4.12). Dividing by T gives the stated asymptotic sampling-covariance estimator.

□

D Extended model selection results

The same bootstrap covariance estimate is used to weight the adjusted moments for every candidate structural order. In finite samples, an estimated covariance matrix can be nearly singular, in which case its inverse may assign excessive weight to small numerical errors. We therefore regularize only numerically small eigenvalues before inversion. This is a numerical safeguard and it does not change the moment conditions, the set of candidate orders, or the penalty. The same regularized weighting matrix is used for every candidate q.

Structural invertibility is imposed directly through the parameterization. For $q \geq 1$, the profile criterion is minimized from deterministic initial values and 12 random starting values, and the best converged admissible solution is retained. The $q = 0$ profile is obtained analytically. The use of multiple starting values reduces the possibility that an order is evaluated using a local, rather than global, minimum of the nonlinear objective.

The bootstrap covariance estimate is based on $B = 199$ valid residual circular-block bootstrap samples with block length $L = 5$. Initial fitted innovations are discarded to reduce initialization effects, the retained innovations are centered, and each bootstrap path is generated with a burn-in of 500 observations. All 5,000 Monte Carlo replications in the formal experiment attained the target of 199 valid bootstrap samples. Thus, none of the reported order selections relied on an incomplete bootstrap covariance estimate.

Appendix Table D.1 reports the complete selection-frequency matrices corresponding

to the structural summary in Table 1. Panel A reports the joint structural (p, q) selection frequencies, while Panel B reports the corresponding reduced-form (p, m) selection frequencies. The additional reduced-form panel helps locate the source of the off-diagonal selections in Panel A. For replications with $\hat{m} > \hat{p}$, the structural order follows directly from $\hat{q} = \hat{m}$, so structural-order errors are inherited from the preliminary ARMA selection. For replications with $\hat{m} = \hat{p}$, the comparison instead reflects the additional role of the bootstrap-weighted profile criterion in selecting among the structural orders compatible with the reduced form.

All 5,000 Monte Carlo replications produced resolved reduced-form and structural orders, and the selected structural profile converged in every case. The diagnostics reported below therefore concern the numerical regularity and local precision of the selected parameter profiles rather than failures to select an order.

E Extended simulation results

Tables E.1 and E.2 report the complete estimator and tuning grids corresponding to Tables 2 and 3. They add S-Moment, S-ARMA-Restricted, alternative values of J for the S-ARMA-W estimators, alternative values of n_T for HR, and the associated success counts. Across both designs, HR performance changes little across the reported values of n_T, whereas the effect of J varies across the S-ARMA-W estimators, and S-Moment shows substantially lower admissibility and larger estimation errors in the more challenging Design 2.

The remaining subsections examine the estimators under heavy-tailed Student-t return innovations and provide additional evidence for higher-order SV(p, q) specifications, including sensitivity to the preliminary HR order and to the magnitude of the structural MA coefficient.

E.1 Student-t return innovations

We repeat Design 1 with standardized Student-t return innovations to assess robustness to heavy-tailed returns. Specifically, z_t follows a Student-t distribution with five degrees of freedom, rescaled to have unit variance:

$$z_t = \frac{\tilde{z}_t}{\sqrt{5/3}}, \qquad \tilde{z}_t \sim t_5.$$

The latent volatility process and the distribution of its innovation v_t are unchanged. Thus, $E[z_t] = 0$ and $\mathrm{Var}(z_t) = 1$. The log-squared transformation remains

$$y_t^* = w_t + \epsilon_t, \qquad \epsilon_t = \log(z_t^2) - c_z.$$

For standardized Student-t_ν innovations with $\nu > 4$, $\kappa_z = \frac{3(\nu-2)}{\nu-4}$. Thus, in the Student-$t_5$ simulations, the S-Moment estimator uses $\kappa_z = 9$ directly. For standardized Student-t_ν

Table D.1: Model-selection frequency matrices for structural SV(p, q) and reduced-form ARMA(p, m) models.

			Selected order (proportions over 500 replications)								
Scenario	True model	Parameters	$(0,0)$	$(0,1)$	$(0,2)$	$(1,0)$	$(1,1)$	$(1,2)$	$(2,0)$	$(2,1)$	$(2,2)$
Panel A: Structural order selection for $y_t \sim \text{SV}(p, q)$											
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**
Panel B: Reduced-form order selection for $y_t^* \sim \text{ARMA}(p, m)$											
S1	ARMA$(0,1)$		0.000	**0.988**	0.012	0.000	0.000	0.000	0.000	0.000	0.000
S2	ARMA$(0,1)$		0.000	**0.988**	0.012	0.000	0.000	0.000	0.000	0.000	0.000
S3	ARMA$(0,2)$		0.000	0.000	**0.944**	0.000	0.056	0.000	0.000	0.000	0.000
S4	ARMA$(1,1)$[a]		0.000	0.000	0.000	0.000	**0.994**	0.006	0.000	0.000	0.000
S5	ARMA$(1,1)$[a]		0.000	0.000	0.000	0.000	**0.984**	0.016	0.000	0.000	0.000
S6	ARMA$(1,1)$		0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000
S7	ARMA$(1,2)$		0.000	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000	0.000
S8	ARMA$(2,2)$[b]		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**
S9	ARMA$(2,2)$[b]		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**1.000**
S10	ARMA$(2,2)$		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 order. All unresolved frequencies are zero and are therefore omitted. Simulations are conducted directly for 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 $c = 1$, $\delta_{\text{IC}} = 0.1$, $\delta_{\text{AR}} = 0.05$, $n_T = \lceil c(\log T)^{1+\delta_{\text{AR}}} \rceil$, and penalty $(\log T)^{1+\delta_{\text{IC}}}/T$.

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 $IC_T(q) = T\left[\hat{C} - c_q(\hat{\beta}(q))\right]' \hat{\Omega}_C^{-1} \left[\hat{C} - c_q(\hat{\beta}(q))\right] + (q+1)\log T$. The residual circular block bootstrap uses $B = 199$ valid replications and block length $L = 5$. The bootstrap weighting matrix is also used to recover the selected structural parameter profile when $\hat{m} > \hat{p}$, although the structural order is then determined directly.

All 5,000 reduced-form and structural orders are resolved, all selected parameter profiles converge, and every bootstrap produces 199 valid replications.

[a] Under $m = \max(p, q)$, SV$(1,0)$ implies a reduced-form ARMA$(1,1)$.

[b] Under $m = \max(p, q)$, SV$(2,0)$ and SV$(2,1)$ imply a reduced-form ARMA$(2,2)$.

Table E.1: Extended Monte Carlo results for SV(1, 1): Design 1

$(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.7, 0.1, 2.5, 2.5)$										
	$T = 1000$					$T = 5000$				
Estimator	ϕ_1	θ_1	σ_v^2	σ_y	Success	ϕ_1	θ_1	σ_v^2	σ_y	Success
Panel A. Bias										
S-Moment	-0.114	-0.073	-0.798	1.163		-0.018	-0.013	-0.657	0.804	
S-ARMA-based	-0.003	-0.026	-0.055	0.016		0.000	-0.007	-0.020	0.001	
S-ARMA-Restricted	-0.002	-0.022	-0.056	0.017		0.000	-0.007	-0.020	0.001	
S-ARMA-W-Mean ($J = 3$)	-0.001	-0.021	-0.043	0.017		0.001	-0.005	-0.017	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-Mean ($J = 10$)	0.029	0.067	-0.169	0.017		0.034	0.064	-0.109	0.000	
S-ARMA-W-LDMean ($J = 3$)	-0.002	-0.024	-0.036	0.017		0.000	-0.006	-0.016	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-LDMean ($J = 10$)	0.023	0.040	-0.102	0.016		0.012	0.016	-0.055	0.000	
S-ARMA-W-Median ($J = 3$)	-0.007	-0.031	-0.045	0.017		0.000	-0.007	-0.018	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-Median ($J = 10$)	-0.030	-0.066	-0.047	0.015		-0.009	-0.022	-0.023	0.001	
S-ARMA-W-OLS ($J = 3$)	-0.014	-0.046	-0.033	0.017		-0.002	-0.010	-0.015	0.001	
S-ARMA-W-OLS ($J = 5$)	-0.017	-0.052	-0.030	0.017		-0.003	-0.012	-0.015	0.001	
S-ARMA-W-OLS ($J = 10$)	-0.022	-0.060	-0.028	0.017		-0.004	-0.014	-0.014	0.001	
HR ($n_T = 3$)	-0.020	-0.053	-0.031	0.005		-0.002	-0.010	-0.014	0.002	
HR ($n_T = 8$)	-0.018	-0.050	-0.029	0.005		-0.002	-0.009	-0.014	0.002	
HR ($n_T = 10$)	-0.017	-0.048	-0.028	0.005		-0.002	-0.009	-0.014	0.002	
Panel B. RMSE										
S-Moment	0.2910	0.4124	1.1306	1.2582	995	0.1700	0.3551	0.9798	0.8807	1285
S-ARMA-based	0.0988	0.2407	0.6144	0.2119	1955	0.0430	0.1021	0.2827	0.0909	2000
S-ARMA-Restricted	0.1004	0.2468	0.6138	0.2117	1963	0.0430	0.1021	0.2827	0.0909	2000
S-ARMA-W-Mean ($J = 3$)	0.0865	0.2112	0.6154	0.2115	1962	0.0357	0.0850	0.2831	0.0909	2000
S-ARMA-W-Mean ($J = 5$)	0.1204	0.2539	0.6169	0.2109	1826	0.0498	0.1057	0.2864	0.0909	1999
S-ARMA-W-Mean ($J = 10$)	0.2117	0.3954	0.6359	0.2133	999	0.1517	0.2939	0.3177	0.0921	1461
S-ARMA-W-LDMean ($J = 3$)	0.0705	0.1956	0.6137	0.2119	1971	0.0297	0.0802	0.2827	0.0909	2000
S-ARMA-W-LDMean ($J = 5$)	0.0848	0.2146	0.6184	0.2127	1911	0.0324	0.0820	0.2833	0.0909	2000
S-ARMA-W-LDMean ($J = 10$)	0.1761	0.3276	0.6260	0.2121	1378	0.1093	0.2112	0.3039	**0.0906**	1801
S-ARMA-W-Median ($J = 3$)	0.0888	0.2164	0.6182	0.2116	1968	0.0384	0.0906	0.2832	0.0909	2000
S-ARMA-W-Median ($J = 5$)	0.0999	0.2288	0.6214	0.2120	1955	0.0413	0.0930	0.2838	0.0909	2000
S-ARMA-W-Median ($J = 10$)	0.1312	0.2652	0.6236	0.2118	1905	0.0643	0.1276	0.2878	0.0909	1997
S-ARMA-W-OLS ($J = 3$)	0.0700	0.2010	0.6209	0.2119	1970	0.0297	0.0813	0.2831	0.0909	2000
S-ARMA-W-OLS ($J = 5$)	0.0690	0.1979	0.6218	0.2119	1969	0.0288	0.0797	0.2834	0.0909	2000
S-ARMA-W-OLS ($J = 10$)	0.0692	0.1969	0.6246	0.2121	1970	0.0288	0.0795	0.2837	0.0909	2000
HR ($n_T = 3$)	**0.0689**	0.1935	0.6004	0.2084	1966	**0.0282**	**0.0758**	0.2787	0.0927	2000
HR ($n_T = 8$)	0.0690	0.1924	0.5987	**0.2081**	1965	0.0283	**0.0758**	0.2787	0.0927	2000
HR ($n_T = 10$)	0.0692	**0.1914**	**0.5978**	**0.2081**	1964	0.0283	**0.0758**	**0.2786**	0.0927	2000

Notes. Each sample size uses $N = 2000$ simulated datasets. Bias and RMSE are computed conditional on successful estimation. Success is the number of successful replications for the estimator and tuning choice shown. Bold entries indicate the smallest RMSE in each parameter column at the reported precision, including ties.

Table E.2: Extended Monte Carlo results for the $SV(1,1)$ model: Design 2

$(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.9, -0.3, 2.5, 2.5)$										
	$T = 1000$					$T = 5000$				
Estimator	ϕ_1	θ_1	σ_v^2	σ_y	Success	ϕ_1	θ_1	σ_v^2	σ_y	Success
Panel A. Bias										
S-Moment	-0.1476	0.3629	-0.5915	53.7674		-0.0773	0.3719	-0.4897	56.5597	
S-ARMA-based	-0.0016	0.0516	0.3131	0.1437		-0.0005	-0.0220	-0.0083	0.0195	
S-ARMA-Restricted	-0.0016	0.0516	0.3131	0.1437		-0.0005	-0.0220	-0.0083	0.0195	
S-ARMA-W-Mean ($J = 3$)	-0.0031	0.0431	0.3189	0.1538		-0.0007	-0.0277	-0.0188	0.0208	
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-Mean ($J = 10$)	-0.0027	0.0527	0.3302	0.1709		-0.0012	-0.0280	-0.0144	0.0178	
S-ARMA-W-LDMean ($J = 3$)	-0.0031	0.0408	0.3124	0.1549		-0.0007	-0.0266	-0.0163	0.0211	
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-LDMean ($J = 10$)	-0.0035	0.0436	0.3252	0.1640		-0.0011	-0.0269	-0.0127	0.0184	
S-ARMA-W-Median ($J = 3$)	-0.0026	0.0444	0.3123	0.1586		-0.0006	-0.0223	-0.0086	0.0193	
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-Median ($J = 10$)	-0.0032	0.0458	0.3214	0.1710		-0.0012	-0.0279	-0.0145	0.0194	
S-ARMA-W-OLS ($J = 3$)	-0.0033	0.0429	0.3213	0.1488		-0.0007	-0.0268	-0.0166	0.0210	
S-ARMA-W-OLS ($J = 5$)	-0.0035	0.0423	0.3234	0.1562		-0.0009	-0.0261	-0.0139	0.0191	
S-ARMA-W-OLS ($J = 10$)	-0.0035	0.0397	0.3154	0.1656		-0.0011	-0.0276	-0.0148	0.0195	
HR ($n_T = 3$)	-0.0039	0.0342	0.2958	0.1126		-0.0004	-0.0215	-0.0044	0.0272	
HR ($n_T = 8$)	-0.0039	0.0352	0.2978	0.1119		-0.0004	-0.0217	-0.0049	0.0270	
HR ($n_T = 10$)	-0.0039	0.0347	0.2969	0.1117		-0.0004	-0.0217	-0.0049	0.0270	
Panel B. RMSE										
S-Moment	0.2476	0.5335	1.2316	83.8890	509	0.1526	0.5381	1.2217	74.8212	561
S-ARMA-based	0.0202	0.2264	0.8614	0.8763	1491	0.0088	0.1606	0.4968	0.3669	1882
S-ARMA-Restricted	0.0202	0.2264	0.8614	0.8763	1491	0.0088	0.1606	0.4968	0.3669	1882
S-ARMA-W-Mean ($J = 3$)	0.0188	0.2147	0.8651	0.8812	1491	0.0080	0.1605	0.5020	0.3677	1909
S-ARMA-W-Mean ($J = 5$)	0.0193	0.2153	0.8665	0.8842	1479	0.0084	0.1558	0.4964	0.3665	1900
S-ARMA-W-Mean ($J = 10$)	0.0228	0.2174	0.8545	0.8841	1421	0.0102	0.1610	0.4917	**0.3656**	1866
S-ARMA-W-LDMean ($J = 3$)	0.0184	0.2159	0.8640	0.8873	1505	0.0078	0.1578	0.4997	0.3675	1906
S-ARMA-W-LDMean ($J = 5$)	0.0185	0.2089	0.8605	0.8789	1482	0.0079	0.1564	0.4980	0.3665	1904
S-ARMA-W-LDMean ($J = 10$)	0.0203	0.2124	0.8610	0.8814	1463	0.0087	0.1550	0.4916	0.3660	1884
S-ARMA-W-Median ($J = 3$)	0.0192	0.2172	0.8649	0.8816	1497	0.0083	0.1537	0.4916	0.3665	1894
S-ARMA-W-Median ($J = 5$)	0.0197	0.2143	0.8630	0.8780	1475	0.0086	0.1565	0.4958	0.3658	1901
S-ARMA-W-Median ($J = 10$)	0.0220	0.2157	0.8586	0.8910	1447	0.0096	0.1602	0.4937	0.3666	1874
S-ARMA-W-OLS ($J = 3$)	0.0186	0.2127	0.8632	0.8783	1491	0.0078	0.1577	0.4994	0.3677	1906
S-ARMA-W-OLS ($J = 5$)	0.0188	0.2125	0.8637	0.8788	1484	0.0080	0.1535	0.4941	0.3661	1898
S-ARMA-W-OLS ($J = 10$)	0.0196	0.2150	0.8656	0.8834	1479	0.0085	0.1563	0.4944	0.3660	1890
HR ($n_T = 3$)	**0.0181**	0.2053	0.8408	0.8764	1533	**0.0077**	**0.1531**	**0.4898**	0.3710	1900
HR ($n_T = 8$)	**0.0181**	**0.2046**	**0.8406**	0.8762	1532	**0.0077**	0.1539	0.4906	0.3710	1901
HR ($n_T = 10$)	**0.0181**	0.2051	0.8407	**0.8759**	1533	**0.0077**	0.1538	0.4906	0.3710	1901

Notes. Each sample size uses $N = 2000$ simulated datasets. Bias and RMSE are computed conditional on successful estimation. Success is the number of successful replications for the estimator and tuning choice shown. Bold entries indicate the smallest RMSE in each parameter column at the reported precision, including ties.

innovations, two distribution-dependent constants change:

$$c_z = \log(\nu - 2) + \psi(1/2) - \psi(\nu/2), \qquad \sigma_\epsilon^2 = \psi'(1/2) + \psi'(\nu/2), \tag{E.1}$$

where ψ and ψ' denote the digamma and trigamma functions. In the simulations, the log-squared estimators approximate these two constants using large auxiliary samples from the standardized Student-t_5 distribution. The resulting estimate of σ_ϵ^2 is used in structural recovery, and the return scale is recovered from $\widehat{\sigma}_y = \exp\left\{\frac{\widehat{\mu}-\widehat{c}_z}{2}\right\}$. For the S-Moment estimator, the distributional adjustment concerns the fourth moment of the return innovations. More generally, suppose that $\{z_t\}$ is i.i.d., independent of the Gaussian latent volatility process, and standardized so that $E[z_t] = 0$ and $E[z_t^2] = 1$. If $\kappa_z := E[z_t^4] < \infty$ is known, the marginal moments satisfy

$$\mu_2 = \sigma_y^2 \exp(\gamma_0/2), \qquad \mu_4 = \kappa_z \sigma_y^4 \exp(2\gamma_0).$$

Thus, the Gaussian fourth-moment constant 3 is replaced by κ_z in the recovery formulas:

$$\gamma_0 = \log\left(\frac{\mu_4}{\kappa_z \mu_2^2}\right), \qquad \sigma_y = \frac{\kappa_z^{1/4}\mu_2}{\mu_4^{1/4}}. \tag{E.2}$$

For $h \geq 1$, the squared-return cross-moments remain $\mu_{2,2}(h) = \sigma_y^4 \exp(\gamma_0 + \gamma_h)$, so the recovery formula $\gamma_h = \log\{\mu_{2,2}(h)/\mu_2^2\}$ and the subsequent structural parameter recovery are unchanged.

The result is presented in Table E.3. The log-squared estimators remain substantially more stable than S-Moment. For every displayed S-ARMA and HR specification, RMSE decreases from $T = 1000$ to $T = 5000$ for all four reported parameters. At $T = 1000$, S-ARMA-W-OLS gives the lowest RMSE for ϕ_1 and θ_1, while HR gives the lowest RMSE for σ_v^2 and σ_y. At $T = 5000$, HR gives the lowest RMSE for ϕ_1, θ_1, and σ_v^2, whereas several S-ARMA specifications have the smallest reported RMSE for σ_y. S-Moment returns only 382 and 663 admissible estimates at the two sample sizes and has larger RMSE than in the Gaussian Design 1 for every reported parameter.

A mechanical feature of the table is that σ_y has nearly identical RMSE across the S-ARMA family. This is because σ_y is recovered from the sample mean of $\log(y_t^2)$ together with the distribution-specific constant $\mathbb{E}[\log(z_t^2)]$, and therefore does not depend on the reduced-form ARMA estimates conditional on the same accepted replications. Small differences across methods arise only from differences in admissibility screening ("Success"). Overall, the results indicate that the proposed ARMA/HR-based pipeline remains accurate and stable under Student-t shocks, while purely moment-based identification becomes considerably more fragile in heavy-tailed environments.

Table E.3: Monte Carlo results for SV(1, 1) with standardized Student-t return innovations

$(\phi_1, \theta_1, \sigma_v^2, \sigma_y) = (0.7, 0.1, 2.5, 2.5), \quad \nu = 5$										
	$T = 1000$					$T = 5000$				
Estimator	ϕ_1	θ_1	σ_v^2	σ_y	Success	ϕ_1	θ_1	σ_v^2	σ_y	Success
Panel A. Bias										
S-Moment	-0.1621	-0.0759	-1.0926	1.8071		-0.0603	-0.0695	-0.8636	1.2712	
S-ARMA-based	-0.0017	-0.0129	-0.0203	0.0116		0.0004	-0.0012	0.0022	0.0032	
S-ARMA-Restricted	-0.0006	-0.0097	-0.0211	0.0114		0.0004	-0.0012	0.0022	0.0032	
S-ARMA-W-Mean ($J = 3$)	-0.0020	-0.0150	-0.0095	0.0108		-0.0002	-0.0016	0.0068	0.0032	
S-ARMA-W-Mean ($J = 5$)	0.0204	0.0273	-0.0355	0.0135		0.0009	0.0008	0.0024	0.0032	
S-ARMA-W-Mean ($J = 10$)	0.0130	0.0460	-0.1186	0.0112		0.0377	0.0761	-0.0991	0.0028	
S-ARMA-W-LDMean ($J = 3$)	-0.0023	-0.0175	-0.0031	0.0114		-0.0001	-0.0018	0.0075	0.0032	
S-ARMA-W-LDMean ($J = 5$)	0.0083	-0.0012	-0.0176	0.0127		0.0004	-0.0009	0.0070	0.0032	
S-ARMA-W-LDMean ($J = 10$)	0.0218	0.0472	-0.0750	0.0114		0.0070	0.0130	-0.0345	0.0033	
S-ARMA-W-Median ($J = 3$)	-0.0085	-0.0251	-0.0064	0.0105		-0.0016	-0.0039	0.0059	0.0032	
S-ARMA-W-Median ($J = 5$)	-0.0143	-0.0357	-0.0118	0.0121		-0.0026	-0.0058	0.0052	0.0032	
S-ARMA-W-Median ($J = 10$)	-0.0327	-0.0595	-0.0054	0.0113		-0.0109	-0.0197	0.0009	0.0032	
S-ARMA-W-OLS ($J = 3$)	-0.0151	-0.0397	0.0043	0.0115		-0.0027	-0.0063	0.0083	0.0032	
S-ARMA-W-OLS ($J = 5$)	-0.0194	-0.0468	0.0075	0.0113		-0.0037	-0.0082	0.0089	0.0032	
S-ARMA-W-OLS ($J = 10$)	-0.0257	-0.0570	0.0110	0.0113		-0.0051	-0.0105	0.0095	0.0032	
HR ($n_T = 3$)	-0.0175	-0.0590	-0.0532	0.0123		-0.0042	-0.0138	-0.0143	-0.0011	
HR ($n_T = 8$)	-0.0156	-0.0561	-0.0544	0.0114		-0.0040	-0.0139	-0.0169	-0.0021	
HR ($n_T = 10$)	-0.0148	-0.0545	-0.0538	0.0114		-0.0040	-0.0139	-0.0169	-0.0021	
Panel B. RMSE and success counts										
S-Moment	0.3905	0.4349	1.3467	1.9123	382	0.2517	0.3824	1.2994	1.3602	663
S-ARMA-based	0.1040	0.2480	0.6335	0.2113	1942	0.0472	0.1114	0.2954	**0.0942**	2000
S-ARMA-Restricted	0.1054	0.2532	0.6339	0.2111	1949	0.0472	0.1114	0.2954	**0.0942**	2000
S-ARMA-W-Mean ($J = 3$)	0.0929	0.2240	0.6403	0.2109	1951	0.0372	0.0895	0.2961	**0.0942**	2000
S-ARMA-W-Mean ($J = 5$)	0.1229	0.2598	0.6421	0.2129	1784	0.0519	0.1111	0.2946	**0.0942**	1999
S-ARMA-W-Mean ($J = 10$)	0.2133	0.3847	0.6747	0.2136	1015	0.1615	0.3139	0.3345	0.0944	1444
S-ARMA-W-LDMean ($J = 3$)	0.0761	0.2064	0.6388	0.2111	1957	0.0313	0.0857	0.2956	**0.0942**	2000
S-ARMA-W-LDMean ($J = 5$)	0.0939	0.2270	0.6422	0.2119	1895	0.0333	0.0882	0.2961	**0.0942**	2000
S-ARMA-W-LDMean ($J = 10$)	0.1805	0.3368	0.6571	0.2126	1368	0.1144	0.2181	0.3039	0.0944	1794
S-ARMA-W-Median ($J = 3$)	0.0941	0.2268	0.6410	0.2109	1950	0.0413	0.0964	0.2960	**0.0942**	2000
S-ARMA-W-Median ($J = 5$)	0.1058	0.2386	0.6434	0.2114	1941	0.0441	0.1023	0.2968	**0.0942**	2000
S-ARMA-W-Median ($J = 10$)	0.1382	0.2672	0.6454	0.2115	1870	0.0664	0.1335	0.3001	**0.0942**	1997
S-ARMA-W-OLS ($J = 3$)	0.0744	0.2045	0.6427	0.2112	1955	0.0314	0.0868	0.2961	**0.0942**	2000
S-ARMA-W-OLS ($J = 5$)	**0.0719**	0.2000	0.6442	0.2112	1955	0.0302	0.0856	0.2964	**0.0942**	2000
S-ARMA-W-OLS ($J = 10$)	0.0732	**0.1992**	0.6463	0.2112	1953	0.0300	0.0849	0.2967	**0.0942**	2000
HR ($n_T = 3$)	0.0730	0.2110	**0.6267**	0.2054	1956	**0.0291**	**0.0815**	**0.2903**	0.0954	2000
HR ($n_T = 8$)	0.0735	0.2124	0.6271	**0.2052**	1956	**0.0291**	0.0817	0.2905	0.0954	2000
HR ($n_T = 10$)	0.0738	0.2125	0.6268	**0.2052**	1956	**0.0291**	0.0817	0.2905	0.0954	2000

Notes. For each reported estimator and tuning configuration, $N = 2000$ Monte Carlo datasets are generated from the same Student-t_5 data-generating process. Bias and RMSE are computed conditional on successful estimation. Success is the number of retained replications for that configuration. The log-squared estimators use large auxiliary Student-t_5 samples to approximate c_z and σ_ϵ^2, while S-Moment uses $E[z_t^4] = 9$. Bold entries indicate the smallest RMSE in each parameter column at the reported precision, including ties.

Table E.4: HR performance across selected SV(p, q) models under alternative n_T

	M(1): SV$(1,0)$			M(2): SV$(0,1)$			M(3): SV$(2,2)$		
	$n_T = 3$	$n_T = 8$	$n_T = 10$	$n_T = 3$	$n_T = 8$	$n_T = 10$	$n_T = 3$	$n_T = 8$	$n_T = 10$
Panel A. Bias									
ϕ_1	-0.0022	-0.0022	-0.0022	–	–	–	NR	-0.0113	-0.0108
ϕ_2	–	–	–	–	–	–	NR	-0.0014	-0.0009
θ_1	–	–	–	0.0122	0.0128	0.0135	NR	-0.0174	-0.0168
θ_2	–	–	–	–	–	–	NR	0.0081	0.0088
σ_v^2	-0.0086	-0.0084	-0.0084	-0.0202	-0.0207	-0.0258	NR	-0.0385	-0.0388
σ_y	0.0000	0.0000	0.0000	-0.0012	-0.0004	0.0000	NR	0.0004	0.0004
Panel B. RMSE									
ϕ_1	0.0153	0.0153	0.0153	–	–	–	NR	0.0773	0.0775
ϕ_2	–	–	–	–	–	–	NR	0.0773	0.0773
θ_1	–	–	–	0.0874	0.0876	0.0860	NR	0.1243	0.1246
θ_2	–	–	–	–	–	–	NR	0.1235	0.1235
σ_v^2	0.3283	0.3284	0.3284	0.2873	0.2884	0.2819	NR	0.2732	0.2732
σ_y	0.1394	0.1394	0.1394	0.0406	0.0423	0.0417	NR	0.0441	0.0441
Success	*2000*	*2000*	*2000*	*1991*	*1989*	*1990*	NR	*1987*	*1987*

Notes. Each reported configuration uses $N = 2000$ simulated datasets at $T = 5000$. 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. Success is the number of successful replications for the corresponding value of n_T. A dash indicates that the parameter is absent from the model; NR denotes a model–tuning configuration not reported in the simulation output.

E.2 Extended simulations for SV(p, q)

Table E.4 examines sensitivity to the preliminary autoregressive order at $T = 5000$. The reported estimates change little over $n_T \in \{3, 8, 10\}$ for SV$(1, 0)$ and SV$(0, 1)$, and over the available values $n_T \in \{8, 10\}$ for SV$(2, 2)$. The corresponding success counts are also similar across the reported values of n_T. A natural interpretation is that when T is large and the model order is low, the HR regression and subsequent mapping rely primarily on low-order dependence features; increasing n_T beyond what is needed to capture those features does not materially change the estimated reduced-form inputs or the admissible backward mapping.

Table E.5 varies the structural MA coefficient in SV$(0, 1)$ while fixing $n_T = 3$. For $\theta_1 = 0.2$, the estimator is almost always admissible even at $T = 1000$. With $\theta_1 = 0.5$, the success rate rises from 85.20% at $T = 1000$ to 99.55% at $T = 5000$ and reaches 100% at $T = 10000$. The $\theta_1 = 0.8$ design is more difficult: the corresponding success rates are 54.75%, 66.95%, and 73.90%. Conditional RMSE decreases with sample size in all three designs.

The decline in admissibility as θ_1 approaches one is consistent with the structural recovery

restriction. For SV$(0,1)$, let $C = \gamma_{y^*}(0) - \sigma_\epsilon^2$ and $D = -\gamma_{y^*}(1)$. Then

$$C = (1+\theta_1^2)\sigma_v^2, \qquad D = \theta_1\sigma_v^2, \qquad C - 2|D| = \sigma_v^2(1-|\theta_1|)^2.$$

An invertible structural MA representation with positive innovation variance requires $C > 2|D|$. With $\sigma_v^2 = 2.5$, this population margin equals 1.6, 0.625, and 0.1 when $\theta_1 = 0.2$, 0.5, and 0.8, respectively. The smaller margin in the last design makes the structural recovery step more sensitive to sampling error.

Table E.5: HR performance in SV$(0,1)$ under alternative θ_1 and sample sizes (fixed $n_T = 3$)

	$\theta_1 = 0.2$			$\theta_1 = 0.5$			$\theta_1 = 0.8$		
	$T = 1000$	$T = 5000$	$T = 10000$	$T = 1000$	$T = 5000$	$T = 10000$	$T = 1000$	$T = 5000$	$T = 10000$
Panel A. Bias									
θ_1	0.0158	0.0024	0.0009	-0.0010	0.0122	0.0054	-0.1541	-0.0654	-0.0361
σ_v^2	-0.0439	-0.0068	0.0003	0.0490	-0.0202	-0.0067	0.5281	0.2159	0.1195
σ_y	0.0010	-0.0011	-0.0003	-0.0082	-0.0012	-0.0003	-0.0258	-0.0100	-0.0050
Panel B. RMSE									
θ_1	0.1187	0.0480	0.0331	0.1493	0.0874	0.0596	0.2072	0.1198	0.0957
σ_v^2	0.4794	0.2150	0.1528	0.5376	0.2873	0.2039	0.7636	0.3907	0.2892
σ_y	0.1009	0.0439	0.0315	0.0913	0.0406	0.0289	0.0892	0.0389	0.0265
Success	*1993*	*2000*	*2000*	*1704*	*1991*	*2000*	*1095*	*1339*	*1478*

Notes. Each configuration uses $N = 2000$ simulated datasets, $\sigma_v^2 = \sigma_y = 2.5$, and preliminary autoregressive order $n_T = 3$. Bias and RMSE are computed conditional on successful estimation. Success is the number of successful replications.

F Extended empirical results

F.1 Summary statistics

Table F.1 reports summary statistics for S&P 500 price, returns, and transformations.

F.2 Additional model-selection diagnostics

This appendix reports the detailed diagnostics underlying the two-stage model-selection result in Section 7.1. In the first stage, we estimate admissible reduced-form ARMA(p,m) specifications over $0 \le p \le 5$ and $p \le m \le 5$ and select the model minimizing

$$IC_T(p,m) = \log\Big(\hat{\sigma}_\eta^2(p,m)\Big) + (p+m)\frac{(\log T)^{1+\delta_{\mathrm{IC}}}}{T}.$$

Table F.1: Extended summary statistics for S&P 500 price, returns, and transformations (1962–2024, $T = 15{,}731$).

Series	Mean	Std. dev.	Median	Skewness	Kurtosis	Min	Max	Range
p_t	1006.87	1217.46	451.37	1.78	5.88	53.49	6090.27	6036.8
r_t	0.03	1.04	0.05	-0.99	28.17	-22.90	10.96	33.9
y_t	0.00	1.04	0.02	-0.99	28.17	-22.93	10.93	33.9
y_t^2	1.08	5.61	0.23	57.25	4963.10	0.00	525.75	525.8
$\log(\lvert y_t\rvert)$	-0.93	1.25	-0.74	-1.15	5.71	-10.19	3.13	13.3
$\log(y_t^2)$	-1.87	2.50	-1.48	-1.15	5.71	-20.38	6.26	26.6
y_t^*	0.00	2.50	0.39	-1.15	5.71	-18.51	8.13	26.6

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 $y_t^* = \log(y_t^2) - T^{-1}\sum_{s=1}^{T}\log(y_s^2)$.

Panel A of Table F.2 reports the complete criterion grid. The minimum is attained at $(\hat{p}, \hat{m}) = (1, 1)$. Because $\hat{m} = \hat{p} = 1$, the reduced-form order is compatible with both SV$(1, 0)$ and SV$(1, 1)$. The second stage therefore compares $q \in \{0, 1\}$ using the bootstrap-weighted profile criterion $IC_T(q) = J_T(q) + (q + 1)\log T$, where $J_T(q)$ is the covariance-standardized distance between the estimated adjusted-moment vector and the closest moment vector implied by the candidate structural model. Panel B reports this comparison for residual circular-block bootstrap lengths $L = 1, 5$, and 10.

Across all three bootstrap block lengths, the structural criterion selects $q = 1$. The SV$(1, 1)$ candidate reproduces the two adjusted first-order moments essentially exactly, whereas imposing $q = 0$ generates a positive profile discrepancy that is sufficiently large to offset the smaller-model penalty.

G Kalman filtering and multi-step forecasting

Here we document the Kalman filtering and multi-step forecasting procedure used in Section 7.2. At each forecast origin, the reduced-form and structural parameters are first estimated from the expanding training sample using the HR and structural-recovery procedures. The Kalman filter is then used to estimate the latent state and generate forecasts.

The transformed observation equation is

$$y_s^* = w_s + \epsilon_s, \qquad \mathrm{Var}(\epsilon_s) = \frac{\pi^2}{2}.$$

Because the centered log-χ_1^2 disturbance ϵ_s is non-Gaussian, the classical Kalman filter is not exact. We therefore adopt a Gaussian quasi-likelihood (QML) approximation by approximating ϵ_s with a mean-zero Gaussian disturbance, $\epsilon_s^G \sim N(0, H)$, for $H = \frac{\pi^2}{2}$.

Table F.2: Detailed model-selection diagnostics for S&P 500 centered log-squared returns

Panel A. Reduced-form HR–IC values

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
$p=0$	1.8363	1.8248	1.8153	1.8080	1.8021	1.7931
$p=1$	—	**1.7361**	1.7368	1.7376	1.7384	1.7392
$p=2$	—	—	1.7378	—	—	—
$p=3$	—	—	—	—	—	—
$p=4$	—	—	—	—	1.7417	1.7421
$p=5$	—	—	—	—	—	1.7418

Panel B. Structural-order comparison conditional on $(\hat{p}, \hat{m}) = (1, 1)$

L	$J_T(0)$	$J_T(1)$	$IC_T(0)$	$IC_T(1)$	Valid/attempted	$\hat{q}$
1	19.124	0.000	28.787	19.327	199/210	**1**
5	27.250	0.000	36.914	19.327	199/205	**1**
10	22.715	0.000	32.378	19.327	199/201	**1**

Notes: Panel A reports $IC_T(p,m) = \log(\hat{\sigma}_\eta^2) + (p+m)(\log T)^{1+\delta_{\text{IC}}}/T$. Entries marked "—" are outside the maintained candidate set $p \le m$ or correspond to reduced-form estimates that fail the stationarity, invertibility, or numerical admissibility checks. The selected reduced form is ARMA$(1,1)$.

Panel B reports $IC_T(q) = J_T(q) + (q+1)\log T$. The covariance weighting matrix is estimated separately for each block length using a residual circular-block bootstrap with $B = 199$ valid replications targeted. "Valid/attempted" reports the number of valid bootstrap draws relative to the number of simulation attempts required to obtain them. Values of $J_T(1)$ smaller than 10^{-18} are reported as zero.

All three structural profiles converge. No numerical failure, boundary flag, or weak-identification flag is triggered.

Conditional on the estimated structural parameters, this gives a linear-Gaussian state-space system for filtering and forecasting. For the general state-space methodology, see Kalman (1960) and Durbin and Koopman (2012, Ch. 4–5). For stochastic volatility filtering based on Gaussian approximations and mixtures, see Harvey et al. (1994) and Kim et al. (1998).

Let $p_w = \max\{p, 1\}$ and define the augmented state vector

$$\alpha_s := \begin{bmatrix} w_s & w_{s-1} & \cdots & w_{s-p_w+1} & v_s & v_{s-1} & \cdots & v_{s-q+1} \end{bmatrix}'.$$

The v-block is omitted when $q = 0$. The state and measurement equations are

$$\alpha_s = T_{pq}\alpha_{s-1} + R_{pq}u_s, \qquad u_s \sim N(0, \sigma_v^2), \tag{G.1}$$

$$y_s^* = Z\alpha_s + \epsilon_s^G, \qquad Z = e_1', \tag{G.2}$$

where the first row of T_{pq} encodes

$$w_s = \sum_{j=1}^{p} \phi_j w_{s-j} - \sum_{j=1}^{q} \theta_j v_{s-j} + u_s.$$

The remaining rows shift the lagged w- and v-components. The innovation-loading vector is

$$R_{pq} = \begin{cases} e_1, & q = 0, \\ e_1 + e_{p_w+1}, & q > 0, \end{cases}$$

so that $u_s = v_s$ enters w_s and, when $q > 0$, also becomes the contemporaneous v_s state. The state innovation covariance is therefore $Q_\alpha = \sigma_v^2 R_{pq} R_{pq}'$.

For a fixed training sample, write

$$a_{s|j} = \mathbb{E}[\alpha_s \mid \mathcal{F}_j^*], \qquad P_{s|j} = \mathrm{Var}(\alpha_s \mid \mathcal{F}_j^*), \qquad \mathcal{F}_j^* = \sigma(y_1^*, \ldots, y_j^*).$$

The implementation initializes the state mean at zero. When the estimated transition matrix is stable, the initial covariance is the stationary solution of

$$P_0 = T_{pq} P_0 T_{pq}' + Q_\alpha;$$

a diffuse covariance is used as a numerical fallback if this solution cannot be computed. The filtering recursion is

$$a_{s|s-1} = T_{pq} a_{s-1|s-1}, \qquad P_{s|s-1} = T_{pq} P_{s-1|s-1} T_{pq}' + Q_\alpha, \tag{G.3}$$

$$\nu_s = y_s^* - Z a_{s|s-1}, \qquad F_s = Z P_{s|s-1} Z' + H, \tag{G.4}$$

$$K_s = P_{s|s-1} Z' F_s^{-1}, \qquad a_{s|s} = a_{s|s-1} + K_s \nu_s, \tag{G.5}$$

$$P_{s|s} = P_{s|s-1} - K_s F_s K_s'. \tag{G.6}$$

The first element of $a_{s|s}$ is the filtered estimate of latent log variance.

Let t denote a forecast origin. Starting from $(a_{t|t}, P_{t|t})$, the code iterates the prediction equations without measurement updates:

$$a_{t+h|t} = T_{pq}^h a_{t|t}, \tag{G.7}$$

$$P_{t+h|t} = T_{pq}^h P_{t|t} (T_{pq}^h)' + \sum_{j=0}^{h-1} T_{pq}^j Q_\alpha (T_{pq}^j)'. \tag{G.8}$$

Define

$$m_{w,t+h|t} = e_1' a_{t+h|t}, \qquad s_{w,t+h|t}^2 = e_1' P_{t+h|t} e_1.$$

Under the Gaussian filtering approximation,

$$w_{t+h} \mid \mathcal{F}_t^* \approx N\Big(m_{w,t+h|t}, s_{w,t+h|t}^2\Big).$$

Since

$$\mathrm{Var}(y_{t+h} \mid w_{t+h}) = \sigma_y^2 e^{w_{t+h}},$$

the return-variance forecast is

$$\widehat{h}_{t+h|t} = \widehat{\sigma}_{y,t}^2 \exp\Big(m_{w,t+h|t} + \frac{1}{2} s_{w,t+h|t}^2\Big). \tag{G.9}$$

The scale parameter is re-estimated at each forecast origin from the same training sample:

$$\widehat{\sigma}^2_{y,t} = \exp(\widehat{\mu}_t - c_z), \qquad \widehat{\mu}_t = \frac{1}{t}\sum_{s=1}^{t} \log\Big([y_s^{(t)}]^2\Big), \qquad c_z = \mathbb{E}[\log(z_s^2)].$$

Equations (G.3)–(G.9) are the filtering and forecasting steps implemented for each SV specification at every expanding-window origin.

References

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

Bradley, R. C. (2005): "Basic Properties of Strong Mixing Conditions: A Survey and Some Open Questions," *Probability Surveys*, 2, 107–144.

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.

Dufour, J.-M. and P. Valéry (2006): "On a Simple Two-Stage Closed-Form Estimator for a Stochastic Volatility in a General Linear Regression," in *Econometric Analysis of Financial and Economic Time Series*, ed. by T. B. Fomby and D. Terrell, Bingley: Emerald Group Publishing, vol. 20A of *Advances in Econometrics*, 259–288.

——— (2009): "Exact and Asymptotic Tests for Possibly Non-Regular Hypotheses on Stochastic Volatility Models," *Journal of Econometrics*, 150, 193–206.

Durbin, J. and S. J. Koopman (2012): *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press, 2nd ed.

Granger, C. W. and M. J. Morris (1976): "Time Series Modelling and Interpretation," *Journal of the Royal Statistical Society: Series A (General)*, 139, 246–257.

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

Kalman, R. E. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 35–45.

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.

Mokkadem, A. (1988): "Mixing Properties of ARMA Processes," *Stochastic Processes and Their Applications*, 29, 309–315.