Supplemental Appendices for

New testing approaches for mean-variance predictability

Gabriele Fiorentini

Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy

<gabriele.fiorentini@unifi.it>

Enrique Sentana

CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain

<sentana@cemfi.es>

Revised: February 2020

A Proofs

Proposition 1

Given the discussion in Supplemental Appendix E, to find the score function, the expected value of the Hessian and the variance of the score of the pseudo log-likelihood function, all we need is the matrix $\mathbf{Z}_{dt}(\boldsymbol{\theta}_s)$, which in turn requires the Jacobian of the conditional mean and covariance functions. In view of (1), we will have that

$$\partial\mu_t(\pi,0,\omega)/\partial\boldsymbol{\theta}' = \begin{pmatrix} 1 & y_{t-1}-\pi & 0 \end{pmatrix}$$

and

$$\partial\sigma_t^2(\pi,0,\omega)/\partial\boldsymbol{\theta}' = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix},$$

whence

$$\mathbf{Z}_{dt}(\pi,0,\omega) = \begin{bmatrix} \omega^{-1/2} & 0 \\ \epsilon_{t-1}(\boldsymbol{\theta}_s,0) & 0 \\ 0 & \frac{1}{2}\omega^{-1} \end{bmatrix}, \tag{A1}$$

so that

$$\mathbf{Z}_{d}(\pi_0,0,\omega_0,\boldsymbol{\eta}_0) = \begin{bmatrix} \omega_0^{-1/2} & 0 \\ 0 & 0 \\ 0 & \frac{1}{2}\omega_0^{-1} \end{bmatrix}. \tag{A2}$$

As a result, the score under the null will be

$$\begin{bmatrix} s_{\boldsymbol{\pi}t}(\pi,0,\omega,\boldsymbol{\eta}) \\ s_{\rho t}(\pi,0,\omega,\boldsymbol{\eta}) \\ s_{\omega t}(\pi,0,\omega,\boldsymbol{\eta}) \end{bmatrix} = \begin{bmatrix} -\omega^{-1/2}\partial f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\eta\right]/\partial\varepsilon^* \\ -\partial f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\eta\right]/\partial\varepsilon^*\cdot\epsilon_{t-1}(\boldsymbol{\theta}_s,0) \\ -\frac{1}{2}\omega^{-1}[\partial f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\eta\right]/\partial\varepsilon^*\cdot\epsilon_t(\boldsymbol{\theta}_s,0)+1] \end{bmatrix}.$$

Given Assumptions 1-3, we can then use standard arguments (see e.g. Newey and McFadden (1994)) to show that

$$\begin{aligned}\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} s_{\rho t}(\hat{\boldsymbol{\phi}}_s,0) &= \frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} s_{\rho t}(\boldsymbol{\phi}_{s\infty},0) + \frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbf{h}_{\rho\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)\sqrt{T}(\hat{\boldsymbol{\phi}}_s-\boldsymbol{\phi}_{s\infty})+o_p(1) \\ &= \frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} s_{\rho t}(\boldsymbol{\phi}_{s\infty},0) - \frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbf{h}_{\rho\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)\left[\frac{1}{T}\sum\nolimits_{t=1}^{T}\mathbf{h}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)\right]^{-1} \\ &\qquad\times\frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T}\mathbf{s}_{\boldsymbol{\phi}_s t}(\boldsymbol{\phi}_{s\infty},0)+o_p(1),\end{aligned}$$

where $\boldsymbol{\phi}_s = (\boldsymbol{\theta}_s',\boldsymbol{\eta}')'$. Hence, the asymptotic variance of $\frac{\sqrt{T}}{T}\sum_{t=1}^{T} s_{\rho t}(\hat{\boldsymbol{\phi}}_s,0)$ will be given by $\mathcal{F}_{\rho\rho}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_\infty;\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0)$, where

$$\mathcal{F}_{\rho\rho} = \mathcal{B}_{\rho\rho} - 2\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}^{-1}\mathcal{B}_{\rho\boldsymbol{\phi}_s}' + \mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}^{-1}\mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}^{-1}\mathcal{A}_{\rho\boldsymbol{\phi}_s}',$$

and $\mathcal{B}_{\rho\rho}$, $\mathcal{A}_{\rho\boldsymbol{\phi}_s}$, etc. are the relevant elements of

$$\begin{aligned}\mathcal{B}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0) &= V[s_{\boldsymbol{\phi}t}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta})|\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0), \\ \mathcal{A}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0) &= -E[h_{\boldsymbol{\phi}\boldsymbol{\phi}t}(\boldsymbol{\theta}_s,0,\boldsymbol{\eta})|\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0).\end{aligned}$$

Tedious but straightforward algebra shows that at $\rho = 0$:

$$
\begin{aligned}
h_{\pi\pi t}(\boldsymbol{\phi}) &= \omega^{-1}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \\
h_{\pi\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-3/2}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s,0) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\} \\
\mathbf{h}_{\pi\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= -\omega^{-1/2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}' \\
h_{\omega\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-2}\{1+\tfrac{3}{2}\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s,0) \\
&\quad +\tfrac{1}{2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t^2(\boldsymbol{\theta}_s,0)\} \\
\mathbf{h}_{\omega\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= -\tfrac{1}{2}\omega^{-2}\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}' \cdot \epsilon_t(\boldsymbol{\theta}_s,0) \\
\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= \partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'
\end{aligned}
$$

Similarly, we can show that at $\rho = 0$

$$
\begin{aligned}
h_{\rho\pi t}(\boldsymbol{\phi}) &= \omega^{-1/2}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s,0) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\} \\
h_{\rho\rho t}(\boldsymbol{\phi}) &= \partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_{t-1}^2(\boldsymbol{\theta}_s,0) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^* \cdot \epsilon_{t-2}(\boldsymbol{\theta}_s,0) \\
h_{\rho\omega t}(\boldsymbol{\phi}) &= \tfrac{1}{2}\omega^{-1}\{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s,0) + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\} \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s,0) \\
\mathbf{h}_{\rho\boldsymbol{\eta} t}(\boldsymbol{\phi}) &= -\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\boldsymbol{\eta} \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s,0)
\end{aligned}
$$

Given that the pseudo-true values of π, ω and $\boldsymbol{\eta}$ are implicitly defined in such a way that

$$
\begin{aligned}
E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^* | \boldsymbol{\varphi}_0\} &= 0, \\
E\{1+\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s,0) | \boldsymbol{\varphi}_0\} &= 0, \\
E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\boldsymbol{\eta} | \boldsymbol{\varphi}_0\} &= \mathbf{0},
\end{aligned}
$$

the law of iterated expectations implies that

$$
\begin{aligned}
E[h_{\pi\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \omega_\infty^{-1}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\
E[h_{\pi\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-3/2}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\
E[\mathbf{h}_{\pi\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= -\omega_\infty^{-1/2}\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1] \\
E[\mathbf{h}_{\omega\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\
E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \mathcal{H}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)
\end{aligned}
$$

and

$$
\begin{aligned}
E[h_{\rho\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \omega_\infty^{-1/2}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s,0) \\
E[h_{\rho\rho t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}^2(\boldsymbol{\theta}_s,0) \\
E[h_{\rho\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s,0) \\
E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= -\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \cdot \epsilon_{t-1}(\boldsymbol{\theta}_s,0)
\end{aligned}
$$

where

$$\begin{aligned}
\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^*|I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\epsilon_t(\boldsymbol{\theta}_s,0)|I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}'|I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\varepsilon^*\cdot\epsilon_t^2(\boldsymbol{\theta}_s,0)|I_{t-1};\boldsymbol{\varphi}_0] \\
\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}_\infty\right]/\partial\varepsilon^*\partial\boldsymbol{\eta}'\cdot\epsilon_t(\boldsymbol{\theta}_s,0)|I_{t-1};\boldsymbol{\varphi}_0]
\end{aligned}$$

and $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}'_{s0}, 0, \boldsymbol{\varrho}'_0)'$.

Consequently,

$$\begin{aligned}
E[h_{\rho\pi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \omega_\infty^{-1/2}\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] \\
E[h_{\rho\rho t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] \\
E[h_{\rho\omega t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] \\
E[\mathbf{h}_{\rho\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0]
\end{aligned}$$

where

$$E[\epsilon_t(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] = E[\omega^{-1/2}(y_t-\pi)|\boldsymbol{\varphi}_0] = E[\omega^{-1/2}(\pi_0+\omega_0^{1/2}\varepsilon_t^*-\pi)|\boldsymbol{\varphi}_0] = \omega^{-1/2}(\pi_0-\pi)$$

and

$$E[\epsilon_t^2(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] = E[\omega^{-1}(y_t-\pi)^2|\boldsymbol{\varphi}_0] = E[\omega^{-1}(\pi_0+\omega_0^{1/2}\varepsilon_t^*-\pi)^2|\boldsymbol{\varphi}_0] = \omega^{-1}[(\pi_0-\pi)^2+\omega_0],$$

so that

$$V[\epsilon_t(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] = \omega^{-1}\omega_0. \tag{A3}$$

Given that $\mathcal{A}_{\rho\boldsymbol{\phi}_s}$ is proportional to the first column of $\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}$, we can immediately show that

$$\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}^{-1} = \begin{pmatrix} E[\epsilon_t(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0]\sqrt{\omega_\infty} & 0 & \mathbf{0}' \end{pmatrix} = E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1' \tag{A4}$$

if we evaluate these expressions at the pseudo true values, where $\mathbf{e}_1$ is the first element of the canonical basis. Therefore, the only elements of $\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$ that we need are the ones corresponding to π and ρ. But since

$$\begin{aligned}
\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) &= E[\mathcal{B}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)|\boldsymbol{\varphi}_\infty], \\
\mathcal{B}_t(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) &= V[\mathbf{s}_{\phi t}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_\infty)|\,I_{t-1};\boldsymbol{\varphi}_\infty] = \mathbf{Z}_t(\boldsymbol{\theta}_\infty)\mathcal{K}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathbf{Z}_t'(\boldsymbol{\theta}_\infty), \\
\mathcal{K}(\boldsymbol{\phi};\boldsymbol{\varphi}) &= V\left[\left.\begin{pmatrix} e_{lt}(\boldsymbol{\phi}) \\ e_{st}(\boldsymbol{\phi}) \\ \mathbf{e}_{rt}(\boldsymbol{\phi})\end{pmatrix}\right|\boldsymbol{\varphi}\right] = \begin{bmatrix} \mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{lr}'(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}'(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi})\end{bmatrix}
\end{aligned}$$

we will have that under the null of $H_0: \rho = 0$,

$$\begin{aligned}
&\begin{bmatrix} \mathcal{B}_{\pi\pi}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{B}_{\pi\rho}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) \\ \mathcal{B}_{\pi\rho}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0) & \mathcal{B}_{\rho\rho}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\end{bmatrix} \\
&= \mathcal{K}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\begin{bmatrix} \omega_\infty^{-1} & \omega_\infty^{-1/2}E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0] \\ \omega_\infty^{-1/2}E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0] & E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\end{bmatrix}.
\end{aligned}$$

Finally we obtain

$$\mathcal{F}_{\rho\rho}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_{\infty};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0)=\mathcal{K}_{ll}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)V[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0],$$

which is precisely the denominator of the R^2 in the regression of $\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty},0),\boldsymbol{\eta}\right]/\partial\varepsilon^*$ on a constant and $\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)$.

We can also use these expressions to derive the asymptotic variance of the pseudo ML estimator of ρ under the null. Specifically, straightforward algebra shows that the "$\rho\rho$" element of the matrix

$$\mathcal{C}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_{\infty})=\mathcal{A}^{-1}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_{\infty})\mathcal{B}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_{\infty})\mathcal{A}^{-1}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_{\infty})$$

will be given by

$$\frac{\mathcal{F}_{\rho\rho}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_{\infty};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0)}{\mathcal{G}^2_{\rho\rho}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_{\infty};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0)},$$

where

$$\mathcal{G}_{\rho\rho}=\mathcal{A}_{\rho\rho}-\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}'_{\rho\boldsymbol{\phi}_s}.$$

But (A4) immediate implies that

$$\begin{aligned}\mathcal{G}_{\rho\rho}(\boldsymbol{\theta}_{s\infty},0,\boldsymbol{\eta}_{\infty};\boldsymbol{\theta}_{s0},0,\boldsymbol{\varrho}_0) &= \mathcal{H}_{ll}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)\left\{E[\epsilon^2_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]-E^2[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\right\}\\ &= \mathcal{H}_{ll}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)V[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0],\end{aligned}$$

whence

$$\sqrt{T}\hat{\rho}_T\to N\left[0,\frac{\mathcal{K}_{ll}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)}{\mathcal{H}^2_{ll}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_0)}\frac{\omega_{\infty}}{\omega_0}\right]$$

in view of (A3). Not surprisingly, this expression nests both the usual Gaussian PML expression, as well as the true ML expression when the information matrix equality holds.

Let us now find the remaining elements of $\mathcal{C}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_{\infty})$. We need to find out an expression for $\mathcal{B}(\boldsymbol{\phi}_{\infty};\boldsymbol{\varphi}_{\infty})$, which is given by the unconditional expected value of

$$\begin{aligned}&\begin{bmatrix}\omega^{-1/2} & 0 & \mathbf{0}\\ \epsilon_{t-1}(\boldsymbol{\theta}_s,0) & 0 & \mathbf{0}\\ 0 & \frac{1}{2}\omega^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{bmatrix}\begin{bmatrix}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}'_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi})\end{bmatrix}\\ &\times\begin{bmatrix}\omega^{-1/2} & \epsilon_{t-1}(\boldsymbol{\theta}_s,0) & 0 & \mathbf{0}\\ 0 & 0 & \frac{1}{2}\omega^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{bmatrix}\\ =&\begin{bmatrix}\omega^{-1/2}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \frac{1}{2}\omega^{-1}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}'_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi})\\ \mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi})\end{bmatrix}\\ &\times\begin{bmatrix}\omega^{-1/2} & \epsilon_{t-1}(\boldsymbol{\theta}_s,0) & 0 & \mathbf{0}\\ 0 & 0 & \frac{1}{2}\omega^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_q\end{bmatrix}\end{aligned}$$

$$
= \left[\begin{array}{cc} \omega^{-1}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}^2(\boldsymbol{\theta}_s,0)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array} \right.
$$

$$
\left. \begin{array}{cc} \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \omega^{-1/2}\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}'_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{4}\omega^{-2}\mathcal{K}_{ss}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \frac{1}{2}\omega^{-1}\mathcal{K}'_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-1}\mathcal{K}_{sr}(\boldsymbol{\phi};\boldsymbol{\varphi}) & \mathcal{K}_{rr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{array} \right].
$$

As for $\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$, we can use the partitioned inverse formula to write

$$
\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \begin{pmatrix} \mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} + \mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}'_{\rho\boldsymbol{\phi}_s}\mathcal{G}^{-1}_{\rho\rho}\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & -\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}'_{\rho\boldsymbol{\phi}_s}\mathcal{G}^{-1}_{\rho\rho} \\ -\mathcal{G}^{-1}_{\rho\rho}\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & \mathcal{G}^{-1}_{\rho\rho} \end{pmatrix}.
$$

But if we use the expression for $\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}$, we will get

$$
\begin{aligned}
\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) &= \begin{pmatrix} \mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & 0 \\ 0 & 0 \end{pmatrix} \\
&+\mathcal{G}^{-1}_{\rho\rho}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1 \\ 1 \end{pmatrix}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}'_1 & 1 \end{pmatrix}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty) = \begin{pmatrix} \mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & 0 \\ 0 & 0 \end{pmatrix} \\
&+\mathcal{G}^{-1}_{\rho\rho}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1 \\ 1 \end{pmatrix}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}'_1 & 1 \end{pmatrix}\begin{pmatrix} \mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & 0 \\ \mathcal{B}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & 0 \end{pmatrix} \\
&+\mathcal{G}^{-1}_{\rho\rho}\begin{pmatrix} \mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & \mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{B}'_{\rho\boldsymbol{\phi}_s} \\ 0 & 0 \end{pmatrix} \\
&\times\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1 \\ 1 \end{pmatrix}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}'_1 & 1 \end{pmatrix} \\
&+\mathcal{F}_{\rho\rho}\mathcal{G}^{-2}_{\rho\rho}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1 \\ 1 \end{pmatrix}\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}'_1 & 1 \end{pmatrix}.
\end{aligned}
$$

But

$$
\begin{pmatrix} -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}'_1 & 1 \end{pmatrix}\begin{pmatrix} \mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & 0 \\ \mathcal{B}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & 0 \end{pmatrix} = \mathbf{0}
$$

because

$$
\mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}'_{\rho\boldsymbol{\phi}_s} = \begin{bmatrix} \omega^{-1}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{bmatrix} E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}
$$

and

$$
\begin{aligned}
&\mathcal{B}'_{\rho\boldsymbol{\phi}_s} - \mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}'_{\rho\boldsymbol{\phi}_s} \\
&= \begin{matrix} \omega^{-1/2}\epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-1}\epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \epsilon_{t-1}(\boldsymbol{\theta}_s,0)\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{matrix} - \begin{bmatrix} \omega^{-1}\mathcal{K}_{ll}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \frac{1}{2}\omega^{-3/2}\mathcal{K}_{ls}(\boldsymbol{\phi};\boldsymbol{\varphi}) \\ \omega^{-1/2}\mathcal{K}_{lr}(\boldsymbol{\phi};\boldsymbol{\varphi}) \end{bmatrix} E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2} = \mathbf{0}.
\end{aligned}
$$

As a result,

$$\mathcal{C}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)=\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{B}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)\mathcal{A}^{-1}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_\infty)$$
$$=\begin{pmatrix}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{B}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} & \mathbf{0}\\ \mathbf{0} & 0\end{pmatrix}+\mathcal{F}_{\rho\rho}\mathcal{G}^{-2}_{\rho\rho}\begin{pmatrix}E^2[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty\mathbf{e}_1\mathbf{e}_1' & -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1\\ -E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1' & 1\end{pmatrix},$$

which means that the PML estimator of ρ will be asymptotically orthogonal to the PML estimators of ω and $\boldsymbol{\eta}$, but not to the PML estimator of π.

To prove the third part of the proposition, we need to find out what would happen for a restricted pseudo ML estimator that fixes the shape parameters to some arbitrary value $\bar{\boldsymbol{\eta}}$. Fortunately, all the previous expressions remain valid after eliminating the rows and columns corresponding to $\boldsymbol{\eta}$, and replacing $\boldsymbol{\theta}_\infty$ by $\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}})=[\pi_\infty(\bar{\boldsymbol{\eta}}),\omega_\infty(\bar{\boldsymbol{\eta}})]$, which are the values that solve the system of equations

$$\begin{aligned}E[\partial\ln f\{\epsilon_t[\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}}),0],\bar{\boldsymbol{\eta}}\}/\partial\varepsilon^*|\boldsymbol{\varphi}_0] &= 0,\\ E[1+\partial\ln f\{\epsilon_t[\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}}),0],\bar{\boldsymbol{\eta}}\}/\partial\varepsilon^*\cdot\epsilon_t[\boldsymbol{\theta}_\infty(\bar{\boldsymbol{\eta}}),0]|\boldsymbol{\varphi}_0] &= 0.\end{aligned}$$

In fact, we would obtain exactly the same expressions even if fixed both ω and $\boldsymbol{\eta}$ to some arbitrary values $\bar{\omega}$ and $\bar{\boldsymbol{\eta}}$, as long as we replaced π_∞ by $\pi_\infty(\bar{\omega},\bar{\boldsymbol{\eta}})$, which would be the value that solves

$$E[\partial\ln f\{\bar{\omega}^{-1/2}[y_t-\pi_\infty(\bar{\omega},\bar{\boldsymbol{\eta}})],\bar{\boldsymbol{\eta}}\}/\partial\varepsilon^*|\boldsymbol{\varphi}_0]=0.$$

□

Proposition 2

Let us first proof that the moment condition (2) continues to hold when the true DGP is (4), in which case

$$\epsilon_t(\pi_\infty,\omega_\infty,0)=\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*].$$

On this basis, we can write the moment condition underlying our proposed test as

$$\omega_\infty^{-1/2}(\mu_0-\pi_\infty)E\left[\left.\frac{\partial\ln f\{\epsilon_t(\boldsymbol{\theta}_{s\infty},0),\boldsymbol{\eta}\}}{\partial\varepsilon^*}\right|\boldsymbol{\varphi}_0\right]+\omega_\infty^{-1/2}E\left[\left.\frac{\partial\ln f\{\epsilon_t(\boldsymbol{\theta}_{s\infty},0),\boldsymbol{\eta}\}}{\partial\varepsilon^*}\sigma_{t-1}\varepsilon_{t-1}^*\right|\boldsymbol{\varphi}_0\right].$$

The first summand is 0 because it is proportional to the moment condition that defines π_∞. In turn, the second summand is also 0 thanks to the zero mean *i.i.d.* assumption on ε_t^*.

Next, we need to find the expected values of the ten different elements of the Hessian evaluated under the null that appear in the proof of Proposition 1. Unfortunately, we cannot directly rely on the law of iterated expectation conditional on the past. Nevertheless, we can still prove that

$$
\begin{aligned}
E[h_{\rho\pi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \omega_\infty^{-1/2}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0] \\
E[h_{\rho\rho t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0] \\
E[h_{\rho\omega t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0] \\
E[\mathbf{h}_{\rho\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]
\end{aligned}
$$

where

$$
E[\epsilon_t(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] = E[\omega^{-1/2}(y_t-\pi)|\boldsymbol{\varphi}_0] = E[\omega^{-1/2}(\mu_0+\sigma_t\varepsilon_t^*-\pi)|\boldsymbol{\varphi}_0] = \omega^{-1/2}(\mu_0-\pi)
$$

and

$$
E[\epsilon_t^2(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] = E[\omega^{-1}(y_t-\pi)^2|\boldsymbol{\varphi}_0] = E[\omega^{-1}(\mu_0+\sigma_t\varepsilon_t^*-\pi)^2|\boldsymbol{\varphi}_0] = \omega^{-1}[(\mu_0-\pi)^2+\sigma^2],
$$

with $\sigma^2 = E(\sigma_t^2)$, so that $V[\epsilon_t(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0] = \omega^{-1}\sigma^2$, as in (A3).

Let us start with the expression for $h_{\rho\pi t}(\boldsymbol{\phi})$:

$$
\begin{aligned}
&E\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_{t-1}\varepsilon_{t-1}^*]\right|\boldsymbol{\varphi}_0\right] \\
&= \omega_\infty^{-1/2}(\mu_0-\pi_\infty)E\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\right|\boldsymbol{\varphi}_0\right] \\
&\quad +\omega_\infty^{-1}E\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\sigma_{t-1}\varepsilon_{t-1}^*\right|\boldsymbol{\varphi}_0\right].
\end{aligned}
$$

The second summand is 0 because ε_{t-1}^* has 0 mean. In contrast, the first summand is clearly seen to be $E[\epsilon_{t-1}(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0]$ times $\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, as required.

Let us now move on to $h_{\rho\rho t}(\boldsymbol{\phi})$. The expectation of its first component is given by

$$
\begin{aligned}
&E\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_{t-1}\varepsilon_{t-1}^*]^2\right|\boldsymbol{\varphi}_0\right] \\
&= \omega_\infty^{-1}(\mu_0-\pi_\infty)^2E\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\right|\boldsymbol{\varphi}_0\right] \\
&\quad +\omega_\infty^{-1}E\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\sigma_{t-1}^2\varepsilon_{t-1}^{*2}\right|\boldsymbol{\varphi}_0\right] \\
&\quad +2\omega_\infty^{-1}(\mu_0-\pi_\infty)\left[\left.\frac{\partial^2\ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\varepsilon^*}\sigma_{t-1}\varepsilon_{t-1}^*\right|\boldsymbol{\varphi}_0\right].
\end{aligned}
$$

We have already seen that the third summand is 0, while the first summand will be

$$
\omega_\infty^{-1}(\mu_0-\pi_\infty)^2\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0).
$$

As for the second one, the zero mean, unit variance *i.i.d.* assumption on ε_t^* together with the definition of σ^2 yields $\omega_\infty^{-1}\sigma^2\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, so that the required expectation becomes $E[\epsilon_{t-1}^2(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0]$ times $\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$.

Nevertheless, we still need to worry about the expected value of the second component of $h_{\rho\rho t}(\boldsymbol{\phi})$, which is given by

$$
\begin{aligned}
& E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_{t-2}\varepsilon_{t-2}^*]\right|\boldsymbol{\varphi}_0\right] \\
= \; & \omega_\infty^{-1/2}(\mu_0-\pi_\infty)E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\right|\boldsymbol{\varphi}_0\right] \\
& +\omega_\infty^{-1}\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\sigma_{t-2}\varepsilon_{t-2}^*\right|\boldsymbol{\varphi}_0\right].
\end{aligned}
$$

But the arguments above immediately imply that both these terms will be 0, as required.

Let us now consider $h_{\rho\omega t}(\boldsymbol{\phi})$. Given that the second term will have 0 mean, we can focus on the following expectations

$$
\begin{aligned}
& \omega_\infty^{-1}E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*][(\mu_0-\pi_\infty)+\sigma_{t-1}\varepsilon_{t-1}^*]\right|\boldsymbol{\varphi}_0\right] \\
= \; & \omega_\infty^{-1}(\mu_0-\pi_\infty)^2E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\right|\boldsymbol{\varphi}_0\right] \\
& +\omega_\infty^{-1}(\mu_0-\pi_\infty)E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\sigma_{t-1}\varepsilon_{t-1}^*\right|\boldsymbol{\varphi}_0\right] \\
& +\omega_\infty^{-1}(\mu_0-\pi_\infty)E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\sigma_t\varepsilon_t^*\right|\boldsymbol{\varphi}_0\right] \\
& +E\left[\left.\frac{\partial \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*}\sigma_t\varepsilon_t^*\sigma_{t-1}\varepsilon_{t-1}^*\right|\boldsymbol{\varphi}_0\right]
\end{aligned}
$$

We have already seen that the first two summands will be 0. For analogous reasons, the fourth one will also be 0. In contrast, the third one will be given by $E[\epsilon_{t-1}(\boldsymbol{\theta}_s,0)|\boldsymbol{\varphi}_0]$ times $\mathcal{H}_{l\omega}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, as required.

Finally, we need to study $\mathbf{h}_{\rho\boldsymbol{\eta}t}(\boldsymbol{\phi})$. But its expected value will be given by

$$
\begin{aligned}
& E\left[\left.\frac{\partial^2 \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\boldsymbol{\eta}}\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_{t-1}\varepsilon_{t-1}^*]\right|\boldsymbol{\varphi}_0\right] \\
= \; & \omega_\infty^{-1/2}(\mu_0-\pi_\infty)E\left[\left.\frac{\partial^2 \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\boldsymbol{\eta}}\right|\boldsymbol{\varphi}_0\right] \\
& \omega_\infty^{-1}E\left[\left.\frac{\partial^2 \ln f\{\omega_\infty^{-1/2}[(\mu_0-\pi_\infty)+\sigma_t\varepsilon_t^*];\boldsymbol{\eta}\}}{\partial\varepsilon^*\partial\boldsymbol{\eta}}\sigma_{t-1}\varepsilon_{t-1}^*\right|\boldsymbol{\varphi}_0\right].
\end{aligned}
$$

Once again, the second summand is clearly equal to 0, while the first one will be given by $-\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\cdot E[\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]$, as desired.

If we combine these expressions with the fact that

$$\begin{aligned}
E[h_{\pi\pi t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \omega_\infty^{-1}\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[h_{\pi\omega t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-3/2}\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[\mathbf{h}_{\pi\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\omega_\infty^{-1/2}\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)-1]\\
E[\mathbf{h}_{\omega\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\\
E[\mathbf{h}_{\boldsymbol{\eta}\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \mathcal{H}_{rr}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)
\end{aligned}$$

it is easy to see that $\mathcal{A}_{\rho\boldsymbol{\phi}_s}$ will be proportional to the first column of $\mathcal{A}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s}$, so that we can immediately show that

$$\mathcal{A}_{\rho\boldsymbol{\phi}_s}\mathcal{A}^{-1}_{\boldsymbol{\phi}_s\boldsymbol{\phi}_s} = \begin{pmatrix} E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\sqrt{\omega_\infty} & 0 & \mathbf{0}' \end{pmatrix} = E[\epsilon_t(\boldsymbol{\theta}_{s\infty},0)|\boldsymbol{\varphi}_0]\omega_\infty^{1/2}\mathbf{e}_1'$$

if we evaluate these expressions at the pseudo true values. The rest of the proof follows the same steps as the proof of Proposition 1, but with $\mathcal{B}_{\boldsymbol{\phi}\boldsymbol{\phi}}$ representing the long run variance of the average scores. □

Proposition 3

The first thing we can show in this context is that the pseudo true value of the mean parameter coincides with the true mean. To understand why, we can use the law of iterated expectations to express the moment condition defining π_∞ evaluated at $\pi_\infty = \mu_0$ as

$$E\left\{\left.\frac{\partial \ln f[\epsilon_t(\mu_0,\omega_\infty,0);\boldsymbol{\eta}]}{\partial\varepsilon^*}\right|\boldsymbol{\varphi}_0\right\} = E\left[E\left\{\left.\frac{\partial \ln f[\epsilon_t(\mu_0,\omega_\infty,0);\boldsymbol{\eta}]}{\partial\varepsilon^*}\right|I_{t-1};\boldsymbol{\varphi}_0\right\}\right].$$

But since

$$\epsilon_t(\mu_0,\omega_\infty,0) = \omega_\infty^{-1/2}\sigma_t\varepsilon_t^*,$$

we can write the conditional expectation as

$$\int_{-\infty}^{\infty}\frac{\partial \ln f[\omega_\infty^{-1/2}\sigma_t\varepsilon_t^*;\boldsymbol{\eta}]}{\partial\varepsilon^*}h(\varepsilon_t^*)d\varepsilon_t^*.$$

The symmetry of the assumed conditional density implies that its derivative with respect to its argument is an odd function, which makes the integral above 0 in view of the symmetry of the true conditional distribution of ε_t^*.

Let us know turn to the moment condition (2) evaluated at $\pi_\infty = \mu_0$. If we use again the law of iterated expectations, we can re-write it as

$$\begin{aligned}
&E\left\{\left.\frac{\partial \ln f[\epsilon_t(\mu_0,\omega_\infty,0);\eta]}{\partial\varepsilon^*}\epsilon_{t-1}(\mu_0,\omega_\infty,0)\right|\boldsymbol{\varphi}_0\right\}\\
&= E\left[\left.\epsilon_{t-1}(\mu_0,\omega_\infty,0)E\left\{\left.\frac{\partial \ln f[\epsilon_t(\mu_0,\omega_\infty,0);\boldsymbol{\eta}]}{\partial\varepsilon^*}\right|I_{t-1};\boldsymbol{\varphi}_0\right\}\right|\boldsymbol{\varphi}_0\right],
\end{aligned}$$

which is 0 for exactly the same reason.

These two results also imply that $s_{\pi t}(\mu_0,\omega_\infty;\boldsymbol{\eta})$ and $s_{\rho t}(\mu_0,\omega_\infty;\boldsymbol{\eta})$ are martingale differences despite the misspecification of the distribution and the disregard for the time-variation σ_t^2, so that the asymptotic variance of their averages will coincide with their unconditional variance.

Once again, we need to look at the Hessian matrix in this case. But the only difference with respect to the conditionally homoskedastic case in Proposition 1 is that the exact expressions for $\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, $\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$, etc. will be a function of σ_t^2. Nevertheless, the symmetry of the true conditional distribution implies that both $\mathcal{H}_{lst}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ and $\mathcal{H}_{srt}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)$ will be zero. But given that

$$\begin{aligned}
E[h_{\rho\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \omega_\infty^{-1/2}\mathcal{H}_{llt}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)\\
E[h_{\rho\rho t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \mathcal{H}_{llt}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty},0)\\
E[h_{\rho\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1}\mathcal{H}_{lst}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)=\mathbf{0}\\
E[\mathbf{h}_{\rho\boldsymbol{\eta} t}(\boldsymbol{\phi}_\infty)|I_{t-1};\boldsymbol{\varphi}_0] &= -\mathcal{H}_{lrt}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)=\mathbf{0}
\end{aligned}$$

we do not need to worry about the sampling uncertainty in estimating ω_∞ and $\boldsymbol{\eta}_\infty$. In general, $E[\mathcal{H}_{llt}(\boldsymbol{\phi}_\infty;\boldsymbol{\varphi}_0)\epsilon_{t-1}(\boldsymbol{\theta}_{s\infty},0)]$ will be different from 0, but if the conditional variance is a symmetric function of ϵ_{t-1}, then this moment will be 0 too. In any case, we can still conduct the usual test by regressing $\partial\ln f[\epsilon_t(\mu_0,\omega_\infty,0);\boldsymbol{\eta}]/\partial\varepsilon$ on a constant and $\epsilon_{t-1}(\mu_0,\omega_\infty,0)$. □

Proposition 4

Consider the following model:

$$\left.\begin{array}{c}
y_t=\pi_0+\sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*,\\
\sigma_t^2(\boldsymbol{\theta})=\omega[1+\gamma(y_{t-1}-\pi)^2],\\
\varepsilon_t^*|I_{t-1};\pi,\omega,\gamma,\boldsymbol{\eta}\sim i.i.d.\ D(0,1,\boldsymbol{\eta}),\\
\text{with density function } f(.,\boldsymbol{\eta})
\end{array}\right\},$$

where the parameters of interest are $\boldsymbol{\phi}=(\boldsymbol{\theta}',\boldsymbol{\eta}')'$, $\boldsymbol{\theta}'=(\boldsymbol{\theta}_s',\gamma)'$ and $\boldsymbol{\theta}_s=(\pi,\omega)'$. In this context, the null hypothesis is $H_0:\gamma=0$.

It is then easy to see that

$$\frac{\partial\mu_t}{\partial\theta'}=\begin{pmatrix}1 & 0 & 0\end{pmatrix}$$

while

$$\frac{\partial\sigma_t^2}{\partial\theta'}=\begin{pmatrix}-2\omega\gamma(x_{t-1}-\pi) & 1+\gamma(x_{t-1}-\pi)^2 & \omega(x_{t-1}-\pi)^2\end{pmatrix}.$$

As a result, the score vector will be

$$\begin{aligned}
s_{\pi t} &= -\frac{1}{\{\omega[1+\gamma(x_{t-1}-\pi)^2]\}^{1/2}}\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\\
&\quad+\frac{\gamma(x_{t-1}-\pi)}{[1+\gamma(x_{t-1}-\pi)^2]}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\},
\end{aligned}$$

$$
\begin{aligned}
s_{\omega t} &= -\frac{1}{2\omega}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\},\\
s_{\gamma t} &= -\frac{(x_{t-1}-\pi)^2}{2[1+\gamma(x_{t-1}-\pi)^2]}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\},\\
s_{\boldsymbol{\eta}t} &= \frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\boldsymbol{\eta}}
\end{aligned}
$$

which under the null of $\gamma=0$ reduces to

$$
\begin{aligned}
s_{\pi t} &= -\frac{1}{\omega^{1/2}}\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]}{\partial\varepsilon^*},\\
s_{\omega t} &= -\frac{1}{2\omega}\left\{1+\epsilon_t(\boldsymbol{\theta}_s,0)\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\},\\
s_{\gamma t} &= -\frac{\omega}{2}\epsilon_{t-1}^2(\boldsymbol{\theta}_s,0)\left\{1+\epsilon_t(\boldsymbol{\theta}_s,0)\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\},\\
s_{\boldsymbol{\eta}t} &= \frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]}{\partial\boldsymbol{\eta}}.
\end{aligned}
$$

Note that we could have obtained the same expressions by using the chain rule for first derivatives since

$$
\begin{aligned}
s_{\omega t} &= -\frac{1-\gamma\omega}{2\omega[1+\gamma(x_{t-1}-\pi)^2]}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\}\\
s_{\alpha t} &= -\frac{(x_{t-1}-\pi)^2-\frac{\omega}{1-\gamma\omega}}{2\omega[1+\gamma(x_{t-1}-\pi)^2]}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\}
\end{aligned}
$$

and

$$
\frac{\partial\begin{pmatrix}\omega\\ \alpha\end{pmatrix}}{\partial\begin{pmatrix}\omega & \gamma\end{pmatrix}}=\begin{pmatrix}(1-\gamma\omega)^{-2} & \omega^2(1-\gamma\omega)^{-2}\\ \gamma & \omega\end{pmatrix}.
$$

Given Assumptions 1-3, we can then use standard arguments (see e.g. Newey and McFadden (1994)) to expand the average score and obtain the asymptotic distribution of the sample analogue to the moment condition (5) evaluated at the pseudo maximum likelihood estimators of the parameters under the null, as in the proof of Proposition 1.

Similarly,

$$
\begin{aligned}
h_{\pi\pi t}(\boldsymbol{\phi}) = {} & \frac{1}{\sigma_t^2}\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}-\frac{\omega\gamma\,(x_{t-1}-\pi)}{\sigma_t^3}\left(\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right)\\
&+\gamma\frac{-1+\gamma(x_{t-1}-\pi)^2}{[1+\gamma(x_{t-1}-\pi)^2]^2}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\}\\
&+\frac{\gamma(x_{t-1}-\pi)}{[1+\gamma(x_{t-1}-\pi)^2]}\left(-\frac{1}{\sigma_t}+\frac{\omega\gamma(x_{t-1}-\pi)}{\sigma_t^2}\epsilon_t(\boldsymbol{\theta})\right)\\
&\times\left[\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right]
\end{aligned}
$$

$$
h_{\pi\omega t}(\boldsymbol{\phi})=-\frac{1}{2\omega}\left(-\frac{1}{\sigma_t}+\frac{\omega\gamma(x_{t-1}-\pi)}{\sigma_t^2}\epsilon_t(\boldsymbol{\theta})\right)\left[\frac{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right]
$$

$$
\begin{aligned}
h_{\pi\gamma t}(\boldsymbol{\phi}) &= \frac{(x_{t-1}-\pi)}{[1+\gamma(x_{t-1}-\pi)^2]^2}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\
&\quad -\frac{(x_{t-1}-\pi)^2}{2[1+\gamma(x_{t-1}-\pi)^2]}\left(-\frac{1}{\sigma_t}+\frac{\omega\gamma(x_{t-1}-\pi)}{\sigma_t^2}\epsilon_t(\boldsymbol{\theta})\right) \\
&\quad \times\left\{\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right\}
\end{aligned}
$$

$$
h_{\pi\boldsymbol{\eta}t}(\boldsymbol{\phi})=\left(-\frac{1}{\sigma_t}+\frac{\omega\gamma(x_{t-1}-\pi)}{\sigma_t^2}\epsilon_t(\boldsymbol{\theta})\right)\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\boldsymbol{\eta}'}
$$

$$
\begin{aligned}
h_{\omega\omega t}(\boldsymbol{\phi}) &= \frac{1}{2\omega^2}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\
&\quad +\frac{1}{4\omega^2}\left[\epsilon_t(\boldsymbol{\theta})\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\varepsilon_t^{*2}(\boldsymbol{\theta})\cdot\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right]
\end{aligned}
$$

$$
h_{\omega\gamma t}(\boldsymbol{\phi})=\frac{(x_{t-1}-\pi)^2}{4\omega[1+\gamma(x_{t-1}-\pi)^2]}\left[\epsilon_t(\boldsymbol{\theta})\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\varepsilon_t^{*2}(\boldsymbol{\theta})\cdot\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right]
$$

$$
h_{\omega\boldsymbol{\eta}t}(\boldsymbol{\phi})=-\frac{1}{2\omega}\epsilon_t(\boldsymbol{\theta})\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\boldsymbol{\eta}}
$$

$$
\begin{aligned}
h_{\gamma\gamma t}(\boldsymbol{\phi}) &= \frac{(x_{t-1}-\pi)^4}{2[1+\gamma(x_{t-1}-\pi)^2]^2}\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\
&\quad +\frac{(x_{t-1}-\pi)^4}{4[1+\gamma(x_{t-1}-\pi)^2]^2}\left[\epsilon_t(\boldsymbol{\theta})\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\varepsilon_t^{*2}(\boldsymbol{\theta})\cdot\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right]
\end{aligned}
$$

$$
h_{\gamma\boldsymbol{\eta}t}(\boldsymbol{\phi})=-\frac{(x_{t-1}-\pi)^2}{2[1+\gamma(x_{t-1}-\pi)^2]}\epsilon_t(\boldsymbol{\theta})\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon_t^*\partial\boldsymbol{\eta}'}
$$

and

$$
h_{\boldsymbol{\eta}\boldsymbol{\eta}t}(\boldsymbol{\phi})=\frac{\partial^2\ln f\left[\varepsilon_t^*(\boldsymbol{\theta}_s,0),\boldsymbol{\eta}\right]}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}.
$$

Under the null of $\gamma=0$ these expressions reduce to

$$
\begin{aligned}
h_{\pi\pi t}(\boldsymbol{\phi}) &= \frac{1}{\omega}\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*} \\
h_{\pi\omega t}(\boldsymbol{\phi}) &= \frac{1}{2\omega^{3/2}}\left[\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right] \\
h_{\pi\gamma t}(\boldsymbol{\phi}) &= \omega^{1/2}\epsilon_{t-1}(\boldsymbol{\theta})\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\
&\quad +\frac{\omega^{1/2}}{2}\epsilon_{t-1}^2(\boldsymbol{\theta})\left\{\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t(\boldsymbol{\theta})\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right\} \\
\mathbf{h}_{\pi\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= -\frac{1}{\omega^{1/2}}\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon_t^*\partial\boldsymbol{\eta}'}
\end{aligned}
$$

$$
\begin{aligned}
h_{\gamma\gamma t}(\boldsymbol{\phi}) &= \frac{1}{2}\omega^2\epsilon_{t-1}^4(\boldsymbol{\theta})\left\{1+\epsilon_t(\boldsymbol{\theta})\cdot\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}\right\} \\
&\quad +\frac{1}{4}\omega^2\epsilon_{t-1}^4(\boldsymbol{\theta})\left[\epsilon_t(\boldsymbol{\theta})\frac{\partial\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*}+\epsilon_t^2(\boldsymbol{\theta})\cdot\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon^*\partial\varepsilon^*}\right] \\
\mathbf{h}_{\gamma\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= -\frac{1}{2}\omega\epsilon_{t-1}^2(\boldsymbol{\theta})\cdot\epsilon_t(\boldsymbol{\theta})\frac{\partial^2\ln f\left[\epsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]}{\partial\varepsilon_t^*\partial\boldsymbol{\eta}'}
\end{aligned}
$$

and

$$h_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) = \frac{\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'}$$

Given that the pseudo-true values of π, ω and $\boldsymbol{\eta}$ are implicitly defined in such a way that

$$\begin{aligned}
E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* | \boldsymbol{\varphi}_0\} &= 0, \\
E\{1 + \partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_{s\infty}, 0) | \boldsymbol{\varphi}_0\} &= 0, \\
E\{\partial \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \boldsymbol{\eta} | \boldsymbol{\varphi}_0\} &= \mathbf{0},
\end{aligned}$$

the law of iterated expectations implies that

$$\begin{aligned}
E[h_{\pi\pi t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \omega_\infty^{-1} \mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[h_{\pi\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-3/2} \mathcal{H}_{ls}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[\mathbf{h}_{\pi\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= -\omega_\infty^{-1/2} \mathcal{H}_{lr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2} [\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) - 1] \\
E[\mathbf{h}_{\omega\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty^{-1} \mathcal{H}_{sr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \\
E[\mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \mathcal{H}_{rr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0)
\end{aligned}$$

and

$$\begin{aligned}
E[h_{\pi\gamma t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{2}\omega_\infty^{-1/2} \mathcal{H}_{ls}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty}, 0)|\boldsymbol{\varphi}_0] \\
E[h_{\omega\gamma t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{4}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) - 1] \cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_s, 0)|\boldsymbol{\varphi}_0] \\
E[h_{\gamma\gamma t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) - 1] \cdot E[\epsilon_{t-1}^4(\boldsymbol{\theta}_s, 0)|\boldsymbol{\varphi}_0] \\
E[\mathbf{h}_{\gamma\boldsymbol{\eta}t}(\boldsymbol{\phi}_\infty)|\boldsymbol{\varphi}_0] &= -\tfrac{1}{2}\omega_\infty \mathcal{H}_{sr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) \cdot E[\epsilon_{t-1}^2(\boldsymbol{\theta}_{s\infty}, 0)|\boldsymbol{\varphi}_0] \\
E[h_{\omega\omega t}(\boldsymbol{\phi}_\infty)|I_{t-1}; \boldsymbol{\varphi}_0] &= \tfrac{1}{4}\omega_\infty^{-2}[\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) - 1]
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{H}_{ll}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* \partial \varepsilon^* | I_{t-1}; \boldsymbol{\varphi}_0] \\
\mathcal{H}_{ls}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* \partial \varepsilon^* \cdot \epsilon_t(\boldsymbol{\theta}_s) | I_{t-1}; \boldsymbol{\varphi}_0] \\
\mathcal{H}_{lr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* \partial \boldsymbol{\eta}' | I_{t-1}; \boldsymbol{\varphi}_0] \\
\mathcal{H}_{ss}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* \partial \varepsilon^* \cdot \epsilon_t^2(\boldsymbol{\theta}_s, 0) | I_{t-1}; \boldsymbol{\varphi}_0] \\
\mathcal{H}_{sr}(\boldsymbol{\phi}_\infty; \boldsymbol{\varphi}_0) &= E[\partial^2 \ln f\left[\epsilon_t(\boldsymbol{\theta}_{s\infty}, 0), \boldsymbol{\eta}_\infty\right]/\partial \varepsilon^* \partial \boldsymbol{\eta}' \cdot \epsilon_t(\boldsymbol{\theta}_s, 0) | I_{t-1}; \boldsymbol{\varphi}_0]
\end{aligned}$$

and $\boldsymbol{\varphi}_0 = (\boldsymbol{\theta}_{s0}', 0, \boldsymbol{\varrho}_0')'$. Finally,

$$E[\epsilon_t^2(\boldsymbol{\theta}_s, 0)|\boldsymbol{\varphi}_0] = E[\omega^{-1}(y_t - \pi)^2|\boldsymbol{\varphi}_0] = E[\omega^{-1}(\pi_0 + \omega_0^{1/2}\varepsilon_t^* - \pi)^2|\boldsymbol{\varphi}_0] = \omega^{-1}[(\pi_0 - \pi)^2 + \omega_0]$$

and

$$\begin{aligned}
E\{[\epsilon_t^4(\boldsymbol{\theta}_s, 0)|\boldsymbol{\varphi}_0\} &= E\{\omega^{-2}[(y_t - \pi)^4|\boldsymbol{\varphi}_0\} = \omega^{-2} E\{[(\pi_0 - \pi) + \omega_0^{1/2}\varepsilon_t^*]^4|\boldsymbol{\varphi}_0\} \\
&= \omega^{-2}[(\pi_0 - \pi)^4 + 6(\pi_0 - \pi)^2\omega_0 + 4\omega_0^{3/2}(\pi_0 - \pi)\varphi(\boldsymbol{\varrho}_0) + \omega_0^2 \kappa(\boldsymbol{\varrho}_0)].
\end{aligned}$$

where $\varphi(\boldsymbol{\varrho}_0) = E(\varepsilon_t^{*3}|\boldsymbol{\varrho}_0)$ and $\kappa(\boldsymbol{\varrho}_0) = E(\varepsilon_t^{*4}|\boldsymbol{\varrho}_0)$ are the skewness and kurtosis coefficients of the true distribution of ε_t^*.

The rest of the proof is entirely analogous to the proof of Proposition 1. □

B Additional results

B.1 Joint tests for mean-variance predictability

In this appendix, we consider joint tests of AR and ARCH effects. Specifically, our alternative in the first-order case will be

$$\left.\begin{array}{c} y_t = \mu_t(\pi_0, \rho_0) + \sigma_t(\boldsymbol{\theta}_0)\varepsilon_t^*, \\ \mu_t(\pi, \rho) = \pi(1-\rho) + \rho y_{t-j}, \\ \sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha_j[y_{t-1} - \mu_{t-1}(\pi, \rho)]^2, \\ \varepsilon_t^*|I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\eta}_0 \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}_0) \end{array}\right\}, \tag{B5}$$

where the parameters of interest are $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta}')'$, with $\boldsymbol{\theta}' = (\boldsymbol{\theta}_s', \rho, \alpha)'$. When the conditional variance $\sigma_t^2(\boldsymbol{\theta})$ is constant ($\alpha = 0$), the above formulation reduces to (1). Similarly, when the levels of the observed variable are unpredictable ($\rho = 0$), the above model simplifies to (5). Finally, the joint null hypothesis of lack of predictability in mean and variance corresponds to $\rho = 0$ and $\alpha = 0$.

In this context, the double length artificial regression of Davidson and MacKinnon (1988) might seem natural. However, there are two potential problems. First, in general the mean and variance regressands, namely $\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s, 0), \boldsymbol{\eta}]/\partial\varepsilon^*$ and $1 + \epsilon_t(\boldsymbol{\theta}_s, 0)\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s, 0), \boldsymbol{\eta}]/\partial\varepsilon^*$, have different variances, which introduces heteroskedasticity. More seriously, those two regressands will be correlated unless the true distribution is symmetric. The solution is a system of seemingly unrelated regression equations (SURE) in which one simultaneously regresses each of those regressands on the corresponding regressors, $\epsilon_{t-1}(\boldsymbol{\theta}_s, 0)$ and $\epsilon_{t-1}^2(\boldsymbol{\theta}_s, 0)$, respectively, and jointly tests the significance of both slope coefficients. In effect, this is a joint moment test of (2) and (6). Under the null, the covariance matrix of those moment conditions is

$$V\left[\begin{array}{c} \epsilon_{t-1}(\boldsymbol{\theta}_{s0}, 0) \\ \frac{1}{2}\epsilon_{t-1}^2(\boldsymbol{\theta}_{s0}, 0) \end{array}\middle| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0\right] \odot V\left[\begin{array}{c} \partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}, 0), \boldsymbol{\eta}_\infty]/\partial\varepsilon^* \\ 1 + \epsilon_t(\boldsymbol{\theta}_{s0}, 0)\partial \ln f[\epsilon_t(\boldsymbol{\theta}_{s0}, 0), \boldsymbol{\eta}_\infty]/\partial\varepsilon^* \end{array}\middle| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\varrho}_0\right],$$

where $\odot$ denotes the Hadamard (or element-by-element) product of two matrices, which reduces to

$$\left[\begin{array}{cc} 1 & \frac{1}{2}\phi_0^2 \\ \frac{1}{2}\phi_0^2 & \frac{1}{4}(\kappa_0 - 1)^2 \end{array}\right]$$

when the assumed distribution is Gaussian but the true one has skewness and kurtosis coefficients ϕ_0 and κ_0, respectively.

Nevertheless, if the true distribution of ε_t^* is symmetric, then it turns out that the joint tests of AR(1)-ARCH(1) in Propositions 1 and 4 is simply the sum of the separate tests:

Proposition 5 *If ε_t^* is symmetrically distributed, then under the joint null hypothesis $H_0 : \rho = 0$ and $\alpha = 0$ the score test statistic*

$$LM_{AR(1)-ARCH(1)}(\boldsymbol{\eta}) = LM_{AR(1)}(\boldsymbol{\eta}) + LM_{ARCH(1)}(\boldsymbol{\eta}),$$

will be distributed as a χ^2 with 2 degrees of freedom as T goes to infinity. This asymptotic null distribution is unaffected if we replace $\boldsymbol{\theta}_s$ and $\boldsymbol{\eta}$ by their joint MLEs.

Proof. The proof is trivial if we combine several results that appear in the proofs of Propositions 1 and 4, respectively, with the fact that the corresponding efficiency bounds are block diagonal between $\boldsymbol{\theta}_s$, ρ and γ when the true distribution of ε_t^* is symmetric. □

Intuitively, the serial correlation orthogonality condition (2) is asymptotically orthogonal to the Arch orthogonality condition (6) because all odd order moments of symmetric distributions are 0, which means that the joint test is simply the sum of its two components.

B.2 Exploiting the persistence of expected returns

Let us now consider a situation in which

$$y_t = \pi(1 - \sum\nolimits_{l=1}^{h} \rho_l) + \sum\nolimits_{l=1}^{h} \rho_l y_{t-l} + \sqrt{\omega}\varepsilon_t^*,$$

with $h > 1$ but finite, so that the null hypothesis of lack of predictability becomes $H_0 : \rho_1 = \ldots = \rho_h = 0$. In view of our previous discussion, it is not difficult to see that under this maintained assumption the score test of $\rho_l = 0$ will be based on the orthogonality condition

$$E\left\{ \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s, 0), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \epsilon_{t-l}(\boldsymbol{\theta}_s, 0) \middle| \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0 \right\} = 0. \tag{B6}$$

In this context, it is straightforward to show that the test against Ar(h) dynamics will be given by the joint test of the moment conditions (B6) for $l = 1, \ldots, h$, whose asymptotic distribution would be a χ_h^2 under the null.

Such a test, though, does not impose any prior knowledge on the nature of the expected return process, other than its lag length is h. Nevertheless, there are theoretical and empirical reasons which suggest that time-varying expected returns should be smooth processes.

A rather interesting example of persistent expected returns is an autoregressive model in which $\rho_l = \rho$ for all l. In this case, we can use the results in Fiorentini and Sentana (1998) to show that the process for expected returns will be given by the following not strictly invertible Arma$(h, h-1)$ process:

$$\mu_{t+1} = \pi(1 - h\rho) + \sum\nolimits_{j=1}^{h} \rho\mu_{t+1-j} + \rho\left[\varepsilon_t + \sum\nolimits_{j=1}^{h-1} \varepsilon_{t-j}\right]. \tag{B7}$$

As long as the covariance stationarity condition $h\rho < 1$ is satisfied, the autocorrelations of the expected return process can be easily obtained from its autocovariance generating function

$$\psi_{\mu\mu}(z) = \frac{\left(1 + \sum_{j=1}^{h-1} z^j\right)\left(1 + \sum_{j=1}^{h-1} z^{-j}\right)}{\left(1 - \rho\sum_{j=1}^{h} z^j\right)\left(1 - \rho\sum_{j=1}^{h} z^{-j}\right)}, \tag{B8}$$

which contrasts with the autocovariance generating function of the observed process

$$\psi_{yy}(z) = \frac{1}{\left(1 - \rho\sum_{j=1}^{h} z^j\right)\left(1 - \rho\sum_{j=1}^{h} z^{-j}\right)}.$$

In this context, we can easily find examples in which the autocorrelations of the observed return process are very small while the autocorrelations of the expected return process are much higher, and decline slowly towards 0. For example, Figure S1 presents the correlograms of y_t and μ_{t+1} on the same vertical scale for $h = 24$ and $\rho = .015$. Note that expression (B8) implies the correlograms of μ_{t+1} and an overlapping sum of h consecutive returns coincide.

This differential behaviour suggests that a test against first order correlation will have little power to detect such departures from white noise, the optimal test being one against an AR(h) process with common coefficients. We shall return to this issue below.

From the econometric point of view, the assumption that $\rho_l = \rho$ for all l does not pose any additional problems. Specifically, it is easy to prove that the relevant orthogonality condition will become

$$E\left\{ \frac{\partial \ln f[\epsilon_t(\boldsymbol{\theta}_s, 0), \boldsymbol{\eta}_0]}{\partial \varepsilon^*} \sum\nolimits_{l=1}^{h} \epsilon_{t-l}(\boldsymbol{\theta}_s, 0) \right| \left. \boldsymbol{\theta}_{s0}, 0, \boldsymbol{\eta}_0 \right\} = 0, \tag{B9}$$

with $h\mathcal{I}_{\rho\rho}(\boldsymbol{\theta}_s, 0, \boldsymbol{\eta})$ being the corresponding asymptotic variance under correct specification.

This moment condition is analogous to the one proposed by Jegadeesh (1989) to test for long run predictability of individual asset returns without introducing overlapping regressands. Cochrane (1991) and Hodrick (1992) discussed related suggestions. The intuition is that if returns contain a persistent but mean reverting predictable component, using a persistent right hand side variable such as an overlapping h-period return may help to pick it up. Not surprisingly, the asymptotic variance is analogous to the so-called Hodrick (1992) standard errors used in tests for long run predictability in univariate OLS regressions with overlapping regressands.

More recently, the Gaussian version of (B9) has also been tested by Moskowitz, Ooi and Pedersen (2012) in their empirical analysis of time series momentum. These authors provide both behavioural and rational justifications for the forecasting ability of lagged compound returns.

It is important to mention that the regressor $\sum_{l=1}^{h} \epsilon_{t-l}(\boldsymbol{\theta}_s, 0)$ will be quite persistent even if returns are serially uncorrelated because of the data overlap. Specifically, the first-order autocorrelation coefficient will be $1 - 1/h$ in the absence of return predictability. Nevertheless, since the correlation between the innovation to the regressor at time $t + 1$ and the innovations $\epsilon_t(\boldsymbol{\theta}_s, 0)$ is $1/\sqrt{h}$ under the null, the size problems that plague predictive regressions should not affect much our test (see Campbell and Yogo (2006)).

Let us now assess the power gains obtained by exploiting the persistence of expected returns. For simplicity we consider Gaussian tests only, and evaluate asymptotic power against *compatible* sequences of local alternatives of the form $\rho_{0T} = \bar{\rho}/\sqrt{T}$. As we show in Supplemental Appendix C, when the true model is (B7), the non-centrality parameter of the Gaussian score test for first order serial correlation is $\bar{\rho}^2$ regardless of h, while the non-centrality parameter of the test that exploits the persistence of the conditional mean will be $h\bar{\rho}^2$. Hence, Pitman's asymptotic relative efficiency of the two tests is precisely h. Figure S2A shows that those differences in non-centrality parameters result in substantive power gains. However, the asymptotic relative efficiency would be exactly reversed in the unlikely event that the true model were an AR(1) but we tested for it by using the moment condition (B9) (see Supplemental Appendix C). Not surprisingly, this would result in substantial power losses, which are illustrated in Figure S2A.

B.3 Construction of the quarterly portfolios

We follow exactly the same procedure as Ken French uses to create annual returns from monthly ones. The first thing we do is to add the monthly gross return on the 1-month Tbill rate to the excess returns of the 6 value-weighted portfolios formed on size and book-to-market, the 6 value-weighted portfolios formed on size and operating profitability, and the 6 value-weighted portfolios formed on size and investment to transform each of them into monthly gross returns. Then we compound the monthly gross returns into quarterly gross returns by multiplication, and subtract the quarterly gross return on the 3-month Tbill (from the FRED database) to obtain our quarterly excess returns. Form those, we create the five FF factors using the appropriate long or short weights.

More formally, let $X_{t,i}^{(K,J,D)}$ be the net % return over month i, year t of some value-weight portfolio, with $i = 1, \ldots, 12$, where $D = SMALL$, BIG, $K = BM$, OP, INV and $J = LOW$, $NEUTRAL$, $HIGH$, with LOW and $HIGH$ denoting growth and value for BM portfolios, weak and robust for OP portfolios, and conservative and aggressive for INV portfolios. We then calculate the quarterly portfolios as:

$$X_{t,I}^{(K,J,D)} = 100 \left[\prod_{i=3(I-1)+1}^{3I} \left(\frac{X_{t,i}^{(K,J,D)}}{100} + 1 \right) - 1 \right],$$

for $I = 1, 2, 3, 4$. Next, we apply the FF factor definitions. Specifically, the small minus big factor is

$$SMB = 1/3(SMB_{BM} + SMB_{OP} + SMP_{INV}),$$

where

$$\begin{aligned} SMB_K &= \frac{X^{(K,LOW,SMALL)} + X^{(K,NEUTRAL,SMALL)} + X^{(K,HIGH,SMALL)}}{3} \\ &\quad - \frac{X^{(K,LOW,BIG)} + X^{(K,NEUTRAL,BIG)} + X^{(K,HIGH,BIG)}}{3}. \end{aligned}$$

Similarly, the high minus low factor is obtained as

$$HML = \frac{X^{(BM,HIGH,SMALL)} + X^{(K,HIGH,BIG)}}{2} - \frac{X^{(BM,LOW,SMALL)} + X^{(K,LOW,BIG)}}{2},$$

the robust minus weak as

$$RMW = \frac{X^{(OP,HIGH,SMALL)} + X^{(OP,HIGH,BIG)}}{2} - \frac{X^{(OP,LOW,SMALL)} + X^{(OP,LOW,BIG)}}{2},$$

and the conservative minus aggressive as

$$CMA = \frac{X^{(INV,LOW,SMALL)} + X^{(INV,LOW,BIG)}}{2} - \frac{X^{(INV,HIGH,SMALL)} + X^{(OP,HIGH,BIG)}}{2}.$$

Finally, the quarterly excess return on the market can be obtained aggregating directly the monthly factor

$$Rm_{t,I} = 100 \left[\prod_{i=3(I-1)+1}^{3I} \left(\frac{Rm_{t,i} + Rf_{t,i}}{100} + 1 \right) - 1 \right] - Rf_{t,I}$$

where $Rf_{t,i}$ and $Rf_{t,I}$ are the one-month and three-month riskfree rate, respectively.

B.4 The symmetry component of the Jarque-Bera (1980) test without imposing normality

Consider a moment test based on the influence function

$$n(y;\pi,\omega)=\epsilon_t^3(\boldsymbol{\theta}_s,0)-3\epsilon_t(\boldsymbol{\theta}_s,0)$$

where $\epsilon_t(\boldsymbol{\theta}_s,0)=\omega^{-1/2}(y_t-\pi)$, evaluated at the sample mean and variance. This influence function coincides with the third Hermite polynomial.

Using standard results (see e.g. Newey and McFadden (1994)), the asymptotic variance of

$$\begin{aligned}&\frac{\sqrt{T}}{T}\sum_{t=1}^{T}n(y_t;\hat{\pi},\hat{\omega})\\ =\;&\frac{\sqrt{T}}{T}\sum_{t=1}^{T}n(y_t;\pi_0,\omega_0)+E\begin{pmatrix}\frac{\partial n(y;\pi_0,\omega_0)}{\partial\pi} & \frac{\partial n(y;\pi_0,\omega_0)}{\partial\omega}\end{pmatrix}\sqrt{T}\begin{pmatrix}\hat{\pi}-\pi_0\\ \hat{\omega}-\omega_0\end{pmatrix}+o_p(1)\end{aligned}$$

But the expected Jacobian matrix evaluated at the true value of the parameters is 0 under symmetry because

$$\begin{aligned}\frac{\partial n(y;\pi,\omega)}{\partial\pi}&=-\frac{3}{\omega^{\frac{1}{2}}}\left[\epsilon_t^2(\boldsymbol{\theta}_s,0)-1\right],\\ \frac{\partial n(y;\pi,\omega)}{\partial\omega}&=-\frac{3}{2\omega}\left[\epsilon_t^2(\boldsymbol{\theta}_s,0)-1\right]\epsilon_t(\boldsymbol{\theta}_s,0).\end{aligned}$$

Therefore, the asymptotic covariance matrix of the sample mean of the third Hermite polynomial evaluated at the sample mean and variance will be the same as if we could evaluate it at the true values. Consequently, a moment test of $H_0: E[n(y;\pi,\omega)]=0$ can be simply computed as the $t-$ratio of the sample mean of $n(y_t;\hat{\pi},\hat{\omega})$.

Interestingly, this moment test coincides with the outer product of the score version of the asymmetry component of the test of the null hypothesis of normality versus generalised hyperbolic alternatives in Mencía and Sentana (2012), which they argue remains valid under as long the true distribution is symmetric.

C Local power calculations

C.1 General results

Let $\mathbf{m}_t(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ denote the h influence functions used to develop the following moment test of $H_0:\boldsymbol{\theta}_2=\mathbf{0}$:

$$M_T=T\bar{\mathbf{m}}_T'(\boldsymbol{\theta}_{10},\mathbf{0})\boldsymbol{\Psi}^{-1}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10},\mathbf{0}),\tag{C10}$$

where $\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10},\mathbf{0})$ is the sample average of $\mathbf{m}_t(\boldsymbol{\theta})$ evaluated under the null, $\boldsymbol{\Psi}$ is the corresponding asymptotic covariance matrix and $\boldsymbol{\theta}_{10}$ the true values of the remaining model parameters. In order to obtain the non-centrality parameter of this test under Pitman sequences of local alternatives of the form $H_0 : \boldsymbol{\theta}_{2T} = \bar{\boldsymbol{\theta}}_2/\sqrt{T}$, it is convenient to linearise $\mathbf{m}_t(\boldsymbol{\theta}_{10}, \mathbf{0})$ with respect to $\boldsymbol{\theta}_2$ around its true value $\boldsymbol{\theta}_{2T}$. This linearisation yields

$$\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}) = \sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{2T}) + \frac{1}{T}\sum\nolimits_{t=1}^{T} \frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_{10}, \boldsymbol{\theta}^*_{2T})}{\partial \boldsymbol{\theta}'_2}\bar{\boldsymbol{\theta}}_2,$$

where $\boldsymbol{\theta}^*_{2T}$ is some "intermediate" value between $\boldsymbol{\theta}_{2T}$ and $\mathbf{0}$. As a result,

$$\sqrt{T}\bar{\mathbf{m}}_T(\boldsymbol{\theta}_{10}, \mathbf{0}) \to N[\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2, \boldsymbol{\Psi}],$$

under standard regularity conditions, where

$$\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0}) = E[\partial \mathbf{m}_t(\boldsymbol{\theta}_{10}, \mathbf{0})/\partial \boldsymbol{\theta}'_2],$$

so that the non-centrality parameter of the moment test (C10) will be

$$\bar{\boldsymbol{\theta}}'_2\mathbf{M}'(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Psi}^{-1}\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2 \tag{C11}$$

when $\boldsymbol{\theta}_{10}$ is known. On this basis, we can easily obtain the limiting probability of M_T exceeding some pre-specified quantile of a central χ^2_h distribution from the cdf of a non-central χ^2 distribution with h degrees of freedom and non-centrality parameter (C11).

Often, though, $\boldsymbol{\theta}_{10}$ will be unknown, and we will have to replace it by some estimator $\bar{\boldsymbol{\theta}}_{1T}$. Let $\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ denote the $\dim(\boldsymbol{\theta}_1)$ influence functions used to estimate $\boldsymbol{\theta}_{10}$ subject to the restriction $\boldsymbol{\theta}_2 = \mathbf{0}$. For convenience, we replace the original influence functions by

$$\mathbf{m}^{\perp}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) - E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_1}\right)\left[E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_1}\right)\right]^{-1}\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2),$$

which are unaffected by the sampling uncertainty in the estimator of $\boldsymbol{\theta}_1$. In this way, the test statistic will be

$$M_T = T\bar{\mathbf{m}}^{\perp\prime}_T(\bar{\boldsymbol{\theta}}_{1T}, \mathbf{0})\boldsymbol{\Upsilon}^{-1}\bar{\mathbf{m}}^{\perp}_T(\bar{\boldsymbol{\theta}}_{1T}, \mathbf{0}),$$

where $\boldsymbol{\Upsilon}$ is the relevant asymptotic covariance matrix, which takes into account the possible (long-run) correlation between $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. As a result, the non-centrality parameter will be

$$\bar{\boldsymbol{\theta}}'_2\mathbf{M}^{\perp\prime}(\boldsymbol{\theta}_{10}, \mathbf{0})\boldsymbol{\Upsilon}^{-1}\mathbf{M}^{\perp}(\boldsymbol{\theta}_{10}, \mathbf{0})\bar{\boldsymbol{\theta}}_2,$$

where

$$\mathbf{M}^{\perp}(\boldsymbol{\theta}_{10}, \mathbf{0}) = E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_2}\right) - E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_1}\right)\left[E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_1}\right)\right]^{-1} E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}'_2}\right).$$

In the special case in which $\bar{\boldsymbol{\theta}}_{1T}$ is the ML estimator of $\boldsymbol{\theta}_{10}$ under the null, and $\mathbf{m}_t(\boldsymbol{\theta}_1, \mathbf{0})$ and the scores corresponding to $\boldsymbol{\theta}_1$ are asymptotically uncorrelated when H_0 is true, as in all our tests under correct specification, then no adjustment will be required because $E[\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}'_1]$ will be 0 by the generalised information matrix equality. In addition, both $\mathbf{M}(\boldsymbol{\theta}_{10}, \mathbf{0})$ and $\boldsymbol{\Psi}$ coincide with the (2,2) block of the information matrix when $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are the scores with respect to $\boldsymbol{\theta}_2$.

If on the other hand $\mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and $\mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ coincide with the scores with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ but these are not uncorrelated under the null, as in our tests under incorrect specification, then we should we work with $\mathbf{m}_t^{\perp}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, although we could still exploit the fact that

$$E\left(\frac{\partial \mathbf{m}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1'}\right)' = E\left(\frac{\partial \mathbf{n}_t(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2'}\right)$$

by the symmetry of the Hessian matrix. In either case, though, the non-centrality parameters of LM and Wald tests will be the same under sequences of local alternatives, at least under the assumption that $\boldsymbol{\theta}_2$ is consistently estimated not only under the null but also under those sequences (see White (1982)).

C.2 Gaussian tests

C.2.1 Serial correlation tests

Let us assume without loss of generality that $\pi = 0$. The first-order serial correlation test is effectively based on the influence functions

$$m_{lt}(\boldsymbol{\theta}_s, \rho) = y_t y_{t-1} - G_{yy}(1)$$

evaluated at $\rho = 0$. But since

$$y_t = \left(1 + \sum\nolimits_{l=1}^{h} \rho L^l\right) \varepsilon_t,$$

we will have that

$$G_{yy}(0) = [1 + (h-1)\rho^2]\sigma^2$$

The Yule-Walker equations of the model considered in (B7) will be given by

$$\begin{array}{rcl}
\frac{G_{yy}(1)}{G_{yy}(0)} & = & \rho\left[1 + \frac{G_{yy}(1)}{G_{yy}(0)} + \ldots + \frac{G_{yy}(h-1)}{G_{yy}(0)}\right] \\
\frac{G_{yy}(2)}{G_{yy}(0)} & = & \rho\left[\frac{G_{yy}(1)}{G_{yy}(0)} + 1 + \ldots + \frac{G_{yy}(h-2)}{G_{yy}(0)}\right] \\
\vdots & & \vdots \\
\frac{G_{yy}(h-1)}{G_{yy}(0)} & = & \rho\left[\frac{G_{yy}(h-2)}{G_{yy}(0)} + \frac{G_{yy}(h-3)}{G_{yy}(0)} + \ldots + \frac{G_{yy}(1)}{G_{yy}(0)}\right]
\end{array}$$

whence

$$G_{yy}(1) = \frac{\rho}{1-(h-1)\rho}[1 + (h-1)\rho^2]\sigma^2.$$

Hence, it trivially follows that

$$M_l(\boldsymbol{\theta}_s, \mathbf{0}) = E[\partial m_{lt}(\boldsymbol{\theta}_s, 0)/\partial \rho] = -\sigma^2.$$

As for the asymptotic covariance matrix, the proof of Proposition 1 implies that if $\rho = 0$, then

$$\sqrt{T} m_{lt}(\boldsymbol{\theta}_s, 0) = \frac{\sqrt{T}}{T}\sum\nolimits_{t=1}^{T} y_t y_{t-1}' \to N(0, \sigma^4)$$

irrespective of the distribution of y_t. As a result, the non-centrality parameter will be ρ^2 regardless of h.

In contrast, the test that uses the influence function

$$y_t \sum_{l=1}^{h} y_{t-l} - \sum_{l=1}^{h} G_{yy}(l)$$

will be asymptotically equivalent to the Wald test based on the Gaussian PML estimator ρ, whose non-centrality parameter is $h\rho^2$, which is clearly bigger than ρ^2 for any $h > 1$.

It is also interesting to study the opposite situation in which we decide to use the influence function that involves $h-$period returns when in fact the true model is an $\textsc{Ar}(1)$. Since $G_{yy}(l) = \rho^l\sigma^2$ in that case, $\sum_{l=1}^{h} G_{yy}(l)$ will be equal to $(1-\rho^{h+1})\sigma^2/(1-\rho)$. Therefore, $M_l(\boldsymbol{\theta}_s, \mathbf{0})$ will also be equal to $-\sigma^2$. But since the asymptotic covariance of the sample average of $y_t \sum_{l=1}^{h} y_{t-l}$ is $h\sigma^4$ under the null, the non-centrality parameter will be $h^{-1}\rho^2$, which is clearly below ρ^2 for any $h > 1$.

C.2.2 GARCH tests

To keep the algebra simple, we assume once again that $\pi = 0$, that the conditional variance has been generated according to a $\textsc{Garch}$(1,1) process and that the conditional distribution has constant kurtosis coefficient κ. The fixed-$\bar{\beta}$ $\textsc{Garch}$ test is based on the following influence function:

$$m_{st}(\sigma^2, \bar{\beta}) = (x_t^2 - \sigma^2) \sum_{j=0}^{\infty} \bar{\beta}^j (x_{t-j}^2 - \sigma^2)$$

As is well known, Bollerslev (1986) showed that a $\textsc{Garch}(1,1)$ model implies the following $\textsc{Arma}(1,1)$ process for x_t^2:

$$(x_t^2 - \sigma^2) = (\alpha + \beta)(x_{t-1}^2 - \sigma^2) + \eta_t - \beta\eta_{t-1},$$

where η_t is the martingale difference sequence $x_t^2 - \sigma_t^2$. As a result,

$$\begin{aligned} V(x_t^2) &= \frac{1 - 2\alpha\beta - \beta^2}{1 - (\alpha + \beta)^2} V(\eta_t), \\ cov(x_t^2, x_{t-1}^2) &= \frac{[1 - (\alpha + \beta)\beta]}{1 - (\alpha + \beta)^2} \alpha V(\eta_t), \end{aligned}$$

and

$$cov(x_t^2, x_{t-j-1}^2) = (\alpha + \beta) cov(x_t^2, x_{t-j}^2) = (\alpha + \beta)^{j-1} cov(x_t^2, x_{t-1}^2)$$

for any $j \geq 1$, so that

$$\begin{aligned} cor(x_t^2, x_{t-1}^2) &= \frac{[1 - (\alpha + \beta)\beta]}{1 - 2\alpha\beta - \beta^2} \alpha, \\ cor(x_t^2, x_{t-j-1}^2) &= (\alpha + \beta)^{j-1} cor(x_t^2, x_{t-1}^2). \end{aligned}$$

But since we know that

$$V(x_t^2) = \frac{1 - 2\alpha\beta - \beta^2}{1 - \kappa\alpha^2 - 2\alpha\beta - \beta^2} (\kappa - 1)\sigma^4$$

when $\kappa\alpha^2 + 2\alpha\beta + \beta^2 < 1$, it immediately follows that

$$V(\eta_t) = \frac{1-(\alpha+\beta)^2}{1-\kappa\alpha^2-2\alpha\beta-\beta^2}(\kappa-1)\sigma^4.$$

As a result, the expected value of $m_{st}(\sigma^2, \bar{\beta})$ under the alternative will be given by

$$\sum\nolimits_{j=0}^{\infty} \bar{\beta}^j (\alpha+\beta)^j E[(x_t^2-\sigma^2)(x_{t-1}^2-\sigma^2)] = \frac{\alpha}{1-\bar{\beta}(\alpha+\beta)} \frac{[1-(\alpha+\beta)\beta]}{1-\kappa\alpha^2-2\alpha\beta-\beta^2}(\kappa-1)\sigma^4.$$

If we expand this expression with respect to α at $\alpha = 0$, we finally obtain

$$\frac{\alpha}{1-\bar{\beta}\beta}(\kappa-1)\sigma^4.$$

Hence, the non-centrality parameter will be

$$\frac{1-\bar{\beta}^2}{(1-\bar{\beta}\beta)^2}\alpha^2.$$

Specifically, for $\bar{\beta} = 0$ the non-centrality parameter will be α^2, while for $\bar{\beta} = 1$ the non-centrality parameter becomes 0 because the regressor has infinite variance while the regressand does not. In fact, $\bar{\beta}$ bigger than $2\beta/(1+\beta^2)$ will result in local power losses relative to $\bar{\beta} = 0$. Not surprisingly, the maximum of this expression is achieved for $\bar{\beta} = \beta$, in which case its value is

$$\frac{\alpha^2}{1-\beta^2},$$

which is bigger than α^2, the more so the closer β is to 1.

Power comparisons To assess the power gains obtained by exploiting the persistence of conditional variances, we compare the Gaussian versions of the Arch(1) and fixed-$\bar{\beta}$ Garch(1,1) tests, and evaluate asymptotic power against *compatible* sequences of local alternatives of the form $\alpha_{0T} = \bar{\alpha}/\sqrt{T}$. Given that the sample variance is consistent for ω, exactly the same results will be obtained if we worked with the transformed sequence $\gamma_{0T} = (\bar{\alpha}\omega_0^{-1})/\sqrt{T} = \bar{\gamma}/\sqrt{T}$.

As we have shown above, when the true model is (B7), the non-centrality parameter of the Gaussian pseudo-score test based on the first order serial correlation coefficient of $\epsilon_t^2(\boldsymbol{\theta}_s, 0)$ is $\bar{\alpha}^2$ regardless of the true value of β. In contrast, the non-centrality parameter of the fixed-$\bar{\beta}$ Garch(1,1) test is $\bar{\alpha}^2(1-\bar{\beta}^2)/(1-\bar{\beta}\beta_0)^2$. Hence, the asymptotic relative efficiency of the two tests is $(1-\bar{\beta}^2)/(1-\bar{\beta}\beta_0)^2$, which is not surprisingly maximised when $\bar{\beta} = \beta_0$. Figure S3A shows that for a realistic value of β_0 these efficiency gains yield substantive power gains when we set $\bar{\beta}$ to its RiskMetrics value of .94

C.3 Student t tests

Under correct specification, the non-centrality parameters are trivial to find because they effectively depend on the $\rho\rho$ or $\alpha\alpha$ elements of the information matrix under the null of mean and variance unpredictability, which we discuss in Lemmas 1 and 2 below. Under distributional misspecification, the calculations are substantially more elaborate.

C.3.1 Normal mixtures

For any given value of the mixing probability λ, the ratio of variances υ and the relative differences in means δ, the first thing we do is to compute the pseudo true values of the Student t pseudo ML estimators under the null, namely π_∞, ω_∞ and η_∞. We obtain these pseudo true values by solving a nonlinear system of three equations that sets to zero the expected value of the scores with respect to π, θ and η. We compute the integrals with respect to the true normal mixture measure as the weighted average of two integrals with respect to the two underlying Gaussian measures, as in Amengual and Sentana (2010). We obtain each of those integrals by Gauss-Hermite quadrature with infinite support using the NAG D01BAF routine with 64 points, $a = \mu_i$ and $b = .5\sigma_i^{-2}$ $(i = 1, 2)$. We solve the resulting nonlinear system of equations in two steps. First, we define a non-uniform grid of 70 values for η between 0.001 and .4995, which is finer close to the two extremes, and then solve the bivariate system for π and ω keeping η fixed. Next, we feed the "best" triplet as starting values for solving the trivariate system using the NAG C05NCF routine.

Once we have thus obtained π_∞, ω_∞ and η_∞, we compute the expected value of the Hessian $(\mathcal{H})$ and variance of the score $(\mathcal{K})$, including the elements involving ρ or γ using the expressions in the proofs of Propositions 1 and 4. We then compute the usual sandwich formulas $\mathcal{H}^{-1}B\mathcal{H}^{-1}$ and take the appropriate diagonal element to obtain the ratio of noncentrality parameters of the Student $t-$based test to the Gaussian one. Although we can repeat these calculations for any possible triplet $(\lambda, \upsilon, \delta)$, in practice we fix $\lambda = .05$ and define a bivariate grid (on a log-scale) on δ and υ of 300×80 points. We then find out the skewness and kurtosis values that those parameters imply using the bounds described in Supplemental Appendix D.1.2.

There are two further controls in the program. On the one hand, when η_∞ is less or equal then 0.001, then we simply set the ratios of noncentrality parameters equal to one. On the other hand, when η_∞ is greater or equal than .4995, then we drop η from the calculations and compute the expected Hessian and variance of the score matrices for the remaining three parameters.

C.3.2 Gram-Charlier expansions

The procedure for the fourth-order Gram-Charlier density is similar to the one we have just described for discrete normal mixtures. The most relevant differences are (i) that the shape parameters of the true measure are now c_3 and c_4, so that we need to find out first the admissible range of values of these parameters which are compatible with a non-negative density; and (ii) the values of a and b in the Gauss-Hermite numerical quadrature NAG D01BAF routine are no longer optimal.

C.4 Power comparisons under correct specification

C.4.1 Serial correlation tests

The following result gives us the necessary ingredients to compare the Gaussian and non-Gaussian tests under correct specification:

Lemma 1 *If the true DGP corresponds to (1) with $\rho_0 = 0$, then the feasible ML estimator of ρ is as efficient as the infeasible ML estimator, which require knowledge of $\boldsymbol{\eta}_0$. In contrast, the inefficiency ratio of the Gaussian PML estimator of ρ is $\mathcal{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$, with $\mathcal{M}_{ll}(\boldsymbol{\eta}_0)$ defined in (E21).*

Proof. The proof is trivial if we combine several results that appear in the proof of Propositions 1. □

This means that Pitman's asymptotic relative efficiency of those serial correlation tests that exploit the non-normality of y_t will be $\mathcal{M}_{ll}^{-1}(\boldsymbol{\eta}_0)$. Figure S2B assesses the power gains against local AR(1) alternatives under the assumption that the true conditional distribution of ε_t^* is a Student t with either 6 or 4.5 degrees of freedom. This figure confirms that the power gains that accrue to our proposed serial correlation tests by exploiting the leptokurtosis of the t distribution are noticeable, the more so the higher the kurtosis of y_t. Similarly, Figure S2C repeats the same exercise for two normal mixtures whose kurtosis coefficients are both 6, and whose skewness coefficients are -.5 and -1.219, respectively. Once again, we can see that there are significant power gains. In this sense, it is worth remembering that since our semiparametric tests are adaptive, they should achieve these gains, at least asymptotically.

C.4.2 Conditional heteroskedasticity tests

The following result gives us the necessary ingredients to compare the Gaussian and non-Gaussian tests under correct specification:

Lemma 2 *If the true DGP corresponds to (5) with $\alpha_0 = 0$, then the feasible ML estimator of α is as efficient as the infeasible ML estimator, which require knowledge of $\boldsymbol{\eta}_0$. In contrast, the inefficiency ratio of the Gaussian PML estimator of α is $4/[(\kappa_0 - 1)\mathcal{M}_{ss}(\boldsymbol{\eta}_0)]$, where $\mathcal{M}_{ss}(\boldsymbol{\eta}_0)$ is defined in (E23).*

Proof. The proof is trivial if we combine several results that appear in the proofs of Propositions 4. □

Lemma 2 then implies that the local non-centrality parameter of the Gaussian test for ARCH is α^2, while the non-centrality parameter of the parametric test for ARCH is $\frac{1}{4}[(\kappa_0 - 1)\mathcal{M}_{ss}(\boldsymbol{\eta}_0)]\alpha^2$. Figure S3B assesses the power gains under the assumption that the true conditional distribution of ε_t^* is a Student t with either 6 or 4.5 degrees of freedom. This figure confirms that the power gains that accrue to our proposed ARCH tests by exploiting the leptokurtosis of the t distribution are in fact more pronounced than the corresponding gains in the mean predictability tests. Similarly, Figure S3C repeats the same exercise for two discrete location scale mixture of normals whose kurtosis coefficients are both 6, and whose skewness coefficients are either -.5 or -1.219. In this case, our tests also yield significant power gains. In this sense, it is worth remembering that since our semiparametric tests are adaptive, they should achieve these gains, at least asymptotically.

D Standardised random variables

D.1 Discrete location scale mixtures of normals

D.1.1 Definition and simulation

Let s_t denote an *i.i.d.* Bernoulli variate with $P(s_t = 1) = \lambda$. If $z_t|s_t$ is *i.i.d.* $N(0,1)$, then

$$\varepsilon_t^* = \frac{1}{\sqrt{1+\lambda(1-\lambda)\delta^2}}\left[\delta(s_t-\lambda)+\frac{s_t+(1-s_t)\sqrt{\upsilon}}{\sqrt{\lambda+(1-\lambda)\upsilon}}z_t\right],$$

where $\delta \in \mathbb{R}$ and $\upsilon > 0$, is a two component mixture of normals whose first two unconditional moments are 0 and 1, respectively. The intuition is as follows. First, note that $\delta(s_t - \lambda)$ is a shifted and scaled Bernoulli random variable with 0 mean and variance $\lambda(1-\lambda)\delta^2$. But since

$$\frac{s_t+(1-s_t)\sqrt{\upsilon}}{\sqrt{\lambda+(1-\lambda)\upsilon}}z_t$$

is a discrete scale mixture of normals with 0 unconditional mean and unit unconditional variance that is orthogonal to $\delta(s_t - \lambda)$, the sum of the two random variables will have variance $1+\lambda(1-\lambda)\delta^2$, which explains the scaling factor.

An equivalent way to define and simulate the same standardised random variable is as follows

$$\varepsilon_t^* = \begin{cases} N[\mu_1^*(\boldsymbol{\eta}), \sigma_1^{*2}(\boldsymbol{\eta})] \text{ with probability } \lambda \\ N[\mu_2^*(\boldsymbol{\eta}), \sigma_2^{*2}(\boldsymbol{\eta})] \text{ with probability } 1-\lambda \end{cases} \tag{D12}$$

where $\boldsymbol{\eta} = (\delta, \upsilon, \lambda)'$ and

$$\begin{aligned}
\mu_1^*(\boldsymbol{\eta}) &= \frac{\delta(1-\lambda)}{\sqrt{1+\lambda(1-\lambda)\delta^2}}, \\
\mu_2^*(\boldsymbol{\eta}) &= -\frac{\delta\lambda}{\sqrt{1+\lambda(1-\lambda)\delta^2}} = -\frac{\lambda}{1-\lambda}\mu_1^*(\boldsymbol{\eta}), \\
\sigma_1^{*2}(\boldsymbol{\eta}) &= \frac{1}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\upsilon]}, \\
\sigma_2^{*2}(\boldsymbol{\eta}) &= \frac{\upsilon}{[1+\lambda(1-\lambda)\delta^2][\lambda+(1-\lambda)\upsilon]} = \upsilon\sigma_1^{*2}(\boldsymbol{\eta}).
\end{aligned}$$

Therefore, we can immediately interpret υ as the ratio of the two variances. Similarly, since

$$\delta = \frac{\mu_1^*(\boldsymbol{\eta}) - \mu_2^*(\boldsymbol{\eta})}{\sqrt{\lambda\sigma_1^{*2}(\boldsymbol{\eta}) + (1-\lambda)\sigma_1^{*2}(\boldsymbol{\eta})}},$$

we can also interpret δ as the parameter that regulates the distance between the means of the two underlying components relative to the mean of the two conditional variances.

We can trivially extended this procedure to define and simulate standardised mixtures with three or more components. Specifically, if we replace the normal random variable in the first branch of (D12) by a k-component normal mixture with mean and variance given by $\mu_1^*(\boldsymbol{\eta})$ and $\sigma_1^{*2}(\boldsymbol{\eta})$, respectively, then the resulting random variable will be a $(k+1)$-component Gaussian mixture with zero mean and unit variance.

Finally, note that we can also use the above expressions to generate a two component mixture of normals with mean π and variance ω^2 as

$$y_t = \begin{cases} N(\mu_1, \sigma_1^2) \text{ with probability } \lambda \\ N(\mu_2, \sigma_2^2) \text{ with probability } 1-\lambda \end{cases}$$

with

$$\begin{aligned} \mu_1 &= \pi + \omega\mu_1^*(\boldsymbol{\eta}) \\ \mu_2 &= \pi + \omega\mu_2^*(\boldsymbol{\eta}) \\ \sigma_1^2 &= \omega\sigma_1^{*2}(\boldsymbol{\eta}), \\ \sigma_2^2 &= \omega\sigma_2^{*2}(\boldsymbol{\eta}). \end{aligned}$$

Interestingly, the expressions for υ and δ above continue to be valid if we replace $\mu_1^*(\boldsymbol{\eta})$, $\mu_2^*(\boldsymbol{\eta})$, $\sigma_1^{*2}(\boldsymbol{\eta})$ and $\sigma_2^{*2}(\boldsymbol{\eta})$ by μ_1, μ_2, σ_1^2 and σ_2^2.

D.1.2 Skewness-kurtosis bounds

In the case of two-component Gaussian mixtures, the parameters λ, δ and υ determine the higher order moments of ε_t^* through the relationship

$$E(\varepsilon_t^{*j}) = \lambda E(\varepsilon_t^{*j}|s_t = 1) + (1-\lambda)E(\varepsilon_t^{*j}|s_t = 0),$$

where $E(\varepsilon_t^{*j}|s_t = 1)$ can be obtained from the usual normal expressions

$$\begin{aligned} E(\varepsilon_t^*|s_t = 1) &= \mu_1^*(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*2}|s_t = 1) &= \mu_1^{*2}(\boldsymbol{\eta}) + \sigma_1^{*2}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*3}|s_t = 1) &= \mu_1^{*3}(\boldsymbol{\eta}) + 3\mu_1^*(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*4}|s_t = 1) &= \mu_1^{*4}(\boldsymbol{\eta}) + 6\mu_1^{*2}(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) + 3\sigma_1^{*4}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*5}|s_t = 1) &= \mu_1^{*5}(\boldsymbol{\eta}) + 10\mu_1^{*3}(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) + 15\mu_1^*(\boldsymbol{\eta})\sigma_1^{*4}(\boldsymbol{\eta}) \\ E(\varepsilon_t^{*6}|s_t = 1) &= \mu_1^{*6}(\boldsymbol{\eta}) + 15\mu_1^{*4}(\boldsymbol{\eta})\sigma_1^{*2}(\boldsymbol{\eta}) + 45\mu_1^{*2}(\boldsymbol{\eta})\sigma_1^{*4}(\boldsymbol{\eta}) + 15\sigma_1^{*6}(\boldsymbol{\eta}) \end{aligned}$$

etc. But since $E(\varepsilon_t^*) = 0$ and $E(\varepsilon_t^{*2}) = 1$ by construction, straightforward algebra shows that the skewness and kurtosis coefficients will be given by

$$E(\varepsilon_t^{*3}) = \frac{3\delta\lambda(1-\lambda)(1-\upsilon)}{[\lambda+(1-\lambda)\upsilon][1+\lambda(1-\lambda)\delta^2]^{3/2}} + \frac{\delta^3(1-\lambda)\lambda(1-2\lambda)}{[1+\lambda(1-\lambda)\delta^2]^{3/2}} = a(\delta,\upsilon,\lambda) \tag{D13}$$

and

$$\begin{aligned} E(\varepsilon_t^{*4}) &= \frac{3[\lambda+(1-\lambda)\upsilon^2]}{[\lambda+(1-\lambda)\upsilon]^2[1+\lambda(1-\lambda)\delta^2]^2} + \frac{6\delta^2\lambda(1-\lambda)[(1-\lambda)+\upsilon\lambda]}{[\lambda+(1-\lambda)\upsilon][1+\lambda(1-\lambda)\delta^2]^2} \\ &\quad + \frac{\delta^4\lambda(1-\lambda)[1-3\lambda(1-\lambda)]}{[1+\lambda(1-\lambda)\delta^2]^2} = b(\delta,\upsilon,\lambda). \end{aligned} \tag{D14}$$

Two issues are worth pointing out. First, $a(\delta, \upsilon, \lambda)$ is an odd function of δ, which means that δ and $-\delta$ yield the same skewness in absolute value. In this sense, if we set $\delta = 0$ then we will obtain a discrete scale mixture of normals, which is always symmetric but leptokurtic. Another way of obtaining discrete normal mixture distributions that are symmetric is by making $\lambda = \frac{1}{2}$ and $\upsilon = 1$. Second, $b(\delta, \upsilon, \lambda)$ is an even function of δ, which implies that δ and $-\delta$ give rise to the same kurtosis. For that reason, in what follows we mostly consider the case of $\delta \geq 0$.

A useful property of two component normal mixtures is that they span the entire unconditional skewness-kurtosis frontier given by the parabola $E(\varepsilon_t^{*4}) \geq 1 + E^2(\varepsilon_t^{*3})$ (see Stuart and Ord (1977)). More specifically, for a fixed value of λ, skewness, which is 0 for $\delta = 0$, reaches its frontier value as $\delta \to \infty$, in which case

$$\lim_{\delta \to \infty} a(\delta, \upsilon, \lambda) = \frac{2(\frac{1}{2} - \lambda)}{\sqrt{\lambda(1-\lambda)}}$$

regardless of υ. Clearly, for $\lambda < .5$ this limiting skewness value is positive, while it is negative for $\lambda > .5$. In any case, we can achieve the mirror point on the frontier as $\delta \to -\infty$.

The corresponding kurtosis values are

$$b(0, \upsilon, \lambda) = \frac{3(\lambda + (1-\lambda)\upsilon^2)}{(\lambda + (1-\lambda)\upsilon)^2} = 3\left(\frac{\lambda(1-\lambda)(1-\upsilon)^2}{(\lambda + (1-\lambda)\upsilon)^2} + 1\right)$$

and

$$\lim_{\delta \to \pm\infty} b(\delta, \upsilon, \lambda) = -3 + \frac{1}{\lambda(1-\lambda)} = 1 + \left(\frac{2(\frac{1}{2} - \lambda)}{\sqrt{\lambda(1-\lambda)}}\right)^2,$$

which again does not depend on υ. Intuitively, the reason is that a standardised two component normal mixture converges in distribution to a standardised Bernoulli random variable with parameter λ as $\delta \to \infty$ regardless of υ. Interestingly, $\lim_{\delta \to \infty} b(\delta, \upsilon, \lambda) = 3$ for $\lambda = \frac{1}{2} \pm \frac{1}{6}\sqrt{3}$.

Nevertheless, to create Figures 2B and 4B, we need to find out the range of skewness and kurtosis that this distribution can generate when λ is fixed. In this sense, notice that kurtosis is always larger or equal than 3 for $\delta = 0$, which reflects the fact that a scale mixture of normals is always leptokurtic. The boundary case is of course $\upsilon = 1$, in which case

$$b(0, 1, \lambda) = 3.$$

In fact, maximum kurtosis when $\delta = 0$ is achieved for $\upsilon = 0$ or for $\upsilon \to \infty$, in which case we obtain either

$$b(0, 0, \lambda) = \frac{3}{\lambda} \text{ or } \lim_{\upsilon \to \infty} b(0, \upsilon, \lambda) = \frac{3}{1-\lambda}.$$

Obviously, this kurtosis can be made arbitrarily large as λ approaches 0 or 1, but it is clearly bounded for fixed λ.

The other interesting cases arise when $\upsilon = 0$ and $\upsilon = 1$. In the first case

$$a(\delta, 0, \lambda) = \delta\,(1-\lambda)\,\frac{3 + (1-2\lambda)\lambda\delta^2}{(1 + \lambda(1-\lambda)\delta^2)^{3/2}}$$

and

$$b(\delta,0,\lambda) = \frac{3}{\lambda(1+\lambda(1-\lambda)\delta^2)^2} + \frac{6\delta^2(1-\lambda)^2}{(1+\lambda(1-\lambda)\delta^2)^2} + \frac{\delta^4\lambda(1-\lambda)(1-3\lambda(1-\lambda))}{(1+\lambda(1-\lambda)\delta^2))^2},$$

while in the second case

$$a(\delta,1,\lambda) = \frac{\delta^3(1-\lambda)\lambda(1-2\lambda)}{(1+\lambda(1-\lambda)\delta^2)^{3/2}}$$

and

$$b(\delta,1,\lambda) = \frac{3}{(1+\lambda(1-\lambda)\delta^2)^2} + \frac{6\delta^2\lambda(1-\lambda)}{(1+\lambda(1-\lambda)\delta^2)^2} + \frac{\delta^4\lambda(1-\lambda)(1-3\lambda(1-\lambda))}{(1+\lambda(1-\lambda)\delta^2))^2}.$$

It turns out that the range of skewness and kurtosis that a standardised mixture of two normals can generate seems to be bounded by the following two parametric curves:

$$(a(\delta,1,\lambda), b(\delta,1,\lambda))$$

and

$$(a(\delta,0,\lambda), b(\delta,0,\lambda)),$$

where the range of δ is $[0,\infty)$. In fact, these curves intersect at the unconditional skewness-frontier boundary when $\delta \to \infty$.

Interestingly, it seems that skewness is always non-negative when $\lambda \leq 1/2$. In contrast, for $\lambda > 1/2$ skewness is initially positive for small values of δ, but then becomes negative as δ increases. In turn, kurtosis bounded from below by 3 when $\lambda \leq \frac{1}{2} - \frac{1}{6}\sqrt{3}$, while it is bounded from above by 3 on the negative skewness side if $\frac{1}{2} \leq \lambda \leq \frac{1}{2} + \frac{1}{6}\sqrt{3}$.

As we explained before, the mirror curves

$$(-a(|\delta|,1,\lambda), b(|\delta|,1,\lambda))$$

and

$$(-a(|\delta|,0,\lambda), b(|\delta|,0,\lambda)),$$

give us the skewness-kurtosis range when δ if negative.

D.2 Gram-Charlier distributions

D.2.1 Definition and moments

The first five raw Hermite polynomials are:

$$\begin{aligned} H_0(z) &= 1, \\ H_1(z) &= z, \\ H_2(z) &= z^2 - 1, \\ H_3(z) &= z^3 - 3z, \\ H_4(z) &= z^4 - 6z^2 + 3. \end{aligned}$$

When $z \sim N(0,1)$, these have 0 mean and are orthogonal to each other. In turn,

$$\begin{aligned} H_2^*(z) &= \frac{z^2-1}{\sqrt{2}}, \\ H_3^*(z) &= \frac{z^3-3z}{\sqrt{6}}, \\ H_4^*(z) &= \frac{z^4-6z^2+3}{\sqrt{24}}. \end{aligned}$$

are called the standardised Hermite polynomials because their variance will be 1 for a standard normal.

The Gram-Charlier density is defined as:

$$\begin{aligned} f(z) &= \phi(z)P(z), && \text{(D15)} \\ \phi(z) &= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}, \\ P(z) &= 1+\frac{\varphi}{\sqrt{6}}H_3^*(z)+\frac{\upsilon}{\sqrt{24}}H_4^*(z) = 1+\frac{\varphi}{6}\left(z^3-3z\right)+\frac{\kappa}{24}\left(z^4-6z^2+3\right). && \text{(D16)} \end{aligned}$$

This density is such that

$$\begin{aligned} E_f(z) &= 0, \\ E_f(z^2) &= 1, \\ E_f(z^3) &= \varphi, \\ E_f(z^4) &= 3+\kappa. \end{aligned}$$

D.2.2 Positivity restrictions

The problem is that $P(z)$ in (D16) can be negative, in which case $f(z)$ in (D15) will not be a proper density.

For a given z, the skewness-excess kurtosis frontier that guarantees positivity must satisfy the following two equations:

$$\begin{aligned} 1+\frac{\varphi}{6}\left(z^3-3z\right)+\frac{\kappa}{24}\left(z^4-6z^2+3\right) &= 0, \\ \frac{\varphi}{2}\left(z^2-1\right)+\frac{\kappa}{6}\left(z^3-3z\right) &= 0. \end{aligned}$$

The first equation, which is given by $P(z)=0$, defines a straight line in (φ,κ) space such that in any neighbourhood of the solution we will find positive and negative densities. In contrast, the second equation, which corresponds to $\partial P(z)/\partial z = 0$, guarantees that we remain in the frontier as we move in (φ,κ) space.

The solution to the above system of equations in terms of φ and κ as a function of z is

$$\begin{aligned} \varphi(z) &= -24\frac{z^3-3z}{z^6-3z^4+9z^2+9}, \\ \kappa(z) &= 72\frac{z^2-1}{z^6-3z^4+9z^2+9}, \end{aligned}$$

where the denominator is

$$d(z) = 4\left(z^3 - 3z\right)^2 - 3\left(z^2 - 1\right)\left(z^4 - 6z^2 + 3\right) = z^6 - 3z^4 + 9z^2 + 9.$$

This solution can be regarded as the parametric representation of the admissible skewness-kurtosis frontier.

The simplest way to find the frontier values is to carry out a grid over z, and for each value of z find out the corresponding values of $\varphi(z)$ and $\kappa(z)$. However, this does not work as expected because we will often end up with two different values of $\varphi(z)$ for the same value of $\kappa(z)$. Following Jondeau and Rockinger (2001), the solution is to restrict the range of z to be $[\sqrt{3}, \infty)$. When $z = \sqrt{3}$, $\varphi(z)$ and $\kappa(z)$ become 0 and 4, respectively. In contrast, when $z \to \infty$ both $\varphi(z)$ and $\kappa(z)$ converge to 0. In practice, the grid should probably be logarithmic between $\sqrt{3}$ and 10^3 or so. The maximum skewness that can be achieved is 1.0493. Obviously, we get the mirror image by changing the sign of z.

D.3 Simulation

A very simple way to simulate random variables with a Gram-Charlier distribution is by using the usual inversion method, which exploits the fact that if Z is a random variable with absolutely continuous distribution function $F_Z(.)$ and quantile function $F_Z^{-1}(.)$, then $U = F_Z(Z)$ is uniformly distributed between 0 and 1, while $F_Z^{-1}(U)$ will follow the distribution of Z.

Given that

$$\int H_k^*(x)\,\phi(x)\,dx = \frac{-1}{\sqrt{k}} H_{k-1}^*(x)\,\phi(x) \qquad k \geq 1 \tag{D17}$$

(see León, Mencía and Sentana (2009)), and that $H_k^*(x)\,\phi(x) \to 0$ when $x \to -\infty$ by virtue of L'Hôpital rule, then

$$\int_{-\infty}^{z} H_k^*(x)\,\phi(x)\,dz = -\frac{1}{\sqrt{k}} H_{k-1}^*(z)\,\phi(z), \qquad k \geq 1. \tag{D18}$$

Consequently,

$$F_Z(z) = \int_{-\infty}^{z} f(x)dx = \int_{-\infty}^{z} \phi(x)P(x)dx = \Phi(z) - \frac{\varphi}{6} H_2^*(z)\,\phi(z) - \frac{\kappa}{24} H_3^*(z)\,\phi(z).$$

In practice, we simulate a uniform variate u, and numerically solve the equation

$$F_Z(z) = u$$

with $\Phi^{-1}(u)$ as starting value.

D.4 Generalised hyperbolic

Let ξ_t denote an *i.i.d.* Generalised Inverse Gaussian (GIG) random variable with parameters $-\nu, \tau$ and 1, or $GIG(-\nu, \tau, 1)$ for short. Mencía and Sentana (2012) show that if $z_t|\xi_t$ is *i.i.d.* $N(0,1)$, then

$$\varepsilon_t^* = c(\beta,\nu,\tau)\beta\left[\frac{\tau\xi_t^{-1}}{R_\nu(\tau)} - 1\right] + \sqrt{\frac{\tau\xi_t^{-1}}{R_\nu(\tau)}}\sqrt{c(\beta,\nu,\tau)}z_t$$

is a standardised Generalised Hyperbolic (GH) distribution with parameters β, ν and τ, where

$$\begin{aligned} c(\beta,\nu,\tau) &= \frac{-1+\sqrt{1+4\beta^2[D_{\nu+1}(\tau)-1]}}{2\beta^2[D_{\nu+1}(\tau)-1]} \\ R_\nu(\tau) &= \frac{K_{\nu+1}(\tau)}{K_\nu(\tau)}, \\ D_{\nu+1}(\tau) &= \frac{K_{\nu+2}(\tau)K_\nu(\tau)}{K_{\nu+1}(\tau)}, \end{aligned}$$

and $K_\nu(.)$ is the modified Bessel function of the third kind. In turn, the GH distribution is a special case of the more general location scale mixtures of normals considered in Mencía and Sentana (2009), in which ξ_t is a positive random variable with an arbitrary distribution.

Mencía and Sentana (2012) also provide expressions for the third and fourth moments of the GH distribution, which in the univariate case reduce to

$$E(\varepsilon_t^{*3}) = c^3(\beta,\nu,\tau)\left[\frac{K_{\nu+3}(\tau)\,K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} - 3D_{\nu+1}(\tau) + 2\right]\beta^3 + 3c^2(\beta,\nu,\tau)\left[D_{\nu+1}(\tau)-1\right]\beta$$

and

$$\begin{aligned} E(\varepsilon_t^{*4}) = c^4(\beta,\nu,\tau)&\left[\frac{K_{\nu+4}(\tau)\,K_\nu^3(\tau)}{K_{\nu+1}^4(\tau)} - 4\frac{K_{\nu+3}(\tau)\,K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} + 6D_{\nu+1}(\tau) - 3\right]\beta^4 \\ +6c^3(\beta,\nu,\tau)&\left[\frac{K_{\nu+3}(\tau)\,K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} - 2D_{\nu+1}(\tau)+1\right]\beta^2 + 3D_{\nu+1}(\tau)\,c^2(\beta,\nu,\tau). \end{aligned}$$

D.4.1 Asymmetric and symmetric versions of the Student t

The asymmetric t distribution is nested within the GH family when $\tau = 0$ and $-\infty < \nu < -2$. If we define $\eta = -1/(2\nu)$, then for $\eta < 1/4$ we will have that

$$\begin{aligned} c(\beta,\nu,\tau) &= \frac{1-4\eta}{2\eta}\,\frac{\sqrt{1+8\beta^2\eta/(1-4\eta)}-1}{2\beta^2}, \\ \lim_{\tau\to\infty}\frac{R_\nu(\tau)}{\tau} &= \lim_{\tau\to\infty}\frac{K_{\nu+1}(\tau)}{\tau K_\nu(\tau)} = \frac{\eta}{1-2\eta}, \\ D_{\nu+1}(\tau) &= \frac{K_{\nu+2}(\tau)K_\nu(\tau)}{K_{\nu+1}(\tau)} = \frac{1-2\eta}{1-4\eta}. \end{aligned}$$

Therefore, we can easily simulate an asymmetric standardised Student t distribution as:

$$\varepsilon_t^* = c(\beta,\nu,\tau)\beta\left[\frac{(1-2\eta)}{\eta\xi_t} - 1\right] + \sqrt{\frac{(1-2\eta)}{\eta\xi_t}}\sqrt{c(\beta,\nu,\tau)}z_t,$$

where $\xi_t \sim i.i.d.\ Gamma$ with mean η^{-1} and variance $2\eta^{-1}$, and $z_t|\xi_t$ is $i.i.d.\ N(0,1)$.

If we further assume that $\eta < 1/8$, then

$$\begin{aligned} \frac{K_{\nu+3}(\tau)\,K_\nu^2(\tau)}{K_{\nu+1}^3(\tau)} &= \frac{(1-2\eta)^2}{(1-4\eta)(1-6\eta)} \\ \frac{K_{\nu+4}(\tau)\,K_\nu^3(\tau)}{K_{\nu+1}^4(\tau)} &= \frac{(1-2\eta)^3}{(1-4\eta)(1-6\eta)(1-8\eta)} \end{aligned}$$

so the skewness and kurtosis coefficients of the asymmetric t distribution will be:

$$E(\varepsilon_t^{*3}) = 16c^3(\beta,\nu,\tau)\frac{\eta^2}{(1-4\eta)(1-6\eta)}\beta^3 + 6c^2(\beta,\nu,\tau)\frac{\eta}{1-4\eta}\beta$$

and

$$E(\varepsilon_t^{*4}) = 12c^4(\beta,\nu,\tau)\frac{\eta^2(10\eta+1)}{(1-4\eta)(1-6\eta)(1-8\eta)}\beta^4$$
$$+12c^3(\beta,\nu,\tau)\frac{\eta(2\eta+1)}{(1-4\eta)(1-6\eta)}\beta^2 + 3\frac{1-2\eta}{1-4\eta}c^2(\beta,\nu,\tau).$$

Not surprisingly, we can obtain maximum asymmetry for a given kurtosis by letting $|\beta| \to \infty$. In contrast, a standardised version of the usual symmetric Student t with $1/\eta$ degrees of freedom is achieved when $\beta = 0$. Since $\lim_{\beta\to 0} c(\beta,\nu,\tau) = 1$, in that case the coefficient of kurtosis becomes

$$E(\varepsilon_t^{*4}) = 3\frac{1-2\eta}{1-4\eta}$$

for any $\eta < 1/4$.

D.4.2 Symmetric Laplace distribution

The asymmetric Laplace distribution is another special case of the GH distribution, which is achieved when $\tau = 0$ and $\nu = 1$. In fact, it is a special case of the asymmetric normal-gamma mixture, which allows ν to be any positive parameter. As is well known, the kurtosis coefficient of a symmetric Laplace distribution is 6. In the univariate case, the Laplace distribution is also a special case of the generalised error distribution (GED) with shape parameter fixed at 1, in contrast to the Gaussian distribution, which is also a special GED case with parameter 2.

The symmetric Laplace distribution is very easy to generate as

$$\varepsilon_t^* = \sqrt{\xi_t} z_t,$$

where ξ_t is an *i.i.d.* exponential (i.e. a *Gamma* with mean 1 and variance 1), and $z_t|\xi_t$ is *i.i.d.* $N(0,1)$. Alternatively, if u_t denotes a $(0,1)$ uniform variate, then we can also simulate a standardised symmetric Laplace random variable ε_t^* as

$$-\frac{1}{\sqrt{2}} sign\left(u_t - \frac{1}{2}\right) \ln\left(1 - 2\left|u_t - \frac{1}{2}\right|\right).$$

In effect, this procedure uses the fact that the absolute value of a Laplace is exponential, with a closed-form quantile function, while its sign is a shifted and scaled Bernoulli random variable that the values ± 1 with probability 1/2 each.

E Econometric methods

E.1 Log-likelihood function, score vector, Hessian and information matrices

Let $\boldsymbol{\phi} = (\boldsymbol{\theta}', \boldsymbol{\eta})'$ denote the $p + r$ parameters of interest, which we assume variation free. Ignoring initial conditions, and assuming that $\sigma_t^2(\boldsymbol{\theta})$ is strictly positive, the log-likelihood function of a sample of size T based on a particular parametric distributional assumption will take the form $L_T(\boldsymbol{\phi}) = \sum_{t=1}^{T} l_t(\boldsymbol{\phi})$, with $l_t(\boldsymbol{\phi}) = d_t(\boldsymbol{\theta}) + \ln f\left[\epsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$, where $d_t(\boldsymbol{\theta}) = -1/2 \ln \sigma_t^2(\boldsymbol{\theta})$, $\varepsilon_t^*(\boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\theta})/\sigma_t(\boldsymbol{\theta})$ and $\varepsilon_t(\boldsymbol{\theta}) = y_t - \mu_t(\boldsymbol{\theta})$.

Let $\mathbf{s}_t(\boldsymbol{\phi})$ denote the score function $\partial l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, and partition it into two blocks, $\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi})$ and $\mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi})$, whose dimensions conform to those of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, respectively. If $\mu_t(\boldsymbol{\theta})$, $\sigma_t^2(\boldsymbol{\theta})$ and $f(\varepsilon^*, \boldsymbol{\eta})$ are differentiable, then we can use the fact that

$$\partial d_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = -\tfrac{1}{2} \cdot \sigma_t^{-2}(\boldsymbol{\theta}) \cdot \partial\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = -\mathbf{Z}_{st}(\boldsymbol{\theta})$$

and

$$\begin{aligned} \partial\varepsilon_t^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta} &= -\sigma_t^{-1}(\boldsymbol{\theta}) \cdot \partial\mu_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} - \tfrac{1}{2}\sigma_t^{-2}(\boldsymbol{\theta}) \cdot \partial\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \cdot \varepsilon_t^*(\boldsymbol{\theta}) \\ &= -\mathbf{Z}_{lt}(\boldsymbol{\theta}) - \mathbf{Z}_{st}(\boldsymbol{\theta})\varepsilon_t^*(\boldsymbol{\theta}), \end{aligned}$$

to show that

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\phi}) &= \frac{\partial d_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + \frac{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial\boldsymbol{\theta}} = \left[\mathbf{Z}_{lt}(\boldsymbol{\theta}), \mathbf{Z}_{st}(\boldsymbol{\theta})\right] \left[\begin{array}{c} e_{lt}(\boldsymbol{\phi}) \\ e_{st}(\boldsymbol{\phi}) \end{array}\right] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathbf{e}_{dt}(\boldsymbol{\phi}), \\ \mathbf{s}_{\boldsymbol{\eta}t}(\boldsymbol{\phi}) &= \partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\boldsymbol{\eta} = \mathbf{e}_{rt}(\boldsymbol{\phi}), \end{aligned}$$

where

$$\begin{aligned} e_{lt}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= -\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^*, \\ e_{st}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= -\left\{1 + \varepsilon_t^*(\boldsymbol{\theta}) \cdot \partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^*\right\}, \end{aligned}$$

depend on the specific distributional assumption.

Let $\mathbf{h}_t(\boldsymbol{\phi})$ denote the Hessian function $\partial\mathbf{s}_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' = \partial^2 l_t(\boldsymbol{\phi})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'$. Assuming twice differentiability of the different functions involved, we will have

$$\begin{aligned} \mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\phi}) &= \frac{\partial\mathbf{Z}_{lt}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} e_{lt}(\boldsymbol{\phi}) + \frac{\partial\mathbf{Z}_{st}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} e_{st}(\boldsymbol{\phi}) + \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial e_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial e_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\theta}'} \quad \text{(E19)} \\ \mathbf{h}_{\boldsymbol{\theta\eta}t}(\boldsymbol{\phi}) &= \mathbf{Z}_{lt}(\boldsymbol{\theta})\frac{\partial e_{lt}(\boldsymbol{\phi})}{\partial\boldsymbol{\eta}'} + \mathbf{Z}_{st}(\boldsymbol{\theta})\frac{\partial e_{st}(\boldsymbol{\phi})}{\partial\boldsymbol{\eta}'} \quad \text{(E20)} \\ \mathbf{h}_{\boldsymbol{\eta\eta}t}(\boldsymbol{\phi}) &= \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}', \end{aligned}$$

where

$$\partial\mathbf{Z}_{lt}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = -\tfrac{1}{2} \cdot \sigma_t^{-3}(\boldsymbol{\theta}) \cdot \partial\mu_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \cdot \partial\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' + \sigma_t^{-1}(\boldsymbol{\theta}) \cdot \partial^2\mu_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}',$$

$$\partial\mathbf{Z}_{st}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' = -\tfrac{1}{2} \cdot \sigma_t^{-4}(\boldsymbol{\theta}) \cdot \partial\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \cdot \partial\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' + \tfrac{1}{2} \cdot \sigma_t^{-2}(\boldsymbol{\theta}) \cdot \partial^2\sigma_t^2(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}',$$

$$\partial e_{lt}(\boldsymbol{\phi})/\partial\boldsymbol{\theta}' = \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \mathbf{Z}_{lt}'(\boldsymbol{\theta}) + \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta}) \cdot \mathbf{Z}_{st}'(\boldsymbol{\theta})$$

$$\begin{aligned} \partial e_{st}(\boldsymbol{\phi})/\partial\boldsymbol{\theta}' &= \left\{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^* + \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta})\right\}\mathbf{Z}_{lt}'(\boldsymbol{\theta}) \\ &+ \left\{\partial \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta}) + \partial^2 \ln f\left[\varepsilon_t^*(\boldsymbol{\theta}), \boldsymbol{\eta}\right]/\partial\varepsilon^*\partial\varepsilon^* \cdot \varepsilon_t^{2*}(\boldsymbol{\theta})\right\} \cdot \mathbf{Z}_{st}'(\boldsymbol{\theta}) \end{aligned}$$

and $\partial^2 \ln f(\varepsilon^*, \eta)/\partial\varepsilon^*\partial\varepsilon^*$, $\partial^2 \ln f(\varepsilon^*, \eta)/\partial\varepsilon^*\partial\boldsymbol{\eta}'$ and $\partial \ln f(\varepsilon^*, \eta)/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'$ depend on the specific distribution assumed for estimation purposes (see Fiorentini, Sentana and Calzolari (2003) for the Student t).

Given correct specification, $\mathbf{e}_t(\boldsymbol{\phi}) = [\mathbf{e}'_{dt}(\boldsymbol{\phi}), \mathbf{e}_{rt}(\boldsymbol{\phi})]'$ evaluated at the true parameter values is an *iid* sequence, and therefore, the score vector $\mathbf{s}_t(\boldsymbol{\phi})$ will be a vector martingale difference sequence. Then, the results in Crowder (1976) imply that, under suitable regularity conditions, the asymptotic distribution of the feasible ML estimator will be $\sqrt{T}(\boldsymbol{\phi}_T - \boldsymbol{\phi}_0) \to N[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\phi}_0)]$, where $\mathcal{I}(\boldsymbol{\phi}_0) = E[\mathcal{I}_t(\boldsymbol{\phi}_0)|\boldsymbol{\phi}_0]$, where

$$\begin{aligned}
\mathcal{I}_t(\boldsymbol{\phi}) &= -E\left[\mathbf{h}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\right] = V\left[\mathbf{s}_t(\boldsymbol{\phi})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}\right] = \mathbf{Z}_t(\boldsymbol{\theta})\mathcal{M}(\boldsymbol{\eta})\mathbf{Z}'_t(\boldsymbol{\theta}), \\
\mathbf{Z}_t(\boldsymbol{\theta}) &= \begin{pmatrix} \mathbf{Z}_{dt}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_{lt}(\boldsymbol{\theta}) & \mathbf{Z}_{st}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_q \end{pmatrix},
\end{aligned}$$

and

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \mathcal{M}_{ll}(\boldsymbol{\eta}) & \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{lr}(\boldsymbol{\eta}) \\ \mathcal{M}_{ls}(\boldsymbol{\eta}) & \mathcal{M}_{ss}(\boldsymbol{\eta}) & \mathcal{M}_{sr}(\boldsymbol{\eta}) \\ \mathcal{M}'_{lr}(\boldsymbol{\eta}) & \mathcal{M}'_{sr}(\boldsymbol{\eta}) & \mathcal{M}_{rr}(\boldsymbol{\eta}) \end{pmatrix},$$

with

$$\mathcal{M}_{ll}(\boldsymbol{\eta}) = V[e_{lt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[\partial^2 \ln f(\varepsilon_t^*; \boldsymbol{\eta})/\partial\varepsilon^*\partial\varepsilon^{*\prime}\middle|\,\boldsymbol{\eta}\right], \tag{E21}$$

$$\mathcal{M}_{ls}(\boldsymbol{\eta}) = E[e_{lt}(\boldsymbol{\phi})e_{st}(\boldsymbol{\phi})'|\boldsymbol{\phi}] = E\left[\varepsilon_t^* \cdot \partial^2 \ln f(\varepsilon_t^*; \boldsymbol{\eta})/\partial\varepsilon^*\partial\varepsilon^{*\prime}\middle|\,\boldsymbol{\eta}\right], \tag{E22}$$

$$\mathcal{M}_{ss}(\boldsymbol{\eta}) = V[e_{st}(\boldsymbol{\phi})|\boldsymbol{\phi}] = E\left[\varepsilon_t^{*2} \cdot \partial^2 \ln f(\varepsilon_t^*; \boldsymbol{\eta})/\partial\varepsilon^*\partial\varepsilon^{*\prime}|\boldsymbol{\eta}\right] - 1, \tag{E23}$$

$$\mathcal{M}_{lr}(\boldsymbol{\eta}) = E[e_{lt}(\boldsymbol{\phi})e'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\varepsilon_t^*; \boldsymbol{\eta})/\partial\varepsilon^*\partial\boldsymbol{\eta}'|\boldsymbol{\eta}\right],$$

$$\mathcal{M}_{sr}(\boldsymbol{\eta}) = E[e_{st}(\boldsymbol{\phi})e'_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\varepsilon_t^*\partial^2 \ln f(\varepsilon_t^*; \boldsymbol{\eta})/\partial\varepsilon^*\partial\boldsymbol{\eta}'|\boldsymbol{\eta}\right],$$

and

$$\mathcal{M}_{rr}(\boldsymbol{\eta}) = V[\mathbf{e}_{rt}(\boldsymbol{\phi})|\boldsymbol{\phi}] = -E\left[\partial^2 \ln f(\varepsilon_t^*; \boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'|\boldsymbol{\phi}\right].$$

In the Student t case, this matrix is simply

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} \frac{\nu(\nu+1)}{(\nu-2)(\nu+3)} & 0 & 0 \\ 0 & \frac{(\nu+1)}{(\nu+3)} & -\frac{6\nu^2}{(\nu-2)(\nu+1)(\nu+3)} \\ 0 & -\frac{6\nu^2}{(\nu-2)(\nu+1)(\nu+3)} & \frac{\nu^4}{4}\left[\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right)\right] - \frac{\nu^4\left[\nu^2+(\nu-4)-8\right]}{2(\nu-2)^2(\nu+1)(\nu+3)} \end{pmatrix}.$$

where $\psi(.)$ is the di-gamma function (see Abramowitz and Stegun (1964)), which under normality reduces to

$$\mathcal{M}(\boldsymbol{\eta}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3/2 \end{pmatrix}.$$

E.2 Gaussian pseudo maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}} L_T(\boldsymbol{\theta}, \mathbf{0})$ denote the Gaussian pseudo-ML (PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ in which $\boldsymbol{\varrho}$ is set to zero. As we mentioned in the introduction, $\tilde{\boldsymbol{\theta}}_T$ remains root-T consistent for $\boldsymbol{\theta}_0$ under correct specification of $\mu_t(\boldsymbol{\theta})$ and $\sigma_t^2(\boldsymbol{\theta})$ even though the conditional distribution of $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is not Gaussian, provided that it has bounded fourth moments. Proposition 2 in Fiorentini and Sentana (2007) derives the asymptotic distribution of the pseudo-ML estimator of $\boldsymbol{\theta}$ when $\boldsymbol{\varepsilon}_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ is *i.i.d.*:

Proposition 6 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ *is i.i.d.* $D(0,1,\varrho_0)$ *with* $\kappa_0 < \infty$*, and the regularity conditions A.1 in Bollerslev and Wooldridge (1992) are satisfied, then* $\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \to N[\mathbf{0}, \mathcal{C}(\boldsymbol{\phi}_0)]$*, where*

$$\mathcal{C}(\boldsymbol{\phi}) = \mathcal{A}^{-1}(\boldsymbol{\phi})\mathcal{B}(\boldsymbol{\phi})\mathcal{A}^{-1}(\boldsymbol{\phi}),$$
$$\mathcal{A}(\boldsymbol{\phi}) = -E\left[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{A}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],$$
$$\mathcal{A}_t(\boldsymbol{\phi}) = -E[\mathbf{h}_{\boldsymbol{\theta\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\mathbf{0})\mathbf{Z}'_{dt}(\boldsymbol{\theta}),$$
$$\mathcal{B}(\boldsymbol{\phi}) = V\left[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}, \mathbf{0})|\boldsymbol{\phi}\right] = E\left[\mathcal{B}_t(\boldsymbol{\phi})|\boldsymbol{\phi}\right],$$
$$\mathcal{B}_t(\boldsymbol{\phi}) = V[\mathbf{s}_{\boldsymbol{\theta}t}(\boldsymbol{\theta}; \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{K}(\kappa)\mathbf{Z}'_{dt}(\boldsymbol{\theta}),$$
$$\text{and } \mathcal{K}(\varphi,\kappa) = V[\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}] = \begin{bmatrix} 1 & \varphi(\varrho) \\ \varphi(\boldsymbol{\varrho}) & \kappa(\boldsymbol{\varrho}) - 1 \end{bmatrix}, \tag{E24}$$

which only depends on $\boldsymbol{\varrho}$ *through the population coefficients of asymmetry and kurtosis*

$$\varphi(\boldsymbol{\varrho}) = E(\varepsilon_t^{*3}|\boldsymbol{\varrho}). \tag{E25}$$
$$\kappa(\boldsymbol{\varrho}) = E(\varepsilon_t^{*4}|\boldsymbol{\varrho}). \tag{E26}$$

Given that $\varphi(\boldsymbol{\varrho}) = 0$ and $\kappa = 2/(\nu - 4)$ for the Student t distribution with ν degrees of freedom, it trivially follows that in that case $\mathcal{B}_t(\boldsymbol{\phi})$ reduces to

$$\frac{1}{\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\mu_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\mu}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\nu-1}{2(\nu-4)}\frac{1}{\sigma_t^4(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}.$$

E.3 Semiparametric estimators of $\boldsymbol{\theta}$

González-Rivera and Drost (1999) obtain the semiparametric efficient score and the corresponding efficiency bound for univariate models:

Proposition 7 *If* $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}_0, \boldsymbol{\varrho}_0$ *is i.i.d.* $(1,0)$ *with density function* $f(\varepsilon_t^*; \boldsymbol{\varrho})$*, where* $\boldsymbol{\varrho}$ *are some shape parameters and* $\boldsymbol{\varrho} = \mathbf{0}$ *denotes normality, such that both its Fisher information matrix for location and scale*

$$\mathcal{M}_{dd}(\boldsymbol{\varrho}) = V\left[\mathbf{e}_{dt}(\boldsymbol{\theta}, \boldsymbol{\varrho})|\mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}\right]$$
$$= V\left\{\left[\begin{array}{c} e_{lt}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \\ e_{st}(\boldsymbol{\theta}, \boldsymbol{\varrho}) \end{array}\right]\middle|\boldsymbol{\theta}, \boldsymbol{\varrho}\right\} = V\left\{\left[\begin{array}{c} -\partial \ln f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial\varepsilon^* \\ -vec\left\{\mathbf{I}_N + \partial \ln f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]/\partial\varepsilon^* \cdot \varepsilon_t^*(\boldsymbol{\theta})\right\} \end{array}\right]\middle|\boldsymbol{\theta}, \boldsymbol{\varrho}\right\}$$

and the matrix of third and fourth order central moments

$$\mathcal{K}(\boldsymbol{\varrho}) = V[\mathbf{e}_{dt}(\boldsymbol{\theta}, \mathbf{0})|\, \mathbf{z}_t, I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\varrho}] \tag{E27}$$

are bounded, then the semiparametric efficient score will be given by:

$$\mathbf{Z}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^{-1}(\varphi, \kappa)\mathbf{e}_{dt}(\boldsymbol{\theta}_0, \mathbf{0})\right], \tag{E28}$$

while the semiparametric efficiency bound is

$$\mathcal{S}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0) - \mathbf{Z}_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0)\left[\mathcal{M}_{dd}(\boldsymbol{\varrho}_0) - \mathcal{K}(0)\,\mathcal{K}^{1}(\varphi, \kappa)\mathcal{K}(0)\right]\mathbf{Z}'_d(\boldsymbol{\theta}_0, \boldsymbol{\varrho}_0), \tag{E29}$$

where $+$ *denotes Moore-Penrose inverses, and* $\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}, \boldsymbol{\varrho}) = E\left[\mathbf{Z}_{dt}(\boldsymbol{\theta})\mathcal{M}_{dd}(\boldsymbol{\varrho})\mathbf{Z}'_{dt}(\boldsymbol{\theta})|\boldsymbol{\theta}, \boldsymbol{\varrho}\right].$

In practice, $f[\varepsilon_t^*(\boldsymbol{\theta}); \boldsymbol{\varrho}]$ has to be replaced by a non-parametric density estimator, which is typically obtained by kernel methods.

Hodgson and Vorkink (2001), Hafner and Rombouts (2007) and other authors have suggested semi-parametric estimators of $\boldsymbol{\theta}$ which limit the admissible distributions of $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}; \boldsymbol{\phi}_0$ to the class of symmetric ones. Proposition 7 in Fiorentini and Sentana (2007) provides the resulting elliptically symmetric semiparametric efficient score and the corresponding efficiency bound:

Proposition 8 *When $\varepsilon_t^*|\mathbf{z}_t, I_{t-1}, \boldsymbol{\phi}_0$ is i.i.d. $s(0{,}1, \boldsymbol{\varrho}_0)$ with $1 < \kappa_0 < \infty$, the elliptically symmetric semiparametric efficient score is given by:*

$$\begin{aligned}\mathring{\mathbf{s}}_{\theta t}(\boldsymbol{\phi}_0) &= \mathbf{Z}_{dt}(\boldsymbol{\theta}_0)\mathbf{e}_{dt}(\boldsymbol{\phi}_0)\\ &\quad -\mathbf{W}_s(\boldsymbol{\phi}_0)\left\{-\left[1+\varepsilon_t(\boldsymbol{\theta}_0)\partial \ln f[\varepsilon_t^*(\boldsymbol{\theta});\boldsymbol{\varrho}]/\partial\varepsilon^*\right] - \frac{2}{\kappa_0-1}\left[\varepsilon_t^2(\boldsymbol{\theta}_0)-1\right]\right\}, \quad \text{(E30)}\end{aligned}$$

where

$$\mathbf{W}_s(\boldsymbol{\phi}_0) = \mathbf{Z}_d(\boldsymbol{\phi}_0)\begin{pmatrix}0\\1\end{pmatrix} = E[\mathbf{Z}_{dt}(\boldsymbol{\theta}_0)|\boldsymbol{\phi}_0]\begin{pmatrix}0\\1\end{pmatrix} = E\left\{\frac{1}{2\sigma_t^2(\boldsymbol{\theta})}\frac{\partial\sigma_t^2(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\bigg|\boldsymbol{\phi}_0\right\}, \quad \text{(E31)}$$

while the elliptically symmetric semiparametric efficiency bound is

$$\mathring{\mathcal{S}}(\boldsymbol{\phi}_0) = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}_0) - \mathbf{W}_s(\boldsymbol{\phi}_0)\mathbf{W}_s'(\boldsymbol{\phi}_0)\cdot\left[\mathcal{M}_{ss}(\boldsymbol{\varrho}_0) - \frac{4}{\kappa_0-1}\right]. \quad \text{(E32)}$$

In practice, $e_{dt}(\boldsymbol{\phi})$ has to be replaced by a semiparametric estimate obtained from the density of ε_t^* that imposes symmetry. The simplest way to do this is by averaging the non-parametric density estimators at ε_t^* and $-\varepsilon_t^*$. Alternatively, one can estimate the common density of $\pm\varepsilon_t^*$ from the density of the Box-Cox transformation $k^{-1}|\varepsilon_t^*|^k - 1$ for some $k \geq 0$.

E.4 Student t-based (pseudo) maximum likelihood estimators

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta},\eta} L_T(\boldsymbol{\theta},\eta)$ denote the t-based pseudo-ML (t-PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ obtained by assuming that the conditional distribution is $t(0,1,\eta)$. Proposition 5 in Fiorentini and Sentana (2019) shows that this estimator is asymptotically equivalent to the Gaussian PML estimator when the conditional distribution is platykurtic. They also show that if the conditional mean and variance can be parametrised as in Linton (1993) and Newey and Steigerwald (1997), then some of the reparametrised mean and variance parameters will be consistently estimated even if the true conditional distribution is not a Student t. In our context, the robustness of the Student t serial correlation tests under conditional symmetry follows from the fact that the only parameter that is inconsistently estimated is ω in those circumstances. More generally, its robustness under possibly asymmetric distributions derives from the fact that we can reparametrise the mean of (1) as $\delta\sqrt{\omega} + \rho y_{t-1}$. Therefore, the t-based ML estimator of ρ continues to be consistent even if the estimators of ω and π are inconsistent. The argument for the α is slightly different, because a Student log-likelihood function can only estimate $\gamma = \alpha/\omega$ consistently in those circumstances. Nevertheless, given that α is 0 under the null, the t-based ML estimator of α continues to be consistent even if the estimators of ω and π are inconsistent.

E.5 Kotz-based (pseudo) maximum likelihood estimators

The original Kotz distribution (see Kotz (1975)) is a member of the spherical family, and thereby symmetric in the univariate case. Its main distinctive characteristic is that ε^{*2} follows a gamma distribution with mean 1 and variance $(3\varkappa_0 + 2)$, where

$$\varkappa = E(\varepsilon^{*4}|\boldsymbol{\eta})/3 - 1$$

is the coefficient of multivariate excess kurtosis of ε^* (see Mardia (1970)), which is trivially 0 under normality. In fact, the Kotz distribution nests the normal distribution when $\varkappa = 0$, in which ε^{*2} follows with a chi square distribution with one degree of freedom, but it can also be either platykurtic ($\varkappa < 0$) or leptokurtic ($\varkappa > 0$), although in the second case the Jensen inequality restriction $E(\varepsilon^{*4}) \geq E(\varepsilon^{*2}) = 1$ implies that $\varkappa \geq -2/3$. Such a nesting provides an analytically convenient generalisation of the normal. Specifically, the kernel of the distribution of ε^{*2} is

$$g(\varepsilon^{*2}; \varkappa) = -\frac{3\varkappa}{2(3\varkappa+2)} \ln \varepsilon^{*2} - \frac{1}{3\varkappa+2}\varepsilon^{*2},$$

while the constant of integration becomes

$$c(\varkappa) = -\ln \Gamma\left(\frac{1}{3\varkappa+2}\right) - \frac{1}{3\varkappa+2}\ln(3\varkappa+2)$$

(see Amengual and Sentana (2011)). Therefore, the density of a leptokurtic Kotz distribution has a pole at 0, and an antimode in the platykurtic case, which is a potential drawback from an empirical point of view.

The contribution of the t^{th} observation to the log-likelihood function is

$$l_t(\boldsymbol{\theta}, \varkappa) = -\frac{1}{2}\ln \sigma_t^2(\boldsymbol{\theta}) + c(\varkappa) + g(\varepsilon_t^{*2}; \varkappa).$$

As a result, the damping factor becomes

$$\delta(\varepsilon^{*2}; \varkappa) = \frac{1}{3\varkappa+2}\left(\frac{3\varkappa}{\varepsilon^{*2}} + 2\right).$$

Let $\tilde{\boldsymbol{\theta}}_T = \arg\max_{\varkappa} L_T(\boldsymbol{\theta}, \varkappa)$ denote the t-based pseudo-ML (t-PML) estimator of the conditional mean and variance parameters $\boldsymbol{\theta}$ obtained by assuming that the conditional distribution is a standardised version of the univariate $Kotz(0, 1, \varkappa)$.

Straightforward algebra shows that the ML estimator of the mean sets to 0 the following moment condition

$$\frac{1}{3\varkappa+2}[3\varkappa\check{\varepsilon}_T^{*-1}(\boldsymbol{\theta}) + 2\bar{\varepsilon}_T^{*}(\boldsymbol{\theta})] = 0,$$

where $\check{\varepsilon}_T^{*-1}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^T \varepsilon_t^{*-1}(\boldsymbol{\theta})$ is the reciprocal of the harmonic mean of the standardised residuals and $\bar{\varepsilon}_T^{*}(\boldsymbol{\theta})$ their arithmetic one. Therefore, the ML estimator makes a combination of the arithmetic and harmonic mean of the standardised residuals equal to 0. In contrast, the ML estimator of the variance can be concentrated out of the log-likelihood function as:

$$\omega(\pi) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \pi)^2$$

Finally, the score with respect to the excess kurtosis parameter $\varkappa$ is

$$s_{\varkappa t}(\boldsymbol{\theta}, \varkappa) = \varepsilon_t^{*2} - \ln \varepsilon_t^{*2} + \left[\psi\left(\frac{1}{3\varkappa+2}\right) + \ln(3\varkappa+2) - 1\right],$$

where $\psi\,(.)$ is the digamma (or Gauss psi) function (see Abramowitz and Stegun (1964)).

We can combine the moments of a gamma and reciprocal gamma random variables to show that

$$\mathrm{M}_{ll}(\varkappa) = \frac{9\varkappa + 2}{(3\varkappa + 1)(3\varkappa + 2)}, \tag{E33}$$

as long as $\varkappa > -1/3$,

$$\mathrm{M}_{ss}(\varkappa) = \frac{\varkappa + 2}{3\varkappa + 2},$$

and $\mathrm{M}_{sr}(\varkappa) = 0 \ \forall \varkappa$, as in the Gaussian case, so that the information matrix is block diagonal between the mean, variance and shape parameters.

To sample the Kotz innovations, we exploit the fact that $\varepsilon_t^* = \sqrt{\xi_t} u_t$, where u_t is a shifted and scaled Bernoulli random variable that the values ± 1 with probability 1/2 each, and ξ_t is a univariate Gamma with mean 1 and variance $(3\varkappa + 2)$.

Like in the Student t case, all mean parameters will be consistently estimated if the true conditional distribution is symmetric, while only ρ will remain consistent under asymmetry. And while ω will be inconsistently estimated unless the true distribution is Kotz, $\gamma = \alpha/\omega$ will be consistently estimated regardless.

E.6 Laplace-based (pseudo) maximum likelihood estimators

The Laplace (or double exponential) distribution, which is also a member of the generalised hyperbolic distribution, contains no shape parameters. As is well known, the ML estimator of the location parameter is given by the sample median, $med(y_1, \ldots, y_T)$. In turn, the estimator of the variance parameter ω is given by the twice the square of the mean absolute deviation around the median. Specifically,

$$\hat{\omega}_T = 2\left[\frac{1}{T}\sum_{t=1}^{T} |y_t - med(y_1, \ldots, y_T)|\right]^2.$$

Although the lack of shape parameters implies that the Laplace distribution is not very flexible, the fact that it is symmetric implies that the robustness properties of the pseudo ML estimators of ρ and γ are exactly the same as in the Student and Kotz-based log-likelihood functions.

E.7 Discrete mixtures of normals-based (pseudo) maximum likelihood estimators

The EM algorithm discussed by Dempster, Laird and Rubin (1977) allows us to obtain initial values as close to the optimum as desired. The recursions are as follows:

$$\hat{\lambda}^{(n} = \frac{1}{T}\sum\nolimits_{t=1}^{T} w(y_t; \boldsymbol{\phi}^{(n-1})$$

$$\hat{\mu}_1^{(n} = \frac{1}{\hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t w(y_t; \boldsymbol{\phi}^{(n-1)}), \quad \hat{\mu}_2^{(n} = \frac{1}{1-\hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t [1-w(y_t; \boldsymbol{\phi}^{(n-1)})],$$

$$\hat{\sigma}_1^{2(n} = \frac{1}{\hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t^2 w(y_t; \boldsymbol{\phi}^{(n-1)}) - \left(\hat{\mu}_1^{(n)}\right)^2, \quad \hat{\sigma}_2^{2(n} = \frac{1}{1-\hat{\lambda}^{(n}} \frac{1}{T} \sum_{t=1}^{T} y_t^2 [1-w(y_t; \boldsymbol{\phi}^{(n-1)})] - \left(\hat{\mu}_2^{(n)}\right)^2,$$

where

$$\begin{aligned} w(y_t; \boldsymbol{\phi}) &= \frac{\frac{\lambda}{\sigma_1} \phi\left(\frac{y_t - \mu_1}{\sigma_1}\right)}{\frac{\lambda}{\sigma_1} \phi\left(\frac{y_t - \mu_1}{\sigma_1}\right) + \frac{1-\lambda}{\sigma_2} \phi\left(\frac{y_t - \mu_2}{\sigma_2}\right)} \\ &= \frac{\frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})} \phi\left[\frac{\varepsilon_t^*(\boldsymbol{\theta}_s, 0) - \mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right]}{\frac{\lambda}{\sigma_1^*(\boldsymbol{\eta})} \phi\left[\frac{\varepsilon_t^*(\boldsymbol{\theta}_s, 0) - \mu_1^*(\boldsymbol{\eta})}{\sigma_1^*(\boldsymbol{\eta})}\right] + \frac{1-\lambda}{\sigma_2^*(\boldsymbol{\eta})} \phi\left[\frac{\varepsilon_t^*(\boldsymbol{\theta}_s, 0) - \mu_2^*(\boldsymbol{\eta})}{\sigma_2^*(\boldsymbol{\eta})}\right]} = w[\epsilon_t^*(\boldsymbol{\theta}_s, 0); \boldsymbol{\eta}] \end{aligned}$$

and $\phi(.)$ denotes the standard normal density.

From those recursions it is easy to check that

$$\begin{aligned} \hat{\pi}^{(n} &= \hat{\mu}_1^{(n} \hat{\lambda}^{(n} + \hat{\mu}_2^{(n} (1 - \hat{\lambda}^{(n)}) = \frac{1}{T} \sum\nolimits_{t=1}^{T} y_t, \\ \hat{\sigma}^{2(n} &= [(\hat{\mu}_1^{(n)})^2 + \hat{\sigma}_1^{2(n)}] \hat{\lambda}^{(n} + [(\hat{\mu}_2^{(n)})^2 + \hat{\sigma}_2^{2(n)}](1 - \hat{\lambda}^{(n)}) - (\hat{\pi}^{(n)})^2 \\ &= \frac{1}{T} \sum\nolimits_{t=1}^{T} y_t^2 - \left(\frac{1}{T} \sum\nolimits_{t=1}^{T} y_t\right)^2, \end{aligned}$$

for all n regardless of the values of $\boldsymbol{\phi}^{(n-1}$. This means that $\hat{\lambda}^{(n}$, $\hat{\upsilon}^{(n} = \hat{\sigma}_2^{2(n} / \hat{\sigma}_1^{2(n}$ and

$$\hat{\delta}^{(n} = \frac{\hat{\mu}_1^{(n} - \hat{\mu}_2^{(n}}{\sqrt{\hat{\lambda}^{(n} \hat{\sigma}_1^{2(n} + (1 - \hat{\lambda}^{(n)}) \hat{\sigma}_2^{2(n}}}$$

will yield the EM recursions for a mixture model parametrised in terms of π, ω^2 and λ, δ and υ, which are the parameters of the standardised version in Supplemental Appendix D.1.

Since the ML estimators constitute the fixed point of the EM recursions, (i.e. $\boldsymbol{\phi} = \boldsymbol{\phi}^{(\infty)}$), another implication of the above result is that $\hat{\pi}$ and $\hat{\omega}$ coincide with the Gaussian PML estimators. As a result, we can maximise the log-likelihood function with respect to λ, δ and υ keeping $\hat{\pi}$ and $\hat{\sigma}^2$ fixed at their Gaussian pseudo ML values. Interestingly, this somewhat surprising result will continue to be true even in a complete log-likelihood situation in which we would observe not only y_t but also s_t. In addition, it is straightforward to prove that the same result holds for finite mixtures of normals with more than two components.

As a result, the ML estimators of π and ω continue to be consistent under distributional misspecification. Similarly, the estimators of ρ and $\alpha = \omega\gamma$ will also remain consistent in that case too, as explained in Fiorentini and Sentana (2019).

Nevertheless, the log-likelihood function of a mixture distribution has a pole for each observation. Specifically, it will go to infinity if we set $\hat{\mu}_1 = y_t$ and let $\hat{\sigma}_1^2$ go to 0. In practice, we deal with this issue by starting the EM algorithm from many different starting values. In addition, there is a trivial identification issue that arises by exchanging the labels of the components. We solve this problem by restricting υ to the range $(0, 1)$ so that the first component is the one with the largest variance.

Additional references

Abramowitz, M. and Stegun, I.A. (1964): *Handbook of mathematic functions*, AMS 55, National Bureau of Standards.

Campbell, J.Y. and Yogo, M. (2006): "Efficient tests of stock return predictability", *Journal of Financial Economics* 81, 27-60.

Cochrane, J.H. (1991): "Volatility tests and efficient markets: a review essay", *Journal of Monetary Economics* 27, 661-676.

Crowder, M.J. (1976): "Maximum likelihood estimation for dependent observations", *Journal of the Royal Statistical Society* B, 38, 45-53.

Davidson R., and MacKinnon, J.G. (1988): "Double-length artificial regressions", *Oxford Bulletin of Economics and Statistics* 50, 203–217.

Dempster, A., Laird, N., and Rubin, D. (1977): "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society B* 39, 1-38.

Fiorentini, G. and Sentana, E. (1998): "Conditional means of time series processes and time series processes for conditional means", *International Economic Review* 39, 1101-1118.

González-Rivera, G. and Drost, F.C. (1999): "Efficiency comparisons of maximum-likelihood-based estimators in GARCH models", *Journal of Econometrics* 93, 93-111.

Hodrick, R. J. (1992): "Dividend yields and stock returns: alternative procedures for inference and measurement", *Review of Financial Studies* 5, 357-386.

Jegadeesh, N. (1989): "On testing for slow decaying components in stock prices", mimeo, Anderson Graduate School of Management, University of California at Los Angeles.

Kotz, S. (1975). "Multivariate distributions at a cross-road", in G. P. Patil, S. Kotz and J.K. Ord (eds.) *Statistical distributions in scientific work*, vol. I, 247-270, Reidel.

Mardia, K. (1970). "Measures of multivariate skewness and kurtosis with applications", *Biometrika*, 57, 519–530.

Mencía, J. and E. Sentana (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, 105-121.

Mencía, J. and E. Sentana (2012): "Distributional tests in multivariate dynamic models with Normal and Student t innovations", *Review of Economics and Statistics* 94, 133-152.

Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H. (2012): "Time series momentum", *Journal of Financial Economics* 104, 228–250.

Newey, W.K. and Steigerwald, D.G. (1997): "Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models", *Econometrica* 65, 587-99.

FIGURE S1: ACF of expected and observed returns ($h = 24, \rho = .015$)

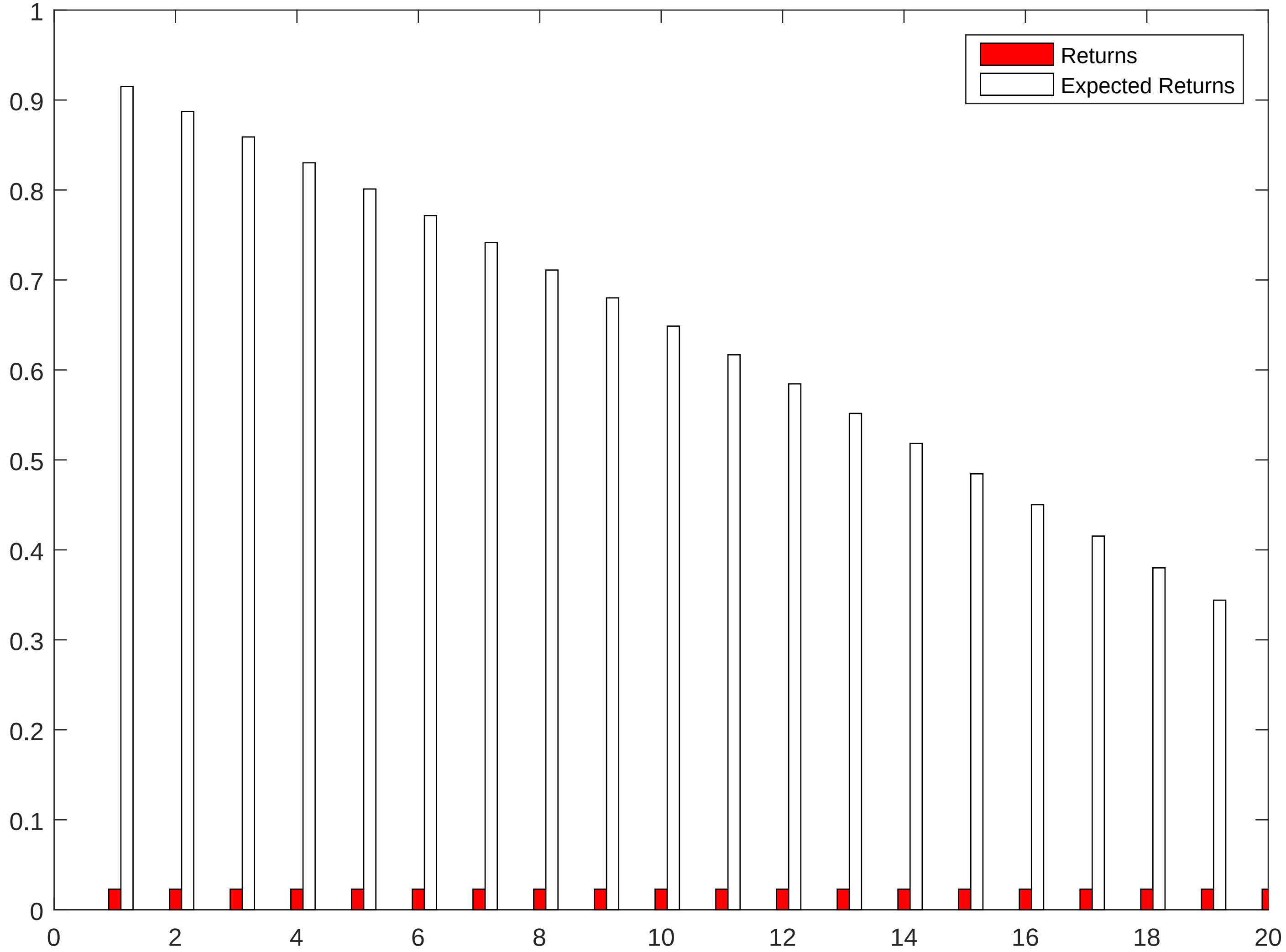
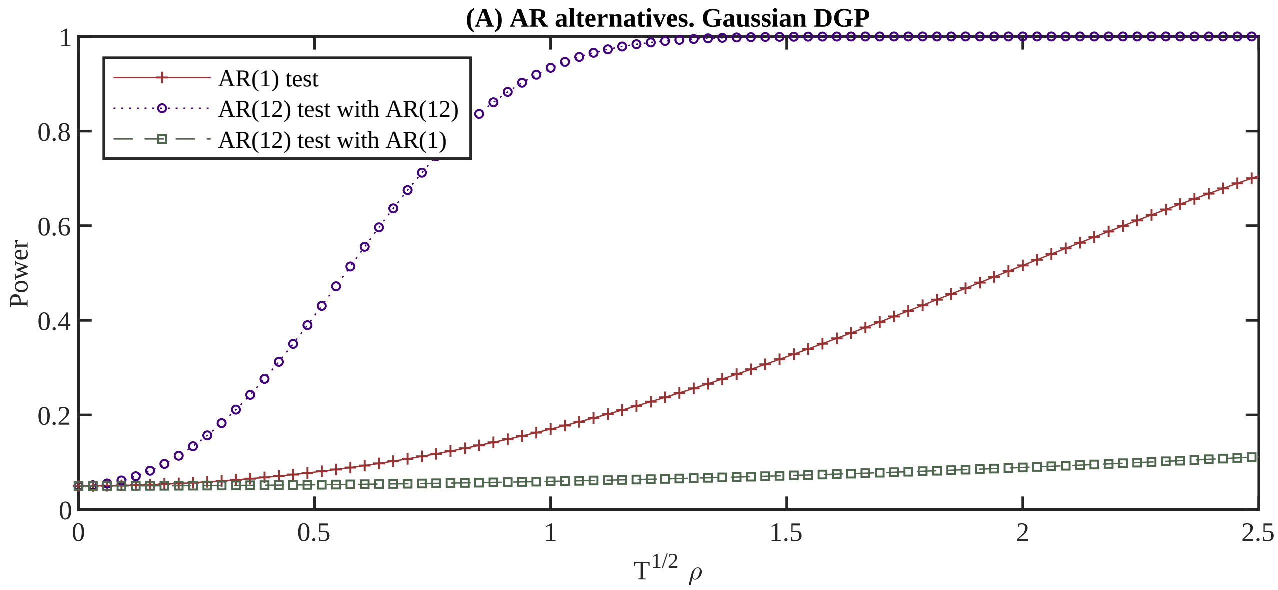
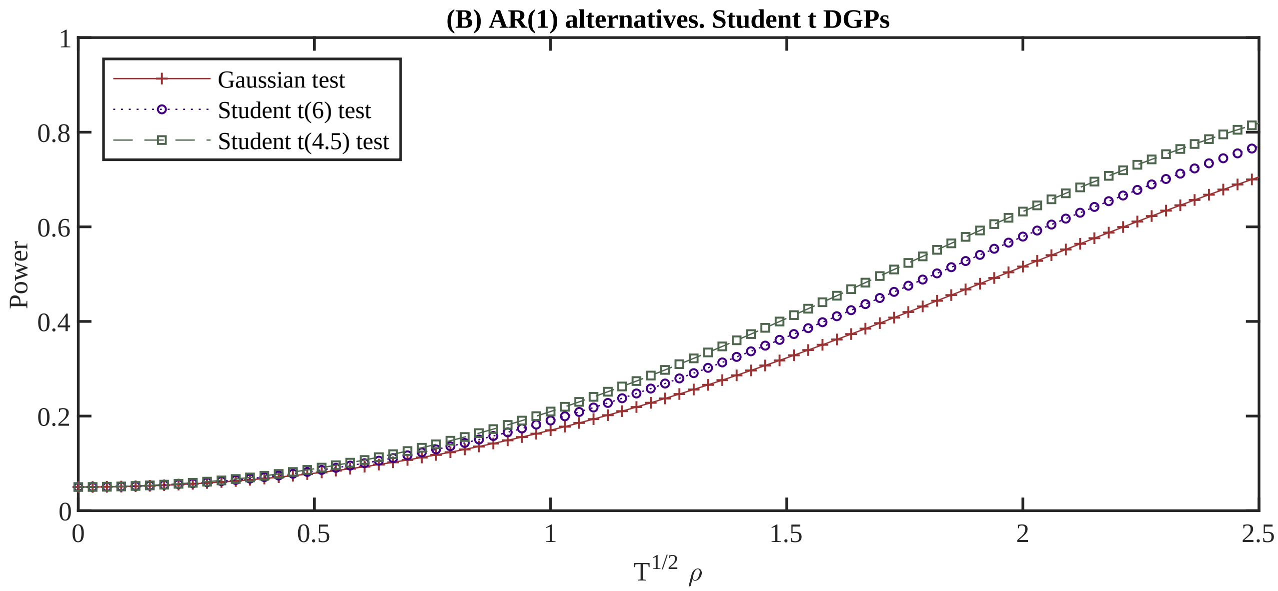


FIGURE S2: Local power of unpredictability in mean tests at 5% level





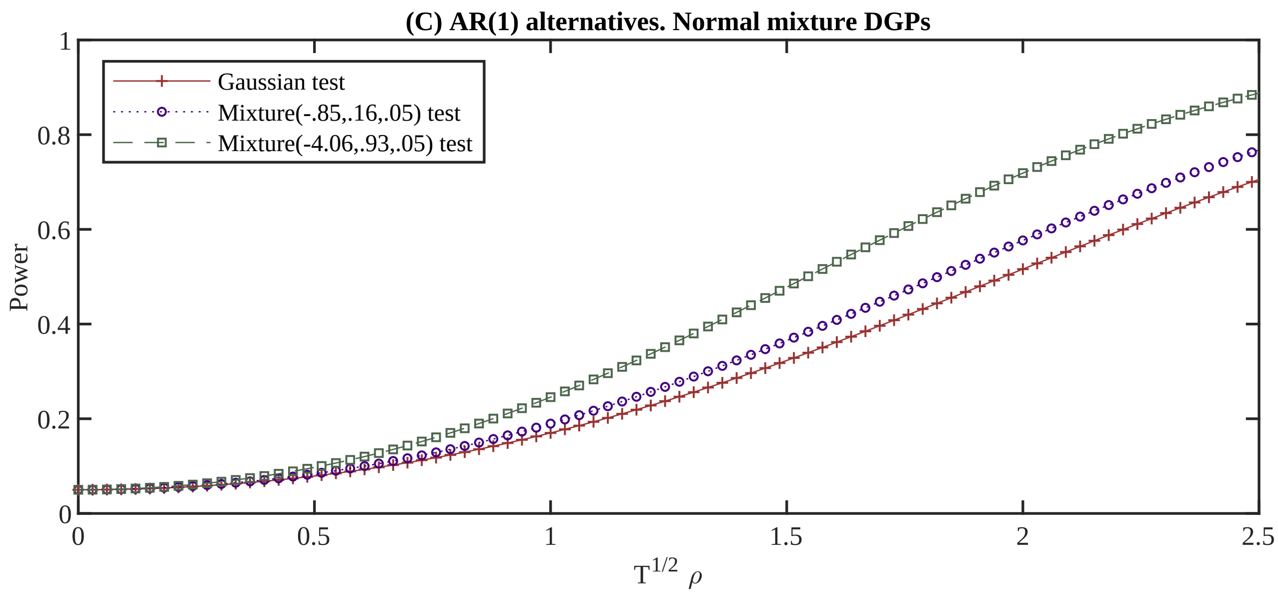


FIGURE S3: Local power of unpredictability in variance tests at 5% level