Supplemental Appendices for

Hypothesis tests with a repeatedly singular information matrix

Dante Amengual
CEMFI
<amengual@cemfi.es>

Xinyue Bei
University of Texas at Austin
<xinyue.bei@austin.utexas.edu>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

Revised: March 2025

B Proof of Lemmas and Propositions

B.1 Lemmata

In lemmas 1–5 we define $LM_n(\boldsymbol{\rho})$ as in (A2) with $\mathcal{S}_n$ and $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$ defined in Assumption 5.

Lemma 1 *If Assumptions 1 and 5.1, 5.2, 5.3 hold, then*

$$(i)\ \boldsymbol{\rho}^{LM} \xrightarrow{p} 0, \quad and \quad (ii)\ n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM}) = O_p(1).$$

Proof. Let us start with Lemma 1.(ii). If we fix $\epsilon > 0$, then by Assumption 5.2, we have that $n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*) = O_p(1)$, which in turn means that there exists an M_1 such that for all $n \geq N$,

$$\Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1) \leq \epsilon. \tag{B1}$$

Next, let $M = (2M_1+1)/e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]$, which is a positive real number because of Assumption 5.3. We can then prove that

$$\Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\}) = 0. \tag{B2}$$

To see (B2), note that if $||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M$ and $||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1$, then we have that

$$\begin{aligned}
&2(n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*))'[n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)] - [n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)]\\
&\leq 2||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \cdot ||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| - e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]||n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM})||^2\\
&\leq ||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| \cdot [2M_1 - e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)||]\\
&< -M = LM_n(\boldsymbol{\phi}^*,\mathbf{0}) - M,
\end{aligned}$$

where the first two inequalities are straightforward, the third one follows from $||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1$ and $||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M = (2M_1+1)/e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]$, while the last one follows from $LM_n(\boldsymbol{\phi}^*,\mathbf{0}) = 0$, which contradicts $\boldsymbol{\rho}^{LM}$ being the minimizer. Thus (B2) holds. As a consequence,

$$\begin{aligned}
\Pr(||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M) =& \Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\})\\
&+ \Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1\}) \qquad \text{(B3)}\\
\leq& \Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1) \leq \epsilon, \qquad \text{(B4)}
\end{aligned}$$

where from (B3) to (B4) we have used (B1) and (B2). Therefore, (B4) trivially implies that for all $\varepsilon > 0$ there exists $M > 0$ such that $\Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \leq \varepsilon$ and, hence, Lemma 1.(i) holds.

As for Lemma 1.(i), for all $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that

$$\Pr\left(\left\|\boldsymbol{\rho}^{LM} - (\boldsymbol{\phi}^*,\mathbf{0})\right\| \geq \epsilon\right) \leq \Pr(||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| \geq n^{\frac{1}{2}}\delta_\epsilon) \to 0,$$

where the inequality follows from Assumption 5.1 and the convergence from Lemma 1.(ii). □

Lemma 2 *If Assumptions 1 and 5.1–4 hold, then* $n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}) = O_p(1)$.

Proof. Assumption 1 implies the consistency of $\hat{\boldsymbol{\rho}}$, while Assumption 5.4 implies that

$$\frac{R_n(\hat{\boldsymbol{\rho}})}{1+n\left\|\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)\right\|^2} = o_p(1).$$

Therefore, for any fixed $\epsilon > 0$ there exists an N such that for all $n > N$,

$$\Pr\left(A_n\right) \geq 1 - \frac{\epsilon}{2}, \tag{B5}$$

with

$$A_n = \left\{\left|\frac{R_n(\hat{\boldsymbol{\rho}})}{1+n\left\|\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)\right\|^2}\right| \leq \frac{1}{6}e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]\right\}.$$

In turn, given that $n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)$ is $O_p(1)$, there exists an M_1 such that for all n,

$$\Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \geq M_1) < \frac{\epsilon}{2}, \tag{B6}$$

so that, letting $M = \max\{(6M_1+3)/e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)], 1\}$, we can then show that

$$\Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| \geq M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\} \cap A_n) = 0. \tag{B7}$$

To prove this by contradiction, suppose we have

$$||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| \geq M,\ \ ||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1 \ \text{ and } \ \left|\frac{R_n(\hat{\boldsymbol{\rho}})}{1+n\left\|\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)\right\|^2}\right| \leq \frac{1}{6}e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]. \tag{B8}$$

Then

$$\begin{aligned}
LR(\hat{\boldsymbol{\rho}}) &= 2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)] - [n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)] + 2R_n\left(\hat{\boldsymbol{\rho}}\right) \\
&\leq 2M_1||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| - e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]n||\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||^2 + \frac{e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]}{3}(1+n||\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||^2) \\
&= ||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||\left\{2M_1 - e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| + \frac{e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]}{3}\left(\frac{1}{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||} + ||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||\right)\right\} \\
&\leq ||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||\left\{2M_1 - e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| + \frac{2e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]}{3}||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||\right\} \\
&= ||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||\left\{2M_1 - \frac{e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]}{3}||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)||\right\} \leq -M = LR(\boldsymbol{\phi}^*, \mathbf{0}) - M,
\end{aligned}$$

which is inconsistent with the definition of $LR(\hat{\boldsymbol{\rho}})$. The first equality follows from Assumption 5, the first inequality from (B8), the next three lines are straightforward, the subsequent inequality follows from $||n^{\frac{1}{2}}\lambda\left(\hat{\boldsymbol{\rho}}\right)|| \geq M \geq (6M_1+3)/e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]$, and the last equality from $LR(\boldsymbol{\phi}^*, \mathbf{0}) = 0$. Therefore,

$$\begin{aligned}
\Pr(||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| \geq M) \leq &\Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| \geq M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\} \cap A_n) \\
&+ \Pr\left(A_n^c\right) + \Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1) \leq \epsilon
\end{aligned}$$

for all $n > N$, where the inequalities follow from (B5), (B6) and (B7). □

Lemma 3 *If Assumptions 1 and 5.1–4 hold, then* $LR_n(\hat{\boldsymbol{\rho}}) = LM_n(\boldsymbol{\rho}^{LM}) + o_p(1)$.

Proof. Given that $\max\{n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}), n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM})\} = O_p(1)$, for all $\epsilon > 0$, there exists an M such that for all n,

$$\Pr(\max\{n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}), n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM})\} \le M) > 1 - \frac{\epsilon}{2}. \tag{B9}$$

Letting $P_n = \{\boldsymbol{\rho} \in \mathrm{P} : n^{\frac{1}{2}} \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \le M\}$, we can use Assumption 5.1 to choose a sequence of $\gamma_n \to 0$ satisfying

$$\inf_{\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\| \ge \gamma_n} \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| > \frac{M}{\sqrt{n}},$$

which in turn implies that $P_n \subset \{\boldsymbol{\rho} \in \mathrm{P} : \|\boldsymbol{\rho} - (\boldsymbol{\phi}^*, \mathbf{0})\| \le \gamma_n\}$. But then,

$$\begin{aligned}
\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| = 2 \sup_{\boldsymbol{\rho}\in P_n} |R_n(\boldsymbol{\rho})| & \\
\le 2(1+M)^2 \sup_{\boldsymbol{\rho}\in\mathbf{P}:\|\boldsymbol{\lambda}(\boldsymbol{\rho})\| \le \frac{M}{\sqrt{n}}} \frac{|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})|}{1 + n\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\|^2} & \\
\le 2(1+M)^2 \sup_{\boldsymbol{\rho}\in\mathbf{P}:\|\boldsymbol{\rho}-(\boldsymbol{\phi}^*,\mathbf{0})\| \le \gamma_n} \frac{|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})|}{1 + n\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\|^2} &= o_p(1),
\end{aligned}$$

where the first line follows from Assumption 5, the second one from the definition of P_n, the third one from $A_n = \{\boldsymbol{\rho} \in \mathrm{P}: n^{\frac{1}{2}} \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \le M\} \subset \{\boldsymbol{\rho} \in \mathrm{P} : \|\boldsymbol{\rho} - (\boldsymbol{\phi}^*, \mathbf{0})\| \le \gamma_n\}$, and the last equality from $\gamma_n \to 0$ and Assumption 5.4. Therefore, there exists an N such that for all $n > N$,

$$\Pr\left(\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon\right) > 1 - \frac{\epsilon}{2}. \tag{B10}$$

As a consequence, we will have that for $n > N$,

$$\begin{aligned}
&\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < \epsilon\right) \\
\ge &\Pr\left(\left\{\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < \epsilon\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) \qquad \text{(B11)} \\
\ge &\Pr\left(\left\{\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) \qquad \text{(B12)} \\
\ge &\Pr\left(\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon\right) + \Pr\left(\{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) - 1 \qquad \text{(B13)} \\
\ge &1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 = 1 - \epsilon, \qquad \text{(B14)}
\end{aligned}$$

where to go from (B11) to (B12) we have used

$$\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| \ge \left|\sup_{\boldsymbol{\rho}\in P_n} LR_n(\boldsymbol{\rho}) - \sup_{\boldsymbol{\rho}\in P_n} LM_n(\boldsymbol{\rho})\right|,$$

from (B12) to (B13) the fact that $\Pr(E_1 \cap E_2) \ge \Pr(E_1) + \Pr(E_2) - 1$, while from (B13) to (B14) we relied on (B9) and (B10). Therefore, for all $\epsilon > 0$, there exists an N such that for all $n > N$,

$$\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < \epsilon\right) > 1 - \epsilon.$$

as desired. □

Lemma 4 *If Assumptions 1, 5.1–4 and 5.7 hold, then* $LR_n(\hat{\rho}) = LM_n(\rho^{LM}) + O_p(n^{-r})$.

Proof. We want to show that for all $\epsilon > 0$ there exists a constant K_ϵ such that for all n,

$$\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| \leq K_\epsilon n^{-r}\right) \geq 1-\epsilon.$$

The proof is almost analogous to the one of Lemma 3. Letting M and P_n be as the ones in that lemma, by Assumption 5.6,

$$\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| = 2\sup_{\boldsymbol{\rho}\in P_n} |R_n(\boldsymbol{\rho})| = O_p(n^{-r}),$$

which is equivalent to saying that there exists an K_ϵ such that for all n,

$$\Pr\left(\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-r}\right) > 1-\frac{\epsilon}{2}. \tag{B15}$$

Thus,

$$\begin{aligned}
&\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < K_\epsilon n^{-r}\right)\\
&\geq \Pr\left(\left\{\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < K_\epsilon n^{-r}\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right)\\
&\geq \Pr\left(\left\{\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-r}\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) && \text{(B16)}\\
&\geq \Pr\left(\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-r}\right) + \Pr\left(\{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) - 1 && \text{(B17)}\\
&\geq 1-\frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 = 1-\epsilon, && \text{(B18)}
\end{aligned}$$

where the last inequality follows from (B9) and (B15). □

Lemma 5 *If Assumptions 1 and 5.1–4 hold, then* $LR_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) = \sup_{(\boldsymbol{\phi},\mathbf{0})\in \boldsymbol{P}} LM_n(\boldsymbol{\phi}, \mathbf{0}) + o_p(1)$. *Moreover, if Assumption 5.7 also holds, then* $LR_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) = \sup_{(\boldsymbol{\phi},\mathbf{0})\in \boldsymbol{P}} LM_n(\boldsymbol{\phi}, \mathbf{0}) + O_p(n^{-r})$.

Proof. The proof is omitted because it is entirely analogous to the proofs of Lemmas 3 and 4 after fixing $\boldsymbol{\theta} = \mathbf{0}$ and changing P to $\{\boldsymbol{\phi} : (\boldsymbol{\phi}, \mathbf{0}) \in$P$\}$. □

Although the following result holds for any proper ordering over $\mathbb{N}^q$, we follow Constantine and Savits (1996) in saying that $\boldsymbol{k}_a \prec \boldsymbol{k}_b$ if at least one the following three conditions hold: (1) $|\boldsymbol{k}_a| < |\boldsymbol{k}_b|$; (2) $|\boldsymbol{k}_a| = |\boldsymbol{k}_b|$ but $k_{a1} < k_{b1}$; (3) $|\boldsymbol{k}_a| = |\boldsymbol{k}_b|$, $k_{a1} = k_{b1}, \ldots k_{aj} = k_{bj}$ but $k_{aj+1} < k_{bj+1}$ for some $1 < j < q$.

Lemma 6 *(Multivariate Faà di Bruno's formula) The arbitrary partial derivative of the composition of functions*

$$l(x_1, ..., x_d) = \log[f(x_1, ..., x_d)]$$

is given by

$$l^{[\boldsymbol{v}]} = \sum_{1 \le h \le \boldsymbol{\iota}_d' \boldsymbol{v}} (-1)^{h+1} \sum_{\substack{s=1:h \\ p_s(\boldsymbol{v},h)}} \prod_{a=1}^{s} \frac{1}{m_a!} \left(\frac{f^{[\boldsymbol{k}_a]}}{f} \right)^{m_a}, \quad where$$

$$p_s(\boldsymbol{v},h) = \left\{ (m_1,...,m_s;\boldsymbol{k}_1,...,\boldsymbol{k}_s) : m_a > 0, \mathbf{0} \prec \boldsymbol{k}_1 \prec ... \prec \boldsymbol{k}_s, \sum_{a=1}^{s} m_a = h \text{ and } \sum_{a=1}^{s} m_a \boldsymbol{k}_a = \boldsymbol{v} \right\}. \quad \text{(B19)}$$

Proof. : See p. 505 of Constantine and Savits (1996).

Proof of Proposition 1

Let

$$\varepsilon_k = u_k - \mathbf{r}_{(k)}(\boldsymbol{\phi}^L)\mathbf{u}_{(k)},$$

where $u_k(\boldsymbol{\varphi}_k^M, \boldsymbol{\varphi}_k^D) = (y_k - \boldsymbol{\varphi}_k^{M\prime}\mathbf{x})/\varphi_k^D$ and $r_{(k)}(\boldsymbol{\phi}^L)$ denotes the coefficients in the theoretical least squares projection of u_k on to (the linear span of) $\mathbf{u}_{(k)} = (u_1, \ldots, u_{k-1}, u_{k+1}, \ldots, u_K)'$. Straightforward calculations allow us to show that

$$\frac{\partial l}{\partial \phi^S} = \mathbf{w}\phi_N(\boldsymbol{\phi}^{S\prime}\mathbf{w}) \left[\frac{d}{\Phi_N(\boldsymbol{\phi}^{S\prime}\mathbf{w})} - \frac{1-d}{\Phi_N(-\boldsymbol{\phi}^{S\prime}\mathbf{w})} \right]$$

$$\frac{\partial l}{\partial \phi_k^M} = \frac{\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]}{\phi_k^D \det\left[\mathbf{R}(\boldsymbol{\phi}^L)\right]} dx\varepsilon_k$$

$$\frac{\partial l}{\partial \phi_k^D} = d\left[a_{kk}\left(u_k^2 - 1\right) + \sum_{h \ne k} a_{kh}\left(u_k u_h - \varphi_{kh}^L\right) \right]$$

$$\frac{\partial l}{\partial \phi_{kj}^L} = d\left[\sum_{h} b_{kj,h}\left(u_h^2 - 1\right) + \sum_{h \ne i} b_{kj,ih}\left(u_i u_h - \varphi_{ih}^L\right) \right]$$

$$\frac{\partial l}{\partial \theta_k} = \frac{\partial^2 l}{\partial \theta_k \partial \theta_j} = 0$$

$$\frac{\partial^3 l}{\partial \theta_k^3} = Cd \det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]^3 \varepsilon_k^3 + A_k \frac{\partial l}{\partial \phi}$$

$$\frac{\partial^3 l}{\partial \theta_k^2 \partial \theta_j} = Cd \det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]^2 \det\left[\mathbf{R}_{(j)}(\boldsymbol{\phi}^L)\right] \varepsilon_k^2 \varepsilon_j + A_{kj} \frac{\partial l}{\partial \phi} \quad \text{and}$$

$$\frac{\partial^3 l}{\partial \theta_k \partial \theta_j \partial \theta_h} = Cd \det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right] \det\left[\mathbf{R}_{(j)}(\boldsymbol{\phi}^L)\right] \det\left[\mathbf{R}_{(h)}(\boldsymbol{\phi}^L)\right] \varepsilon_k \varepsilon_j \varepsilon_h + A_{kjh} \frac{\partial l}{\partial \phi},$$

where $\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)$ the $(K-1) \times (K-1)$ matrix obtained from $\mathbf{R}(\boldsymbol{\phi}^L)$ after eliminating its k^{th} row and column,

$$C = \frac{1}{\det[\mathbf{R}(\boldsymbol{\phi}^L)]^3} \, \mathrm{d}^2 \left[\frac{\phi_N(x)}{\Phi_N(x)} \right] \Bigg|_{x = \phi^S \varepsilon},$$

and a_{kh}, $b_{kj,ih}$, A_k, A_{kj}, A_{kjh} for $k, j, h = 1, ..., K$ are some terms whose detailed expressions (available on request) are irrelevant for the proof.

Thus, we have that the test depends on the influence function

$$\sum_k \frac{1}{6}\upsilon_k^{\dagger 3}\frac{\partial^3 l}{\partial \theta_k^3}+\sum_{j\neq k}\frac{1}{2}\upsilon_k^{\dagger 2}\upsilon_j^{\dagger}\frac{\partial^3 l}{\partial \theta_k^2\partial\theta_j}+\sum_{h\neq j\neq k}\upsilon_k^{\dagger}\upsilon_h^{\dagger}\upsilon_j^{\dagger}\frac{\partial^3 l}{\partial\theta_k\partial\theta_j\partial\theta_h}$$
$$\propto\left\{\sum_k d\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]w_k\upsilon_k^{\dagger}\right\}^3+A^{\dagger}\frac{\partial l}{\partial\boldsymbol{\phi}}\propto dH_3\left(\frac{\boldsymbol{\upsilon}'\boldsymbol{\varepsilon}}{\sqrt{\boldsymbol{\upsilon}\boldsymbol{\upsilon}}}\right)+A\frac{\partial l}{\partial\boldsymbol{\phi}},$$

so that, by suitably choosing υ in the last expression,

$$\sum_k d\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]\varepsilon_k\upsilon_k^{\dagger}\propto d\boldsymbol{\upsilon}'\boldsymbol{\varepsilon},$$

which allows to show that the test has form in (13). □

Proof of Proposition 2

For those observations with $d=1$, we can write

$$\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\left(\boldsymbol{\varphi}^D\right)^{-1}\left(\mathbf{y}-\boldsymbol{\varphi}^M\mathbf{x}\right)=\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}u_S+\mathbf{z}^{\dagger}$$

where $\mathbf{z}^{\dagger}\sim N\left(\mathbf{0},\mathbf{I}_K\right)$ by construction. Given that the test is based on the standardized residuals, the statistics which use either $\mathbf{y}$ or

$$\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\left(\boldsymbol{\varphi}^D\right)^{-1}\left(\mathbf{y}-\boldsymbol{\varphi}^M\mathbf{x}\right)$$

as inputs are numerically the same. Therefore, for any $\boldsymbol{\upsilon}$ we will have that

$$\begin{aligned}\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\left(\boldsymbol{\varphi}^D\right)^{-1}\left(\mathbf{y}-\boldsymbol{\varphi}^M\mathbf{x}\right)&=\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}u_S+\boldsymbol{\upsilon}^{\dagger\prime}\mathbf{z}^{\dagger}\\&\propto u_S+\frac{1}{\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}}\boldsymbol{\upsilon}^{\dagger\prime}\mathbf{z}^{\dagger},\end{aligned}$$

which implies that the distribution of the test statistic conditional on $\mathbf{x}$ and $\mathbf{w}$ is determined by the unconditional distribution of

$$\left\{\left[\frac{\boldsymbol{\upsilon}^{\dagger\prime}}{\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}}\mathbf{z}^{\dagger}\right]_{\boldsymbol{\upsilon}^{\dagger}\neq\mathbf{0}},u_S\right\}. \tag{B20}$$

Next, let

$$\boldsymbol{\ell}=\frac{\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}}{\sqrt{\boldsymbol{\vartheta}'\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1}\boldsymbol{\vartheta}}}\quad\text{and}\quad\nu=\sqrt{\boldsymbol{\vartheta}'\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1}\boldsymbol{\vartheta}},$$

with $\boldsymbol{\ell}'\boldsymbol{\ell}=1$, which means that $\mathbf{I}-\boldsymbol{\ell}\boldsymbol{\ell}'$ has rank $K-1$, so that the singular value decomposition implies the existence of a $(K-1)\times K$ matrix $\mathbf{A}$ with full row rank such that $\mathbf{A}'\mathbf{A}=\mathbf{I}-\boldsymbol{\ell}'\boldsymbol{\ell}$. Defining $\boldsymbol{\upsilon}'=\boldsymbol{\upsilon}^{\dagger\prime}[(\ \boldsymbol{\ell}\quad\mathbf{A}\ )^{-1}$, we then have that

$$\frac{1}{\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^L\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}}\boldsymbol{\upsilon}^{\dagger\prime}\mathbf{z}^{\dagger}=\frac{\boldsymbol{\upsilon}^{\dagger\prime}\begin{bmatrix}\boldsymbol{\ell}'\\\mathbf{A}\end{bmatrix}^{-1}}{\boldsymbol{\upsilon}^{\dagger\prime}\begin{bmatrix}\boldsymbol{\ell}'\\\mathbf{A}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{\ell}'\\\mathbf{A}\end{bmatrix}\boldsymbol{\ell}\nu}\begin{bmatrix}\boldsymbol{\ell}'\\\mathbf{A}\end{bmatrix}\mathbf{z}^{\dagger}=\frac{\boldsymbol{\upsilon}'}{\boldsymbol{\upsilon}'\mathbf{e}_1\nu}\mathbf{z},$$

which in turn implies that

$$\left\{\left[\frac{\boldsymbol{\upsilon}^{\dagger\prime}}{\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^{L}\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}}\mathbf{z}^{\dagger}\right]_{\boldsymbol{\upsilon}^{\dagger}\neq\mathbf{0}}, u_S\right\} \sim \left\{\left[\frac{\boldsymbol{\upsilon}'}{\boldsymbol{\upsilon}'\mathbf{e}_1\nu}\mathbf{z}\right]_{\boldsymbol{\upsilon}\neq\mathbf{0}}, u_S\right\},$$

where

$$\mathbf{z}=\begin{bmatrix}\boldsymbol{\ell}'\\ \mathbf{A}\end{bmatrix}\mathbf{z}^{\dagger}, \mathbf{z}|\mathbf{x},\mathbf{w}\sim N(\mathbf{0},\mathbf{I}_K),$$

which confirms that the power will depend on ν exclusively. Finally, the Woodbury formula implies that we can rewrite ν as

$$\begin{aligned}\boldsymbol{\vartheta}'\left[\mathbf{R}\left(\boldsymbol{\phi}^{L}\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1}\boldsymbol{\vartheta} &= \boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}+\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}\left[1-\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}\right]\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}\\ &= \frac{\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}}{1-\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}},\end{aligned}$$

which confirms the exclusive role played by $\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\phi}^{L}\right)\boldsymbol{\vartheta}$. □

Proof of Proposition 3

Let $\tilde{u}_i=(y_i-\tilde{\phi}^{L})/\sqrt{\tilde{\phi}^{V}}$ and $\tilde{\mathrm{H}}_k=\sum_{i=1}^{n}H_k(\tilde{u}_i)$. Then it is easy to see that $\tilde{\mathrm{H}}_1=\tilde{\mathrm{H}}_2=0$, which in turn implies that

$$LR_n=\sup_{\boldsymbol{\theta}\in\Theta}\left\{2\mathcal{S}'_{\boldsymbol{\theta},n}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})-n\boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\}+O_p(n^{-1/8}) \tag{B21}$$

by virtue of Theorem 3, with $S_{\boldsymbol{\theta}}=(\tilde{\mathrm{H}}_3,\tilde{\mathrm{H}}_4)'$,

$$\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})=\left(-2\sqrt{3}\theta_1\theta_2,-\sqrt{6}\theta_2^2+\frac{\sqrt{6}}{9}\theta_1^4\right)',$$

and $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}=\mathbf{I}_2$. The rate $O_p(n^{-1/8})$ follows from the 8th-order Taylor expansion. Finally, after some tedious calculations available on request, we can verify that the conditions for Theorem 3 are satisfied in this example. In fact, we can further simplify the right-hand side of (B21) as follows. First, it is easy to see that an upper bound will be given by

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\{2\mathcal{S}'_{\boldsymbol{\theta},n}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})-n\boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\}\leq\frac{1}{n}\mathcal{S}'_{\boldsymbol{\theta},n}\mathcal{I}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\mathcal{S}_{\boldsymbol{\theta},n}=\frac{1}{n}\tilde{\mathrm{H}}_3^2+\frac{1}{n}\tilde{\mathrm{H}}_4^2.$$

Second, we can construct θ_1 and θ_2 such that

$$\begin{cases}-2\sqrt{3}\sqrt{n}\theta_1\theta_2=n^{-1/2}\tilde{\mathrm{H}}_3\\ -\sqrt{6}\sqrt{n}\theta_2^2+\frac{\sqrt{6}}{9}\sqrt{n}\theta_1^4=n^{-1/2}\tilde{\mathrm{H}}_4+O_p(n^{-1/4})\end{cases} \tag{B22}$$

which implies that a lower bound is

$$\frac{1}{n}\tilde{\mathrm{H}}_3^2+\frac{1}{n}\tilde{\mathrm{H}}_4^2+O_p(n^{-1/4}).$$

Specifically, if $n^{-1/2}\tilde{\mathrm{H}}_4\geq 0$, then we solve for

$$\begin{cases}-2\sqrt{3}\sqrt{n}\theta_1\theta_2=n^{-1/2}\tilde{\mathrm{H}}_3\\ \frac{\sqrt{6}}{9}\sqrt{n}\theta_1^4=n^{-1/2}\tilde{\mathrm{H}}_4\end{cases}$$

which gives $\theta_1 = O_p(n^{-1/8})$, $\theta_2 = O_p(n^{-3/8})$, so that (B22) holds with $\sqrt{n}\theta_2^2 = O_p(n^{-1/4})$. On the other hand, if $n^{-1/2}\tilde{\mathrm{H}}_4 < 0$, then we solve for

$$\begin{cases} -2\sqrt{3}\sqrt{n}\theta_1\theta_2 = n^{-1/2}\tilde{\mathrm{H}}_3 \\ -\sqrt{6}\sqrt{n}\theta_2^2 = n^{-1/2}\tilde{\mathrm{H}}_4 \end{cases}$$

which gives $\theta_1 = O_p(n^{-1/4})$, $\theta_2 = O_p(n^{-1/4})$, so that (B22) holds with $\sqrt{n}\theta_1^4 = O_p(n^{-1/2})$.

Therefore, we end up with

$$LR_n = \frac{1}{n}\tilde{\mathrm{H}}_3^2 + \frac{1}{n}\tilde{\mathrm{H}}_4^2 + O_p(n^{-1/8}),$$

as desired. □

Proof of Proposition 4

In this example,

$$\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{bmatrix} -2\sqrt{3}\theta_1\theta_2 \\ \sqrt{6}(\frac{1}{9}\theta_1^4 - \theta_2^2) \end{bmatrix}, \ \ \boldsymbol{\Lambda} = \mathbb{R}^2, \ \text{and} \ \ \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}'(\tilde{\boldsymbol{\phi}}) = \mathbf{I}_2.$$

Therefore, under the sequence

$$\lim_{n\to\infty} \sqrt{n}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_\infty) = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}\,,$$

we will have

$$\begin{aligned} GET_n &\xrightarrow{d} \sup_{\boldsymbol{\lambda}_{\boldsymbol{\theta}}\in\Lambda} \left\{2\left(S + \boldsymbol{\lambda}_{\infty,\boldsymbol{\theta}}\right)'\boldsymbol{\lambda}_{\boldsymbol{\theta}} - \boldsymbol{\lambda}_{\boldsymbol{\theta}}'\boldsymbol{\lambda}_{\boldsymbol{\theta}}\right\} \\ &= \left(S + \boldsymbol{\lambda}_{\infty,\boldsymbol{\theta}}\right)'\left(S + \boldsymbol{\lambda}_{\infty,\boldsymbol{\theta}}\right) \end{aligned}$$

as claimed. □

Proof of Proposition 5

The proof is entirely analogous to the proof of Proposition 8 in Amengual et al (2025), so we omit it for the sake of brevity. □

C Reparametrizations

C.1 Sequential reparametrization method

In what follows we explain how to obtain the reparametrization alluded to in section 2.1 using a sequential approach. To do so, we make the following:

Assumption 10 *1) The asymptotic covariance matrix of the sample averages of* $(\mathbf{s}_{\boldsymbol{\varphi}}, \mathbf{s}_{\boldsymbol{\vartheta}_1})$ *evaluated at* $(\boldsymbol{\varphi}, \mathbf{0})$ *scaled by* $\sqrt{n}$ *has full rank.*
2) $\left.\frac{\partial^{\boldsymbol{\iota}_{q_r}'\mathbf{j}_{q_r}}1}{\partial\boldsymbol{\vartheta}_r^{\mathbf{j}_{q_r}}}\right|_{(\boldsymbol{\varphi},\mathbf{0})} = 0$*, for all index vectors such that* $\boldsymbol{\iota}_{q_r}'\mathbf{j}_{q_r} < r-1$.
3) There exists a set of coefficients $\{m_{\mathbf{j}_{q_r},k}\}_{\boldsymbol{\iota}_{q_r}'\mathbf{j}_{q_r}=r-1,k=1,\ldots,p-q_r}$ *which may be functions of* $\boldsymbol{\varphi}$ *such that*

$$m_{\mathbf{j}_{q_r},1}s_{\varphi_1} + \ldots + m_{\mathbf{j}_{q_r},p}s_{\varphi_p} + m_{\mathbf{j}_{q_r},p+1}s_{\vartheta_{11}} + \ldots + m_{\mathbf{j}_{q_r}.p+q_1}s_{\vartheta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}_{q_r}'\mathbf{j}_{q_r}}1}{\partial\boldsymbol{\vartheta}^{\mathbf{j}_{q_r}}} = 0$$

for all $\boldsymbol{\iota}_{q_r}'\mathbf{j}_{q_r} = r-1$*, where the default argument is* $(\boldsymbol{\varphi},\mathbf{0})$.

In this context, a convenient way of reparametrizing the model from $(\boldsymbol{\varphi},\boldsymbol{\vartheta})$ to $(\boldsymbol{\phi},\boldsymbol{\theta})$ is as follows:

$$\varphi_1 = \phi_1 + \sum_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}=r-1} \frac{m_{\mathbf{j}_{q_r},1}}{\mathbf{j}_{q_r}!}\boldsymbol{\theta}_r^{\mathbf{j}_{q_r}}, \ldots, \varphi_p = \phi_p + \sum_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}=r-1} \frac{m_{\mathbf{j}_{q_r},p}}{\mathbf{j}_{q_r}!}\boldsymbol{\theta}_r^{\mathbf{j}_{q_r}}, \tag{C23}$$

$$\vartheta_{11} = \theta_{11} + \sum_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}=r-1} \frac{m_{\mathbf{j}_{q_r},p+1}}{\mathbf{j}_{q_r}!}\boldsymbol{\theta}_r^{\mathbf{j}_{q_r}}, \ldots, \vartheta_{1q_1} = \theta_{1q_1} + \sum_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}=r-1} \frac{m_{\mathbf{j}_{q_r}p+q_1}}{\mathbf{j}_{q_r}!}\boldsymbol{\theta}_r^{\mathbf{j}_{q_r}}, \tag{C24}$$

$$\vartheta_{r1} = \theta_{r1}, ..., \vartheta_{rq_r} = \theta_{rq_r}. \tag{C25}$$

Then, if we use Faà di Bruno (1859) formulas, which generalize the usual chain rule to higher-order derivatives, we can show that

$$\frac{\partial^{r-1}l}{\partial \boldsymbol{\theta}_r^{\mathbf{j}_{q_r}}} = m_{\mathbf{j}_{q_r}1}s_{\varphi_1} + ... + m_{\mathbf{j}_{q_r},p}s_{\varphi_p} + m_{\mathbf{j}_{q_r}p+1}s_{\vartheta_{11}} + ... + m_{\mathbf{j}_{q_r}p+q_1}s_{\vartheta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}}\mathrm{l}}{\partial \boldsymbol{\vartheta}^{\mathbf{j}_{q_r}}} = 0$$

for all $\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r} = r-1$ as desired, where the default argument is again $(\boldsymbol{\varphi},\mathbf{0})$.

Finally, we need to check whether $\sum_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}=r} \frac{\boldsymbol{\lambda}^{\mathbf{j}_{q_r}}}{\mathbf{j}_{q_r}!}\frac{\partial^r l}{\partial \boldsymbol{\theta}^{\mathbf{j}_{q_r}}}$ evaluated at $(\boldsymbol{\phi},\mathbf{0})$ is linearly independent of $(\mathbf{s}_{\boldsymbol{\phi}},\mathbf{s}_{\boldsymbol{\theta}_1})$ for all $\lambda_1^2+\cdots+\lambda_{q_r}^2=1$, so that Theorem 1 applies. Otherwise, we can check whether:

1) there exists a new set of coefficients $\{m^\dagger_{\mathbf{j}_{q_r}k}\}_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}=r,k=1,\ldots,p+q_1}$ which may be functions of $\boldsymbol{\phi}$ such that

$$m_1^{\dagger\mathbf{j}_{q_r}}s_{\phi_1} + \ldots + m^\dagger_{\mathbf{j}_{q_r},p}s_{\phi_p} + m^\dagger_{\mathbf{j}_{q_r},p+1}s_{\theta_{11}} + \ldots + m^\dagger_{\mathbf{j}_{q_r},p+q_1}s_{\theta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{q_r}}l}{\partial \theta^{\mathbf{j}_{q_r}}} = 0 \tag{C26}$$

when evaluated under the null, in which case we can do a further reparametrization from $(\boldsymbol{\phi},\boldsymbol{\theta})$ to $(\boldsymbol{\phi}^\dagger,\boldsymbol{\theta}^\dagger)$ in such a way that we set all the r^{th} partial derivatives with respect to $\boldsymbol{\theta}^\dagger$ to zero, or

2) we can use Theorem 3, which covers far more general cases.

C.2 Numerical invariance to reparametrization

For simplicity of notation, consider the simple case in which $r=2$ and $\boldsymbol{\theta}=\boldsymbol{\theta}_2$, so that we can omit the subscript 2 from $\boldsymbol{\theta}$ henceforth, and we also drop the subscript i from the contributions of each observation to the log-likelihood function.

Define $\boldsymbol{\varrho}=(\boldsymbol{\varphi},\boldsymbol{\vartheta})$ as the original parameter vector, where $\boldsymbol{\varphi}$ and $\boldsymbol{\vartheta}$ are vectors of dimension p and q, respectively. In what follows, $(\boldsymbol{\varphi},\mathbf{0})$ are the omitted arguments for all the relevant quantities that depend on $(\boldsymbol{\varphi},\boldsymbol{\vartheta})$.

We maintain that Assumption 3 holds with $r=2$ for the original parameters $\boldsymbol{\varrho}$, so that 1) the asymptotic variance of the sample average of $s_{\boldsymbol{\varphi}}$ has full rank, 2) there is a $q\times p$ matrix $\mathbf{M}$ of possible functions of $\boldsymbol{\varphi}$ such that

$$\mathbf{M}\mathbf{s}_{\boldsymbol{\varphi}i}(\boldsymbol{\varphi},\mathbf{0}) + \mathbf{s}_{\boldsymbol{\vartheta}i}(\boldsymbol{\varphi},\mathbf{0}) = \mathbf{0} \tag{C27}$$

holds, and 3) the asymptotic variance of the sample average of

$$\left[\mathbf{s}_{\boldsymbol{\varphi}}, \boldsymbol{\upsilon}'\begin{pmatrix}\mathbf{M}'\\ \mathbf{I}_q\end{pmatrix}'\frac{\partial^2 \mathrm{l}}{\partial\boldsymbol{\varrho}\partial\boldsymbol{\varrho}'}\begin{pmatrix}\mathbf{M}'\\ \mathbf{I}_q\end{pmatrix}\boldsymbol{\upsilon}\right]$$

has full rank under the null for all $\boldsymbol{\upsilon}$ such that $||\boldsymbol{\upsilon}|| \neq 0$. If we reparametrize from $\boldsymbol{\varrho}$ to $\boldsymbol{\rho}$ as

$$\boldsymbol{\varphi} = \boldsymbol{\phi} + \mathbf{M}'\boldsymbol{\theta}, \quad \text{and} \quad \boldsymbol{\vartheta} = \boldsymbol{\theta},$$

then, we can easily check that

$$\frac{\partial l}{\partial \boldsymbol{\phi}} = \frac{\partial \mathrm{l}}{\partial \boldsymbol{\varphi}}, \tag{C28}$$

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \mathbf{M}\frac{\partial \mathrm{l}}{\partial \boldsymbol{\varphi}} + \frac{\partial \mathrm{l}}{\partial \boldsymbol{\vartheta}} = \mathbf{M}\mathbf{s}_{\boldsymbol{\varphi}i} + \mathbf{s}_{\boldsymbol{\vartheta}i} = \mathbf{0}, \tag{C29}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = [\mathbf{M}, \mathbf{I}_q]\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varrho}\partial \boldsymbol{\varrho}'}\begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}.$$

In addition, (C28) and (C29) hold when evaluated under the null, with

$$\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\boldsymbol{\upsilon} = \boldsymbol{\upsilon}'\begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}'\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varrho}\partial \boldsymbol{\varrho}'}\begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}\boldsymbol{\upsilon}$$

linearly independent of $\partial l/\partial \boldsymbol{\phi}$, which implies that Assumption 3 is satisfied with $r = 2$ for the transformed parameters $\boldsymbol{\rho} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$ too. Therefore, we can apply Theorem 1, which yields $\mathrm{GET}_n^{\boldsymbol{\rho}} = \sup_{||\boldsymbol{\upsilon}||\neq 0} ET_n^{\boldsymbol{\rho}}(\boldsymbol{\upsilon})$, where

$$\begin{aligned} ET_n^{\boldsymbol{\rho}}(\boldsymbol{\upsilon}) &= \frac{[\boldsymbol{\upsilon}'\mathbb{H}(\tilde{\boldsymbol{\varphi}})\boldsymbol{\upsilon}]^2\,\mathbf{1}\,[\boldsymbol{\upsilon}'\mathbb{H}(\tilde{\boldsymbol{\varphi}})\boldsymbol{\upsilon} \geq \mathbf{0}]}{V(\boldsymbol{\upsilon}, \tilde{\boldsymbol{\varphi}})}, \\ \mathbb{H}(\boldsymbol{\varphi}) &= \begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}'\left.\frac{\partial^2 \mathrm{l}(\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho}\partial \boldsymbol{\varrho}'}\right|_{(\boldsymbol{\varphi},\mathbf{0})}\begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}, \end{aligned} \tag{C30}$$

and $\quad V_\eta(\boldsymbol{\upsilon}, \boldsymbol{\varphi}) = V[\boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}] - Cov[\boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}, \mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\varphi})]Var^{-1}[\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\varphi})]Cov[\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\varphi}), \boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}]$ is the adjusted variance of $\boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}$.

Consider now an alternative reparametrization from $\boldsymbol{\varrho}$ to $\boldsymbol{\rho}^\dagger$ characterized by

$$\boldsymbol{\varrho}' = \begin{pmatrix} \boldsymbol{\varphi}' & \boldsymbol{\vartheta}' \end{pmatrix} = \begin{bmatrix} \mathbf{g}^{\boldsymbol{\phi}}(\boldsymbol{\phi}^\dagger, \boldsymbol{\theta}^\dagger)' & \mathbf{g}^{\boldsymbol{\theta}}(\boldsymbol{\phi}^\dagger, \boldsymbol{\theta}^\dagger)' \end{bmatrix} = \mathbf{g}(\boldsymbol{\rho}^\dagger),$$

where $g(\cdot)$ is a vector of second-order continuously differentiable functions that represent a suitable diffeomorphism, at least locally around the null. Such an alternative reparametrization must also ensure that: (i) $\mathbf{s}_{\boldsymbol{\phi}^\dagger}$ has full rank, (ii) $\mathbf{s}_{\boldsymbol{\theta}^\dagger}$ is identically 0 at $H_0 : \boldsymbol{\theta}^\dagger = 0$, and (iii) $\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^\dagger\partial \boldsymbol{\theta}^{\dagger\prime}}\boldsymbol{\upsilon}$ is linearly independent of $\mathbf{s}_{\boldsymbol{\phi}^\dagger}$ for all $||\boldsymbol{\upsilon}|| \neq 0$.

Given that the first order derivative of $\boldsymbol{\phi}^\dagger$ under the null is given by

$$\frac{\partial l}{\partial \boldsymbol{\phi}^\dagger} = \frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\phi}^\dagger}\mathbf{s}_{\boldsymbol{\varphi}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\phi}^\dagger}\mathbf{s}_{\boldsymbol{\vartheta}} = \left(\frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\phi}^\dagger} - \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\phi}^\dagger}\mathbf{M}\right)\mathbf{s}_{\boldsymbol{\varphi}},$$

where we have used the chain rule in the first equality and (C27) in the second one, we need to assume that

$$\det\left(\frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\phi}^\dagger} - \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\phi}^\dagger}\mathbf{M}\right) \neq 0 \tag{C31}$$

for $\partial l/\partial \boldsymbol{\phi}^\dagger$ to have full rank. Similarly, given that (C27) and the chain rule imply that

$$\frac{\partial l}{\partial \boldsymbol{\theta}^\dagger} = \frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\theta}^\dagger}\mathbf{s}_{\boldsymbol{\varphi}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\theta}^\dagger}\mathbf{s}_{\boldsymbol{\vartheta}} = \left(\frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\theta}^\dagger} - \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\theta}^\dagger}\mathbf{M}\right)\mathbf{s}_{\boldsymbol{\varphi}},$$

we must also assume that

$$\frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\theta}^{\dagger}} = \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\theta}^{\dagger}}\mathbf{M} \tag{C32}$$

to ensure that $\partial l/\partial \boldsymbol{\theta}^{\dagger} = \mathbf{0}$ under the null irrespective of $\boldsymbol{\phi}^{\dagger}$ because $\mathbf{s}_{\boldsymbol{\varphi}}$ has full rank.

Let us now turn to condition (iii), for which we first need to compute the corresponding second-order derivatives. Applying the chain rule once again, we obtain

$$\begin{aligned}\frac{\partial^2 l}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} &= \frac{\partial \mathrm{l}}{\partial \boldsymbol{\varphi}'}\frac{\partial^2 \mathbf{g}^{\boldsymbol{\phi}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \theta_j^{\dagger}}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\frac{\partial \mathbf{g}^{\boldsymbol{\phi}}}{\partial \theta_i^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\varphi}'}\frac{\partial \mathbf{g}^{\boldsymbol{\phi}}}{\partial \theta_i^{\dagger}}\\ &+ \frac{\partial \mathrm{l}}{\partial \boldsymbol{\vartheta}'}\frac{\partial^2 \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \theta_j^{\dagger}}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\vartheta}'}\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}}.\end{aligned}$$

In this context, (C32) and (C27) imply that

$$\begin{aligned}\frac{\partial^2 l}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} &= \mathbf{s}_{\boldsymbol{\varphi}}'\frac{\partial^2 \mathbf{g}^{\boldsymbol{\phi}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\mathbf{M}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'}\mathbf{M}'\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\varphi}'}\mathbf{M}'\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}}\\ &- \mathbf{s}_{\boldsymbol{\varphi}}'\mathbf{M}'\frac{\partial^2 \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\mathbf{M}\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\vartheta}'}\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}}\\ &= \mathbf{s}_{\boldsymbol{\varphi}}'\left(\frac{\partial^2 \mathbf{g}^{\boldsymbol{\phi}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} - \mathbf{M}'\frac{\partial^2 \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}}\right) + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \theta_j^{\dagger}}\begin{pmatrix}\mathbf{M}'\\ \mathbf{I}_q\end{pmatrix}'\frac{\partial^2 \mathrm{l}}{\partial \boldsymbol{\varrho}\partial \boldsymbol{\varrho}'}\begin{pmatrix}\mathbf{M}'\\ \mathbf{I}_q\end{pmatrix}\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}}\end{aligned}$$

when evaluated at the null, so

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}} = \left\{\mathbf{s}_{\boldsymbol{\varphi}}'\left(\frac{\partial^2 \mathbf{g}^{\boldsymbol{\phi}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}} - \mathbf{M}'\frac{\partial^2 \mathbf{g}^{\boldsymbol{\theta}}}{\partial \theta_i^{\dagger}\partial \theta_j^{\dagger}}\right)\right\}_{ij} + \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\theta}^{\dagger}}\mathbb{H}\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger}}.$$

Hence, (C30) implies that

$$\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\upsilon} = \mathbf{s}_{\boldsymbol{\varphi}}'\mathbf{a} + \boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}, \text{ for all } \boldsymbol{\upsilon} \neq \mathbf{0}$$

when evaluated at the null, where $\mathbf{a} = (a_1, \ldots, a_q)'$ with

$$a_i = \boldsymbol{\upsilon}'\left(\frac{\partial^2 \mathbf{g}_i^{\boldsymbol{\phi}}}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}} - \mathbf{M}'\frac{\partial^2 \mathbf{g}_i^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\right)\boldsymbol{\upsilon} \text{ and } \boldsymbol{\upsilon}^{\dagger} = \frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger\prime}}\boldsymbol{\upsilon}.$$

If we further assume that

$$\det\left(\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger\prime}}\right) \neq 0, \tag{C33}$$

then $\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\upsilon}$ will be linearly independent of $s_{\boldsymbol{\phi}^{\dagger}}$ for all $\boldsymbol{\upsilon}^{\dagger}$ such that $\|\boldsymbol{\upsilon}^{\dagger}\| \neq 0$ because (a) $\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}$ is linearly independent of $\mathbf{s}_{\boldsymbol{\varphi}}$ and (b) $\mathbf{s}_{\boldsymbol{\phi}^{\dagger}}$ is a linear combination of $\mathbf{s}_{\boldsymbol{\varphi}}$.

Therefore, once we guarantee that (C31), (C32) and (C33) hold, the parametrization from $\boldsymbol{\varrho}$ to $\boldsymbol{\rho}^{\dagger}$ satisfies the rank deficiency condition in Assumption 3 with $r = 2$.

Finally, let us define the adjusted asymptotic variance of $\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\upsilon}$ as

$$\begin{aligned}V_{\eta^{\dagger}}^{\dagger}(\boldsymbol{\upsilon}, \boldsymbol{\phi}^{\dagger}) &= Var\left(\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\upsilon}\right) - Cov\left(\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\upsilon}, \mathbf{s}_{\boldsymbol{\phi}^{\dagger}}\right)Var^{-1}(\mathbf{s}_{\boldsymbol{\phi}^{\dagger}})Cov\left(\mathbf{s}_{\boldsymbol{\phi}^{\dagger}}, \boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\upsilon}\right)\\ &= Var(\mathbf{s}_{\boldsymbol{\varphi}}'\mathbf{a} + \boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}) - Cov(\mathbf{s}_{\boldsymbol{\varphi}}'\mathbf{a} + \boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}, \mathbf{a}'\mathbf{s}_{\boldsymbol{\varphi}})Var^{-1}(\mathbf{a}'\mathbf{s}_{\boldsymbol{\varphi}})Cov(\mathbf{a}'\mathbf{s}_{\boldsymbol{\varphi}}, \mathbf{s}_{\boldsymbol{\varphi}}'\mathbf{a} + \boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger})\\ &= Var(\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}) - Cov(\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}, \mathbf{s}_{\boldsymbol{\varphi}})Var^{-1}(\mathbf{s}_{\boldsymbol{\varphi}})Cov(\mathbf{s}_{\boldsymbol{\varphi}}, \boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}\boldsymbol{\upsilon}^{\dagger}) = V_{\eta}(\boldsymbol{\upsilon}^{\dagger}, \boldsymbol{\phi}).\end{aligned}$$

Then, we will have that

$$
\begin{aligned}
ET_n^{\boldsymbol{\rho}^\dagger}(\boldsymbol{\upsilon}) &= \frac{\left[\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial\boldsymbol{\theta}^\dagger\partial\boldsymbol{\theta}^\dagger}(\tilde{\boldsymbol{\rho}}^\dagger)\boldsymbol{\upsilon}\right]^2\mathbf{1}\left[\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial\boldsymbol{\theta}^\dagger\partial\boldsymbol{\theta}^\dagger}(\tilde{\boldsymbol{\rho}}^\dagger)\boldsymbol{\upsilon}\geq 0\right]}{V_{\eta^\dagger}^\dagger(\boldsymbol{\upsilon},\boldsymbol{\phi}^\dagger)} \\
&= \frac{[\mathbf{s}_{\boldsymbol{\varphi}}'(\tilde{\boldsymbol{\varphi}})\mathbf{a}+\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}(\tilde{\boldsymbol{\varrho}})\boldsymbol{\upsilon}^\dagger]^2\mathbf{1}\left[\mathbf{s}_{\boldsymbol{\varphi}}'(\tilde{\boldsymbol{\varphi}})\mathbf{a}+\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}(\tilde{\boldsymbol{\varrho}})\boldsymbol{\upsilon}^\dagger\geq 0\right]}{V_\eta(\boldsymbol{\upsilon}^\dagger,\boldsymbol{\phi})} \\
&= \frac{[\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}(\tilde{\boldsymbol{\varrho}})\boldsymbol{\upsilon}^\dagger]^2\mathbf{1}\left[\boldsymbol{\upsilon}^{\dagger\prime}\mathbb{H}(\tilde{\boldsymbol{\varrho}})\boldsymbol{\upsilon}^\dagger\geq 0\right]}{V_\eta(\boldsymbol{\upsilon}^\dagger,\boldsymbol{\phi})} = ET_n^{\boldsymbol{\rho}}(\boldsymbol{\upsilon}^\dagger),
\end{aligned}
$$

where the third equality follows from the fact that $\mathbf{s}_{\boldsymbol{\varphi}}(\tilde{\boldsymbol{\varphi}})=\mathbf{0}$. Given that the mapping from $\boldsymbol{\upsilon}$ to $\boldsymbol{\upsilon}^\dagger$ is bijective, taking the sup will finally imply that

$$
\text{GET}_n^{\boldsymbol{\rho}^\dagger}=\sup_{||\boldsymbol{\upsilon}||\neq 0}ET_n^{\boldsymbol{\rho}^\dagger}(\boldsymbol{\upsilon})=\sup_{||\boldsymbol{\upsilon}^\dagger||\neq 0}ET_n^{\boldsymbol{\rho}}(\boldsymbol{\upsilon}^\dagger)=\text{GET}_n^{\boldsymbol{\rho}},
$$

as desired.

D Example 3: Testing Gaussian vs Hermite copulas

D.1 The model and its log-likelihood function

The validity of the Gaussian copula in finance has been the subject of considerable debate. As a result, it is not surprising that several authors have considered more flexible copulas. For example, Amengual and Sentana (2020) look at the Generalized Hyperbolic copula, a location-scale Gaussian mixture which nests the popular Student t copula discussed by Fan and Patton (2014), which in turn nests the Gaussian one. In this section, we consider Hermite copulas instead, which can potentially provide much more flexible alternatives.

As is well known, Hermite polynomial expansions of the multivariate normal pdf can be understood as Edgeworth-like expansions of its characteristic function. They are based on multivariate Hermite polynomials of order $\boldsymbol{\iota}_K'\mathbf{j}$ where $\mathbf{j}\in N^K$, which are defined as differentials of the multivariate normal density:

$$
H_{\mathbf{j}}(\mathbf{x},\boldsymbol{\varphi})=f_{NK}(\mathbf{x};\mathbf{R})^{-1}\left(\frac{-\partial}{\partial\mathbf{x}}\right)^{\mathbf{j}}f_{NK}(\mathbf{x};\mathbf{R}), \tag{D34}
$$

$\boldsymbol{\varphi}=vecl(\mathbf{R})$, and $\mathbf{R}$ is a positive definite correlation matrix.

To keep the expressions manageable, we only consider explicitly pure fourth-order expansions in the bivariate case. We could also include third-order Hermite polynomials, but at a considerable cost in terms of notation. Similarly, extensions to higher dimensions would be tedious but otherwise straightforward.

We say that (x_1,x_2) follow a pure fourth-order Hermite expansion of the Gaussian distribution when their joint density function is given by

$$
f_H(x_1,x_2;\varphi,\boldsymbol{\vartheta})=f_{N2}\left[\begin{pmatrix}x_1\\x_2\end{pmatrix};\begin{pmatrix}1&\varphi\\\varphi&1\end{pmatrix}\right]P(x_1,x_2;\varphi,\boldsymbol{\vartheta}), \tag{D35}
$$

where

$$P(x_1, x_2; \varphi, \boldsymbol{\vartheta}) = 1 + \sum_{j=0}^{4} \vartheta_{j+1} H_{4-j,j}(x_1, x_2; \varphi)$$

with φ being the correlation between x_1 and x_2, which we assume is different from 0, and $\vartheta_1, \ldots, \vartheta_5$ the coefficients of the expansion. The leading term in (D35) is the normal pdf and the remaining terms represent departures from normality. Indeed, $f_H(x_1, x_2; \varphi, \boldsymbol{\vartheta})$ reduces to a Gaussian distribution when $\boldsymbol{\vartheta} = \mathbf{0}$.

It is then straightforward to show that the corresponding marginal distributions are given by

$$\left.\begin{array}{l} f_H(x_1; \vartheta_1) = \phi(x_1)[1 + \vartheta_1 H_{40}(x_1, x_2)] \\ f_H(x_2; \vartheta_5) = \phi(x_2)[1 + \vartheta_5 H_{04}(x_1, x_2)] \end{array}\right\}, \tag{D36}$$

where $\phi(.)$ the standard normal pdf and $H_{40}(x_1, x_2)$ and $H_{04}(x_1, x_2)$ are the (non-standardized) fourth-order univariate Hermite polynomials for x_1 and x_2, respectively.

Hermite expansion copulas are based on Hermite expansion distributions: letting $\mathbf{y} = (y_1, y_2)$ denote the original data, defining $\mathbf{u} = (u_1, u_2) = [F_1(y_1), F_2(y_2)]$ as the uniform ranks of $\mathbf{y}$, and finally $\mathbf{x} = (x_1, x_2) = [F_H^{-1}(u_1; \vartheta_1), F_H^{-1}(u_2; \vartheta_5)]$, where $F_H^{-1}(.; \vartheta_i)$ are the inverse cdfs (or quantile functions) of the univariate fourth-order Hermite expansions with parameter ϑ_i in (D36).

Consequently, the pdf of the pure fourth-order Hermite expansion copula is

$$\frac{f_H(x_1, x_2; \boldsymbol{\varrho})}{f_H(x_1; \vartheta_1) f_H(x_2; \vartheta_5)} = \frac{\phi_2(x_1, x_2; \varphi)[1 + \sum_{j=0}^{4} \vartheta_{j+1} H_{4-j,j}(x_1, x_2; \varphi)]}{\phi_1(x_1)[1 + \vartheta_1 H_{40}(x_1, x_2)]\phi_1(x_2)[1 + \vartheta_5 H_{04}(x_1, x_2)]}.$$

D.2 The null hypothesis and the GET test statistic

Straightforward calculations allow us to show that in this case

$$s_{\vartheta_1}(\varphi, \mathbf{0}) + 3\varphi s_{\vartheta_2}(\varphi, \mathbf{0}) + 3\varphi^2 s_{\vartheta_3}(\varphi, \mathbf{0}) + \varphi^3 s_{\vartheta_4}(\varphi, \mathbf{0}) = 0,$$
$$s_{\vartheta_5}(\varphi, \mathbf{0}) + 3\varphi s_{\vartheta_4}(\varphi, \mathbf{0}) + 3\varphi^2 s_{\vartheta_3}(\varphi, \mathbf{0}) + \varphi^3 s_{\vartheta_2}(\varphi, \mathbf{0}) = 0,$$

and, hence, our proposed reparametrization, namely

$$\varphi = \phi, \quad \vartheta_1 = \theta_{21}, \quad \vartheta_2 = \theta_{11} + 3\phi\theta_{21} + \phi^3\theta_{22},$$
$$\vartheta_3 = \theta_{12} + 3\phi^2\theta_{21} + 3\phi^2\theta_{22}, \quad \vartheta_4 = \theta_{13} + 3\phi\theta_{22} + \phi^3\theta_{21}, \quad \vartheta_5 = \theta_{22},$$

confines the singularity to the scores of θ_{21} and θ_{22}. Therefore, we need to obtain the second order derivatives with respect to θ_{21} and θ_{22}. Specifically, we can prove that the asymptotic covariance matrix of

$$\frac{\partial l}{\partial \phi}, \frac{\partial l}{\partial \theta_{11}}, \frac{\partial l}{\partial \theta_{12}}, \frac{\partial l}{\partial \theta_{13}}, \frac{\partial^2 l}{\partial \theta_{21}^2}, \frac{\partial^2 l}{\partial \theta_{22}^2} \text{ and } \frac{\partial^2 l}{\partial \theta_{21} \partial \theta_{22}}$$

scaled by $\sqrt{n}$ has full rank. Although the algebra is a bit messy, after orthogonalizing those second derivatives with respect to the score of ϕ to eliminate the effect of the sampling uncertainty in estimating this correlation coefficient under the null, we can express the three second-order derivatives as linear combinations of all the even-order multivariate Hermite polynomials of (x_1, x_2) up to the 8^{th} order, whose coefficients depend on φ, as we explain below.

Let $\theta_{21} = \upsilon_1\eta$ and $\theta_{22} = \upsilon_2\eta$ with $\upsilon_1^2 + \upsilon_2^2 = 1$, and consider the simplified null hypothesis $H_0 : \theta_{11} = \theta_{12} = \theta_{13} = \eta = 0$. Then, it is easy to see that the GET statistic will be

$$\frac{1}{n}S'_{1n}V_{11}^{-1}S_{1n} + \frac{1}{n}\sup_{\|\boldsymbol{\upsilon}\|=1}(V_{\eta\eta} - V_{\eta 1}V_{11}^{-1}V_{1\eta})^{-1}D_n^2\mathbf{1}\left[D_n > 0\right], \tag{D37}$$

where

$$\begin{aligned} D_n(\boldsymbol{\phi}, \eta, \boldsymbol{\upsilon}) &= H_{\eta n}(\boldsymbol{\phi}, \eta, \boldsymbol{\upsilon}) - V_{\eta 1}(\boldsymbol{\phi}, \eta, \boldsymbol{\upsilon})V_{11}^{-1}(\boldsymbol{\phi})S_{1n}(\boldsymbol{\phi}, \mathbf{0}), \\ H_{\eta n}(\boldsymbol{\phi}, \eta, \boldsymbol{\upsilon}) &= \sum_{i=1}^{n}(\upsilon_1\ \upsilon_2)\begin{bmatrix} h_{\theta_{21}\theta_{21},i}(\boldsymbol{\rho}) & h_{\theta_{21}\theta_{22},i}(\boldsymbol{\rho}) \\ h_{\theta_{21}\theta_{22},i}(\boldsymbol{\rho}) & h_{\theta_{22}\theta_{22},i}(\boldsymbol{\rho}) \end{bmatrix}\begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix}, \\ S_{1n}(\boldsymbol{\phi}, \mathbf{0}) &= [S_{\theta_{11}}(\boldsymbol{\phi}, \mathbf{0}), S_{\theta_{12}}(\boldsymbol{\phi}, \mathbf{0}), S_{\theta_{13}}(\boldsymbol{\phi}, \mathbf{0})]', \end{aligned}$$

where the omitted arguments are $(\tilde{\phi}, \mathbf{0}, \boldsymbol{\upsilon})$ for D_n, $(\tilde{\phi}, \boldsymbol{\upsilon})$ for $V_{\eta\eta}$, $V_{\eta 1}$ and $V_{1\eta}$, $(\tilde{\phi}, \mathbf{0})$ for $S_{1,n}$ and $\tilde{\phi}$ for V_{11}.

As a consequence, the asymptotic distribution of GET is bounded above by a χ_6^2 distribution because of the six influence functions, while it is bounded below by a 50:50 mixture of χ_3^2 and χ_4^2 because θ_{11}, θ_{12} and θ_{13} are first-order identified parameters and an even-order derivative of η is involved.

D.3 Computational details

D.3.1 Influence functions

In practice, the calculation of the GET statistic requires explicit expressions for all the different ingredients that appear in (D37). Tedious but straightforward algebra implies that

$$\frac{\partial l}{\partial \phi} = (0, 1, 0) \cdot \mathbf{H}_2(x_1, x_2; \phi),$$

$$\frac{\partial l}{\partial \theta_{11}} = H_{31}(x_1, x_2; \phi), \quad \frac{\partial l}{\partial \theta_{12}} = H_{22}(x_1, x_2; \phi), \quad \frac{\partial l}{\partial \theta_{13}} = H_{13}(x_1, x_2; \phi),$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \theta_{21}^2} &= (0, 6\phi, 0) \cdot \mathbf{H}_2(x_1, x_2; \phi) + (0, 18\phi, 36\phi^2, 18\phi^3, 0) \cdot \mathbf{H}_4(x_1, x_2; \phi) \\ &+ (0, 9\phi, 36\phi^2, 54\phi^3, 36\phi^4, 9\phi^5, 0) \cdot \mathbf{H}_6(x_1, x_2; \phi) \\ &+ (0, \phi, 6\phi^2, 15\phi^3, 20\phi^4, 15\phi^5, 6\phi^6, \phi^7, 0) \cdot \mathbf{H}_8(x_1, x_2; \phi), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \theta_{21}\partial \theta_{22}} &= -(0, 6\phi^3, 0) \cdot \mathbf{H}_2(x_1, x_2; \phi) - \left[0, 18\phi^3, 18\left(\phi^4 + \phi^2\right), 18\phi^3, 0\right] \cdot \mathbf{H}_4(x_1, x_2; \phi) \\ &- \left[0, 9\phi^3, 18\left(\phi^4 + \phi^2\right), 9\left(\phi^5 + 4\phi^3 + \phi\right), 18\left(\phi^4 + \phi^2\right), 9\phi^3, 0\right] \cdot \mathbf{H}_6(x_1, x_2; \phi) \\ &- \left[0, \phi^3, 3\left(\phi^4 + \phi^2\right), 3\left(\phi^5 + 3\phi^3 + \phi\right), \phi^6 + 9\phi^4\right. \\ &\left. +9\phi^2 + 1, 3\left(\phi^5 + 3\phi^3 + \phi\right), 3\left(\phi^4 + \phi^2\right), \phi^3, 0\right] \cdot \mathbf{H}_8(x_1, x_2; \phi) \end{aligned}$$

and

$$\frac{\partial l}{\partial \theta_{22}^2} = (0, 6\phi, 0) \cdot \mathbf{H}_2(x_1, x_2; \phi) + \left(0, 18\phi^3, 36\phi^2, 18\phi, 0\right) \cdot \mathbf{H}_4(x_1, x_2; \phi)$$
$$+ \left(0, 9\phi^5, 36\phi^4, 54\phi^3, 36\phi^2, 9\phi, 0\right) \cdot \mathbf{H}_6(x_1, x_2; \phi)$$
$$+ \left(0, \phi^7, 6\phi^6, 15\phi^5, 20\phi^4, 15\phi^3, 6\phi^2, \phi, 0\right) \cdot \mathbf{H}_8(x_1, x_2; \phi),$$

where $\mathbf{H}_k(x_1, x_2; \phi) = [H_{k0}(x_1, x_2; \phi), H_{k-1,1}(x_1, x_2; \phi), ..., H_{0,k}(x_1, x_2; \phi)]'$.

D.3.2 Positivity of the Hermite expansion of the Gaussian copula

The foregoing derivations, though, ignore that the positivity of the Hermite copula density for all values of $\mathbf{y}$ imposes highly nonlinear inequality constraints on the elements of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ with $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12}, \theta_{13})'$ and $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})'$, the latter being the ones affecting the marginal distributions after reparametrization. Therefore, Assumption 1.1 fails because $\boldsymbol{\rho}_0$ lies at the boundary of the admissible parameter space. Nevertheless, we can still derive an LR-equivalent test. Specifically, given that under the null hypothesis of a Gaussian copula the UMLE estimators of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ converge at rates $n^{-\frac{1}{2}}$ and $n^{-\frac{1}{4}}$, respectively, the elements of the sequence $\boldsymbol{\theta}_{1n}$ are negligible and we simply need to find the asymptotes of the feasible set for $(\theta_{21}, \theta_{22})$. Let $\theta_{21} = \eta \upsilon_1 = \eta \sin(\omega)$ and $\theta_{22} = \eta \upsilon_2 = \eta \cos(\omega)$ with $\omega \in [0, 2\pi)$ to ensure a unit norm for $\boldsymbol{\upsilon} = (\upsilon_1, \upsilon_2)'$. These parameters lead to a positive density when η is small enough if and only if $\omega \in (\omega_l, \omega_u)$, with ω_l and ω_u defined in (D40) and, therefore, an asymptotically equivalent GET statistic that imposes positivity of the Hermite expansion copula under admissible alternatives local to the null will be given by

$$\frac{1}{n} S_{1n}' V_{11}^{-1} S_{1n} + \frac{1}{n} \sup_{\omega \in (\omega_l, \omega_u)} D_n' \left(V_{\eta\eta} - V_{\eta 1} V_{11}^{-1} V_{1\eta}\right)^{-1} D_n \mathbf{1}\left[D_n > 0\right], \tag{D38}$$

which is asymptotically equivalent to the LR test by imposing positivity because a zero density gives rise to an infinitely penalized log-likelihood. Importantly, our test is again far more computationally convenient than the LR test because the positivity constraints effectively become linear under local alternatives.

To justify these claims, it is convenient to remember that in the original parametrization, $P(x_1, x_2; \varphi, \boldsymbol{\vartheta})$ is equal to

$$1 + \vartheta_1 H_{40}(x_1, x_2; \varphi) + \vartheta_2 H_{31}(x_1, x_2; \varphi) + \vartheta_3 H_{22}(x_1, x_2; \varphi) + \vartheta_4 H_{13}(x_1, x_2; \varphi) + \vartheta_5 H_{04}(x_1, x_2; \varphi).$$

But as mentioned before, $\boldsymbol{\theta}_1$ is dominated, at least asymptotically. For that reason, we first discuss the positivity constraint on $\boldsymbol{\theta}_2$ when $\boldsymbol{\theta}_1 = \mathbf{0}$, and then explain how to simplify the asymptotic positivity constraint and the extremum test statistic.

Let $x_2 = tx_1$, $\theta_{22} = k\theta_{21}$, $k \geq 0$ so that the polynomial that multiplies the Gaussian pdf simplifies to

$$\begin{aligned} \tilde{P}(x_1, \phi, k, t, \theta_{21}) &= P[x_1, tx_1; \phi, (\theta_{21}, 0, 0, 0, k\theta_{21})'] \\ &= 1 + 3\theta_{21} C_0(k) + \frac{3\theta_{21}}{1-\phi^2} C_2(k, t, \phi) x_1^2 + \frac{\theta_{21}}{1-\phi^2} C_4(k, t, \phi) x_1^4, \text{ where} \end{aligned}$$

$C_0(k) = k+1,\ C_2(k,t,\phi) = k\left(\phi^2-2\right)t^2+(k+1)\,\phi t+\phi^2-2$ and $C_4(k,t,\phi) = kt^4-k\phi t^3-\phi t+1$.

It is easy then to see that $\min_x \tilde{P}(x,\phi,k,t,\theta_{21})$ is finite if and only if (i) $C_4(k,t,\phi) > 0$ or (ii) $C_4(k,t,\phi) = 0$ and $C_2(k,t,\phi) \geq 0$. In addition, when θ_{21} is very small under either (i) or (ii), we have that $\min_x \tilde{P}(x,\phi,k,t,\theta_{21})$ is greater than 0. Thus, we need to find a set $K(\phi)$ such that $\forall \phi \neq 0$, $\forall k \in K(\phi) \subseteq [0,+\infty)$ and $\forall t \in R$, we have either (i) or (ii), and, thereby, we need $C_4(k,t,\phi) = kt^4 - k\phi t^3 - \phi t + 1 \geq 0\ \forall t$.

To guarantee the positivity of this expression, we need $k > 0$. If the discriminant of $C_4(k,t,\phi)$ is positive, then $C_4(\cdot,t,\cdot) = 0$ has either only real or only complex roots, while if the discriminant is negative, then $C_4(\cdot,t,\cdot) = 0$ will have both two real and two complex roots, while if the discriminant is zero, then at least two roots must be equal. Therefore, to ensure discriminant of $C_4(k,t,\phi)$ to be non-negative, we need to find two functions, $lb(\phi)$ and $ub(\phi)$ such that $lb(\phi) < k < ub(\phi)$ if and only if the discriminant is positive while $k \in \{lb(\phi), ub(\phi)\}$ if and only if the discriminant is zero. Moreover, $lb(\phi) \in (0,1)$, $ub(\phi) \in (1,+\infty)$, and $lb(\phi)ub(\phi) = 1$. The proof of these statements is as follows:

Specifically, we can show that

$$Disc_t[C_4(k,t,\phi)] = -k^2[27k^2\phi^4 + 2k\left(2\phi^6+3\phi^4+96\phi^2-128\right) + 27\phi^4],$$

so that the solution to

$$Disc_t[C_4(k,t,\phi)] = 0$$

is given by
$$\begin{cases} lb(\phi) = -\dfrac{2\phi^6+3\phi^4+96\phi^2+2(\sqrt{\left(\phi^2-4\right)^3\left(\phi^2-1\right)\left(\phi^2+8\right)^2}-64)}{27\phi^4} \\ ub(\phi) = -\dfrac{2\phi^6+3\phi^4+96\phi^2-2(\sqrt{\left(\phi^2-4\right)^3\left(\phi^2-1\right)\left(\phi^2+8\right)^2}+64)}{27\phi^4} \end{cases}$$

Thus, when $k \in [lb(\phi), ub(\phi)]$, the discriminant is positive and we simply need to check whether $C_4(k,t,\phi) \geq 0$. First, consider $\phi > 0$ and $C_4(k,t,\phi) = kt^3(t-\phi)-\phi t+1$. When $t \geq \phi$, $C_4(k,t,\phi)$ is increasing in k. In this context, we can prove that $\min_{t\geq\phi} C_4[lb(\phi),t,\phi] = 0$. In turn, when $t \in [0,\phi)$, $C_4(k,t,\phi)$ is decreasing in k, and we have $\min_{t\geq\phi} C_4[ub(\phi),t,\phi] = 0$. Finally, when $t < 0$, it is obvious that $C_4(k,t,\phi) > 0$.

However, when either $k = lb(\phi)$ or $k = ub(\phi)$, we have t_l, t_u defined by $C_4[lb(\phi),t_l,\phi] = 0$ and $C_4[ub(\phi),t_u,\phi] = 0$, respectively, so that

$$C_2[lb(\phi),t_l,\phi] < 0 \ \text{ and } C_2[ub(\phi),t_u,\phi] < 0 \ \text{ for all } \phi,$$

which in turn implies that $k \in \{lb(\phi), ub(\phi)\}$ does not hold.

Therefore, when $\boldsymbol{\theta}_1 = \mathbf{0}$, the asymptotes of the feasible set near 0 are $\theta_{22} = lb(\phi)\theta_{21}$ and $\theta_{22} = ub(\phi)\theta_{21}$, and Theorem 1 implies that

$$LR = ET(\boldsymbol{\theta}^{ET}) + O_p(n^{-\frac{1}{2r}}) \tag{D39}$$

with

$$ET_n(\boldsymbol{\theta}) = 2\begin{pmatrix} n^{\frac{1}{2}}\boldsymbol{\theta}_1 \\ n^{\frac{1}{2}}\theta_{21}^2 \\ n^{\frac{1}{2}}\theta_{21}\theta_{22} \\ n^{\frac{1}{2}}\theta_{22}^2 \end{pmatrix}'\begin{pmatrix} n^{-\frac{1}{2}}S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0}) \\ n^{-\frac{1}{2}}H_{\theta_{21}\theta_{21}}(\tilde{\phi},\mathbf{0}) \\ n^{-\frac{1}{2}}H_{\theta_{21}\theta_{22}}(\tilde{\phi},\mathbf{0}) \\ n^{-\frac{1}{2}}H_{\theta_{22}\theta_{22}}(\tilde{\phi},\mathbf{0}) \end{pmatrix} - \begin{pmatrix} n^{\frac{1}{2}}\boldsymbol{\theta}_1 \\ n^{\frac{1}{2}}\theta_{21}^2 \\ n^{\frac{1}{2}}\theta_{21}\theta_{22} \\ n^{\frac{1}{2}}\theta_{22}^2 \end{pmatrix}' V_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\phi})\begin{pmatrix} n^{\frac{1}{2}}\boldsymbol{\theta}_1 \\ n^{\frac{1}{2}}\theta_{21}^2 \\ n^{\frac{1}{2}}\theta_{21}\theta_{22} \\ n^{\frac{1}{2}}\theta_{22}^2 \end{pmatrix},$$

$$\boldsymbol{\theta}^{ET} = \arg\max_{\boldsymbol{\theta}\in\Theta} ET_n(\boldsymbol{\theta}),$$

where Θ is the set of parameters that satisfies the positivity constraint. Unfortunately, $ET_n(\theta^{ET})$ is not very easy to calculate because Θ is difficult to characterize explicitly. For that reason, we below show that

$$ET_n(\boldsymbol{\theta}^{ET}) = GET_n + o_p(1), \text{ where}$$

$$GET_n = \frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0}) + \sup_{\omega\in(\omega_l,\omega_u)}\frac{1}{n}\frac{D^2(\tilde{\phi},\boldsymbol{\upsilon})\mathbf{1}[D(\tilde{\phi},\boldsymbol{\upsilon})\geq 0]}{V_{22}(\tilde{\phi},\boldsymbol{\upsilon}) - V_{21}(\tilde{\phi},\boldsymbol{\upsilon})V_{11}^{-1}(\tilde{\phi})V_{12}(\tilde{\phi},\boldsymbol{\upsilon})},$$

with $\upsilon_1 = \sin(\omega)$ and $\upsilon_2 = \cos(\omega)$ so that $\|\boldsymbol{\upsilon}\| = 1$, and

$$\omega_l = \arctan[lb(\tilde{\phi})], \quad \omega_u = \arctan[ub(\tilde{\phi})], \tag{D40}$$

so that letting $\theta_{21} = \upsilon_1\eta$ and $\theta_{22} = \upsilon_2\eta$, then

$$ET_n(\boldsymbol{\theta}_1,\eta,\boldsymbol{\upsilon}) = 2\begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}'\begin{pmatrix} \mathcal{S}_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0}) \\ \mathcal{S}_{\boldsymbol{\theta}_2}(\tilde{\phi},\mathbf{0},\boldsymbol{\upsilon}) \end{pmatrix} - n\begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}'\begin{bmatrix} V_{11}(\tilde{\phi}) & V_{12}(\tilde{\phi},\boldsymbol{\upsilon}) \\ V_{21}(\tilde{\phi},\boldsymbol{\upsilon}) & V_{22}(\tilde{\phi},\boldsymbol{\upsilon}) \end{bmatrix}\begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}, \tag{D41}$$

with $$\mathcal{S}_{\boldsymbol{\theta}_2}(\phi,0,\boldsymbol{\upsilon}) = \begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix}'\begin{bmatrix} H_{\theta_{21}\theta_{21}}(\phi,\mathbf{0}) & H_{\theta_{21}\theta_{22}}(\phi,\mathbf{0}) \\ H_{\theta_{21}\theta_{22}}(\phi,\mathbf{0}) & H_{\theta_{22}\theta_{22}}(\phi,\mathbf{0}) \end{bmatrix}\begin{pmatrix} \upsilon_1 \\ \upsilon_2 \end{pmatrix}.$$

Similarly, letting $\tilde{\eta} = \max\{\eta^{ET}, n^{-k}\}$ with $\frac{1}{4} < k < \frac{1}{2}$ it is easy to notice that

$$ET_n(\boldsymbol{\theta}_1^{ET},\tilde{\eta},\boldsymbol{\upsilon}^{ET}) = ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) + o_p(1). \tag{D42}$$

Next, considering $(\theta_1^*,\eta^*,\upsilon^*) = \text{argmax}_{pc\wedge\{\eta\geq n^{-k}\}} ET_n(\theta_1,\eta,\upsilon)$, where $pc = \{(\boldsymbol{\theta}_1,\eta\upsilon_1,\eta\upsilon_2)\in\Theta\}$, it is easy to see that w.p.a.1,

$$ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) \geq ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*) \geq ET_n(\boldsymbol{\theta}_1^{ET},\tilde{\eta},\boldsymbol{\upsilon}^{ET}) \tag{D43}$$

because $(\theta_1^{ET},\eta^{ET},\upsilon^{ET}) = \arg\max_{pc} ET_n(\theta_1,\eta,\upsilon)$ is over a larger feasible set, and the event $(\theta_1^{ET},\tilde{\eta},\upsilon^{ET})\in pc$ and $\{\tilde{\eta}\geq n^{-k}\}$ happens w.p.a.1. Combining (D42) and (D43), we get

$$ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*) = ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) + o_p(1), \tag{D44}$$

so we only need to calculate $(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*)$.

Next, noticing that there exists a $k'\in(k,\frac{1}{2})$ such that

$$\lim_n \Pr(\|\boldsymbol{\theta}_1^*\| < n^{-k'} < n^{-k} \leq \eta^*) = 1, \tag{D45}$$

$\boldsymbol{\theta}_1^*$ becomes asymptotically irrelevant for the positivity constraints because it is effectively unrestricted and, consequently, (D45) implies that the only relevant restriction will affect the direction of $\boldsymbol{\theta}_2$.

In view of (D41), the first order condition for $\boldsymbol{\theta}_1^*$ for given η^* and $\boldsymbol{\upsilon}^*$ implies that

$$n^{\frac{1}{2}}\boldsymbol{\theta}_1^*(\eta^*,\boldsymbol{\upsilon}^*)=V_{11}^{-1}(\tilde{\phi})[n^{-\frac{1}{2}}S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})-V_{12}(\tilde{\phi},\boldsymbol{\upsilon}^*)n^{\frac{1}{2}}\eta^{*2}]$$

and, hence, if we substitute $\theta_1^*(\eta^*,\upsilon^*)$ in the expression for $ET(\theta_1,\eta,\upsilon)$, we end up with

$$\begin{aligned}ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*)&=\frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})\\&-n^{\frac{1}{2}}\eta^{*2}[V_{22}(\tilde{\phi},\boldsymbol{\upsilon}^*)-V_{21}(\tilde{\phi},\boldsymbol{\upsilon}^*)V_{11}^{-1}(\tilde{\phi})V_{12}(\tilde{\phi},\boldsymbol{\upsilon}^*)]n^{\frac{1}{2}}\eta^{*2}\\&+2n^{\frac{1}{2}}\eta^{*2}[n^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{\theta}_2}(\tilde{\phi},\mathbf{0},\boldsymbol{\upsilon}^*)-V_{21}(\tilde{\phi},\boldsymbol{\upsilon}^*)V_{11}^{-1}(\tilde{\phi})n^{-\frac{1}{2}}S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})].\end{aligned}\tag{D46}$$

Given that (D46) is quadratic in η^{*2}, if take into account the restriction $\eta^*\geq n^{-k}$, we obtain

$$\eta^*(\boldsymbol{\upsilon}^*)=\max\left\{n^{-\frac{1}{4}}\sqrt{[V_{22}(\tilde{\phi},\boldsymbol{\upsilon}^*)-V_{21}(\tilde{\phi},\boldsymbol{\upsilon}^*)V_{11}^{-1}(\tilde{\phi})V_{12}(\tilde{\phi},\boldsymbol{\upsilon}^*)]n^{-\frac{1}{2}}D(\tilde{\phi},\boldsymbol{\upsilon}^*)\mathbf{1}[D(\tilde{\phi},\boldsymbol{\upsilon}^*)\geq 0]},n^{-k}\right\},$$

where $D(\phi,\boldsymbol{\upsilon})=S_{\boldsymbol{\theta}_2}(\phi,\mathbf{0},\boldsymbol{\upsilon}^*)-V_{21}(\phi,\boldsymbol{\upsilon})V_{11}^{-1}(\phi)S_{\boldsymbol{\theta}_1}(\phi,\mathbf{0})$. Finally, if we replace the previous expression for $\eta^*(\boldsymbol{\upsilon}^*)$ into (D46), we end up with

$$\begin{aligned}ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*)&=\frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})\\&+\underbrace{\frac{1}{n}\frac{D^2(\tilde{\phi},\boldsymbol{\upsilon}^*)\mathbf{1}[D(\tilde{\phi},\boldsymbol{\upsilon}^*)\geq 0]}{V_{22}(\tilde{\phi},\boldsymbol{\upsilon}^*)-V_{21}(\tilde{\phi},\boldsymbol{\upsilon}^*)V_{11}^{-1}(\tilde{\phi})V_{12}(\tilde{\phi},\boldsymbol{\upsilon}^*)}}_{\text{part 2}}+o_p(1).\end{aligned}\tag{D47}$$

But since part 2 in (D47) is a function of $\boldsymbol{\upsilon}^*$, which by definition is a maximizer of ET_n, we end up with

$$\begin{aligned}ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*)&=\frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})\\&+\sup_{\omega\in(\omega_l,\omega_u)}\frac{1}{n}\frac{D^2(\tilde{\phi},\boldsymbol{\upsilon})\mathbf{1}[D(\tilde{\phi},\boldsymbol{\upsilon})\geq 0]}{V_{22}(\tilde{\phi},\boldsymbol{\upsilon})-V_{21}(\tilde{\phi},\boldsymbol{\upsilon})V_{11}^{-1}(\tilde{\phi})V_{12}(\tilde{\phi},\boldsymbol{\upsilon})}+o_p(1),\end{aligned}$$

which, in view of (D44), confirms that

$$\begin{aligned}ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET})&=\frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0})\\&+\sup_{\omega\in(\omega_l,\omega_u)}\frac{1}{n}\frac{D^2(\tilde{\phi},\boldsymbol{\upsilon})\mathbf{1}[D(\tilde{\phi},\boldsymbol{\upsilon})\geq 0]}{V_{22}(\tilde{\phi},\boldsymbol{\upsilon})-V_{21}(\tilde{\phi},\boldsymbol{\upsilon})V_{11}^{-1}(\tilde{\phi})V_{12}(\tilde{\phi},\boldsymbol{\upsilon})}+o_p(1).\end{aligned}$$

D.4 Simulation evidence

For simplicity, we assume the marginal distributions are known, so that we can directly work with the uniform ranks, which we immediately convert into Gaussian ranks (see Amengual and Sentana (2020)). We estimate the correlation parameter, whose true value we set to 0.5 under both the null and alternative hypotheses using the Gaussian rank correlation in Amengual, Sentana and Tian (2022), which effectively imposes the null. As alternative hypotheses, we consider two Hermite expansion copulas: one with $\boldsymbol{\vartheta}'=(0.03,0,0,0,0)$ (H_{a_1}) and another with $\boldsymbol{\vartheta}' = (0.02, 0, 0, 0, 0.02)$ (H_{a2}). While the second one generates a copula density which is symmetric around the 45° line, the first one does not. In any event, both departures from the Gaussian copula are rather mild, as they only involve one or two parameters different from 0.

If the correlation coefficient were known, we could again compute exact critical values under the null for any sample size to any degree of accuracy by repeatedly simulating samples of *iid* bivariate normals with correlation φ. In practice, though, we fix the correlation coefficient to its estimated value in each sample in what is effectively the parametric bootstrap procedure described in section 2.3 (see Appendix D.1 in Amengual and Sentana (2015) for further details).

In Table 3 we compare the results of our tests with three alternative procedures: KS, which denotes the non-parametric Kolmogorov–Smirnov test for copula models (see Rémillard (2017)), KT-AS, which is the Kuhn-Tucker test based on the score of a symmetric Student t copula evaluated under Gaussianity (see Amengual and Sentana (2020)), and GMM, which refers to the moment test based on the underlying influence functions in GET.

Following the structure of Table 1, the first three columns of Table 3 report rejection rates under the null at the 1%, 5% and 10% levels for $n = 400$ (Panel A) and $n = 1600$ (Panel B). The results make clear that the parametric bootstrap works remarkably well for both sample sizes. In turn, the last six columns present the rejection rates at the same levels for the two alternatives. By and large, the behavior of the different test statistics is in accordance with expectations: first, when the sample size is large our proposal is the most powerful as it is designed to direct power against alternatives in which the copula follows a Hermite expansion of the Gaussian one and, second, its non-parametric competitor has close to trivial power in samples of 400 observations, a situation that improves marginally when $n = 1600$. Interestingly, the Kuhn-Tucker version of the Gaussian versus Student t copula test in Amengual and Sentana (2020) performs quite well when n is large in spite of not being designed for these alternatives. Importantly, GET does a better job than the moment test based on the influence functions S_n implied by the higher-order expansion of the log-likelihood on which it is based, which is partly due to the fact that it takes into account the partially one-sided nature of the alternatives.

Finally, it is important to mention that in this example the log-likelihood function under the alternative is particularly difficult to maximize over the five parameters involved. In fact, we systematically encounter multiple local maxima in samples of up to 100,000 observations even if we fix the correlation parameter to its true value and use global optimization methods, which forced us to repeat the calculations over a huge grid of initial values. For that reason, we have only computed the Gaussian rank correlation coefficient between the LR test and GET across ten such simulated samples, obtaining a high value of .96.

E Example 4: Purely non-linear predictive regression

E.1 The model and its log-likelihood function

Consider the following extension of the nonlinear regression model in Bottai (2003), where n observations on $\mathbf{y} = (y_1, y_2, y_3)$ are drawn from a joint distribution characterized by $f(\mathbf{y};\boldsymbol{\theta}) = f(y_3|y_1, y_2; \boldsymbol{\theta})f(y_1, y_2)$ with $f(y_1, y_2)$ fixed and known, and

$$f(y_3|y_1, y_2; \boldsymbol{\theta}) = \phi \left[y_3 - \exp(\theta_1 y_1 + \theta_2 y_2) + \theta_1 y_1 + \theta_2 y_2 + \frac{1}{2}\theta_2^2 y_2^2 \right], \tag{E48}$$

with $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ unknown. This model has an interesting interpretation in the context of predictive regressions. Specifically, a Taylor expansion of the exponential function immediately shows that the mean predictability of y_3 does not come from the terms that also enter outside the exponent (namely, y_1, y_2 and y_2^2) but rather, from higher order powers of the two regressors as well as their cross-products. Therefore, model (E48) provides an interesting functional form for predictive regressions of variables such as financial returns when a researcher believes in predictability but not through standard linear terms (see for example Spiegel (2008) and the references therein for a discussion of return predictability).

E.2 The null hypothesis and the GET test statistic

In the case of a single regressor, Bottai (2003) showed that the nullity of the information matrix is one when the regressand is unpredictable. Not surprisingly, the information matrix has several rank deficiencies under the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ in the multiple regressor case. The relevant derivatives of the log-likelihood function with respect to θ_1, θ_2 evaluated under H_0 are

$$\frac{\partial l}{\partial \theta_1} = 0, \quad \frac{\partial l}{\partial \theta_2} = 0,$$
$$\frac{\partial^2 l}{\partial \theta_1^2} = y_1^2(y_3 - 1), \quad \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} = y_1 y_2 (y_3 - 1), \quad \frac{\partial^2 l}{\partial \theta_2^2} = 0 \quad \text{and} \quad \frac{\partial^3 l}{\partial \theta_2^3} = y_2^3(y_3 - 1).$$

Therefore, we have a situation in which the degree of underidentification is different for the two regression coefficients. But since Assumption 5 is satisfied with $C = \{(2,0), (1,1), (0,3)\}$, a straightforward application of Theorem 2 implies that

$$LR_n = \text{GET}_n + O_p(n^{-\frac{1}{6}})$$
$$= \sup_{\theta_1,\theta_2} 2(\theta_1^2, \theta_1\theta_2, \theta_2^3) \begin{pmatrix} L_n^{[2,0]} \\ L_n^{[1,1]} \\ L_n^{[0,3]} \end{pmatrix} - n(\theta_1^2, \theta_1\theta_2, \theta_2^3) \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} & \mathcal{I}_{13} \\ \mathcal{I}_{21} & \mathcal{I}_{22} & \mathcal{I}_{23} \\ \mathcal{I}_{31} & \mathcal{I}_{32} & \mathcal{I}_{33} \end{pmatrix} \begin{pmatrix} \theta_1^2 \\ \theta_1\theta_2 \\ \theta_2^3 \end{pmatrix} + O_p(n^{-\frac{1}{6}}), \tag{E49}$$

where
$$\begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} & \mathcal{I}_{13} \\ \mathcal{I}_{21} & \mathcal{I}_{22} & \mathcal{I}_{23} \\ \mathcal{I}_{31} & \mathcal{I}_{32} & \mathcal{I}_{33} \end{pmatrix} = \lim_{n\to\infty} Var \left[\sqrt{n} \begin{pmatrix} l^{[2,0]} \\ l^{[1,1]} \\ l^{[0,3]} \end{pmatrix} \right].$$

In this case, though, we need to obtain the maximum with respect to θ_1 and θ_2 over the entire Euclidean space of dimension 2 rather than over the unit circle.

Nevertheless, we can provide an asymptotically equivalent but much simpler statistic. Let $p_1 = \sqrt{n}(\theta_1^{ET})^2$, $p_2 = \sqrt{n}\theta_1^{ET}\theta_2^{ET}$ and $p_3 = \sqrt{n}(\theta_2^{ET})^3$. It is then straightforward to show that

$$n^{\frac{1}{6}} p_1 p_3^{\frac{2}{3}} = p_2^2.$$

As a result, we must have that either p_1 or p_3 are negligible when n is large because p_2 is $O_p(1)$ from Lemma 1 in Appendix B. If p_1 is negligible, then (E49) is asymptotically equivalent to

$$\begin{aligned} supLM_{1n} &= \sup_{\theta_1,\theta_2} 2(\theta_1\theta_2, \theta_2^3)\begin{pmatrix} L_n^{[1,1]} \\ L_n^{[0,3]} \end{pmatrix} - n(\theta_1\theta_2, \theta_2^3)\begin{pmatrix} \mathcal{I}_{22} & \mathcal{I}_{23} \\ \mathcal{I}_{32} & \mathcal{I}_{33} \end{pmatrix}\begin{pmatrix} \theta_1\theta_2 \\ \theta_2^3 \end{pmatrix} \\ &= \frac{1}{n}(L_n^{[1,1]}, L_n^{[0,3]})\begin{pmatrix} \mathcal{I}_{22} & \mathcal{I}_{23} \\ \mathcal{I}_{32} & \mathcal{I}_{33} \end{pmatrix}^{-1}\begin{pmatrix} L_n^{[1,1]} \\ L_n^{[0,3]} \end{pmatrix}. \end{aligned}$$

If instead p_3 is negligible, then (E49) becomes asymptotically equivalent to

$$\begin{aligned} supLM_{2n} &= \sup_{\theta_1,\theta_2} 2(\theta_1^2, \theta_1\theta_2)\begin{pmatrix} L_n^{[2,0]} \\ L_n^{[1,1]} \end{pmatrix} - n(\theta_1^2, \theta_1\theta_2)\begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}\begin{pmatrix} \theta_1^2 \\ \theta_1\theta_2 \end{pmatrix} \\ &= \frac{1}{n}\left\{\frac{(L_n^{[1,1]})^2}{\mathcal{I}_{22}} + \frac{(L_n^{[2,0]} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}L_n^{[1,1]})^2}{\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21}}\mathbf{1}[L_n^{[2,0]} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}L_n^{[1,1]} > 0]\right\}. \end{aligned}$$

Consequently, we could obtain an asymptotically equivalent statistic up to a term of order $o_p(1)$ by simply retaining $\text{GET}_n = \max\{supLM_{1n}, supLM_{2n}\}$.

In addition to computational advantages, it turns out that the asymptotic distribution of our test is easy to obtain. Specifically, letting

$$Z_{1n} = n^{-\frac{1}{2}}\frac{L_n^{[2,0]} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}L_n^{[1,1]}}{\sqrt{\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21}}}, \quad Z_{2n} = n^{-\frac{1}{2}}\frac{L_n^{[1,1]}}{\sqrt{\mathcal{I}_{22}}} \quad \text{and} \quad Z_{3n} = n^{-\frac{1}{2}}\frac{L_n^{[0,3]} - \mathcal{I}_{32}\mathcal{I}_{22}^{-1}L_n^{[1,1]}}{\sqrt{\mathcal{I}_{33} - \mathcal{I}_{32}\mathcal{I}_{22}^{-1}\mathcal{I}_{23}}}, \text{ where}$$

$$\begin{pmatrix} Z_{1n} \\ Z_{2n} \\ Z_{3n} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 & r_{13} \\ 0 & 1 & 0 \\ r_{13} & 0 & 1 \end{pmatrix}\right] \text{ and}$$

$$r_{13} = \frac{\mathcal{I}_{13} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{23}}{\sqrt{\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21}}\sqrt{\mathcal{I}_{33} - \mathcal{I}_{32}\mathcal{I}_{22}^{-1}\mathcal{I}_{23}}},$$

then, $\sup LM_{1n} = Z_{2n}^2 + Z_{3n}^2$ and $\sup ML_{2n} = Z_{2n}^2 + Z_{1n}^2 1\left[Z_{1n} \geq 0\right]$. As a consequence,

$$\text{GET}_n \xrightarrow{d} \max\{Z_1^2\mathbf{1}\left\{Z_1 \geq 0\right\}, Z_3^2\} + Z_2^2.$$

That is, the asymptotic distribution of GET_n will be a χ_2^2 50% of the time (when $Z_1 < 0$) and the sum of a χ_1^2 with the largest of two other possibly dependent $\chi_1^{2\prime}s$ (when $Z_1 \geq 0$). If we further assume that the regressors y_1 and y_2 are two independent normals with 0 means and variances σ_1^2 and σ_2^2, respectively, then the $Z's$ will be three independent standard normals.

E.3 Simulation evidence

As alternatives, we consider $\theta_1 = 0.3$, $\theta_2 = 0$ (H_{a1}) and $\theta_1 = 0$, $\theta_2 = 0.5$ (H_{a2}) in model (E48). And like in the normal versus SNP example, if we maintain that y_1 and y_2 are uncorrelated, we can compute exact critical values for any sample size to any degree of accuracy by repeatedly drawing *iid* spherical normal vectors (y_1, y_2, y_3), which effectively imposes the null.

In Table 4 we compare the results of the two versions of our tests discussed above, with the GMM test mentioned at the end of section 2.2 and two simple alternative procedures. First, a standard LM test based on pseudo-Gaussian ML that checks the joint significance of y_1^2 and y_1y_2 in the OLS regression of y_3 on a constant and these two variables, which are the transformations of the predictors missing from the part outside the exponent in the conditional mean specification. And second, a closely related LM test based on pseudo-Gaussian ML which augments the previous regression with the following four cubic terms y_1^3, $y_1^2y_2$, $y_1y_2^2$ and y_2^3. We refer to these tests as OLS_1 and OLS_2, respectively.

The first three columns of Table 4 report rejection rates under the null at the 1%, 5% and 10% levels for $n = 400$ (Panel A) and $n = 1600$ (Panel B) for the first alternative hypothesis we consider while the last three do the same for the second one. Once again, the behavior of the different test statistics is in accordance with expectations. In particular, our proposed statistics are the most powerful in both cases. Part of the reason has to do with the fact that the linear regressions only provide an approximation to the true non-linear conditional expectation. However, the fraction of the theoretical variance of y_3 explained by y_1^2, y_1y_2, y_1^3, $y_1^2y_2$, $y_1y_2^2$ and y_2^3 is essentially the same as the fraction explained by the true conditional mean in H_{a2}. As a result, the superior power of our tests relative to OLS_2 comes from the reduction in degrees of freedom.

Given that in this case our test has a relatively standard asymptotic distribution –namely, a 50:50 mixture of χ_2^2 and the sum of χ_1^2 with the larger of two other independent χ_1^2's– we can also compute Davidson and MacKinnon (1998)'s p-value discrepancy plots to assess the finite sample reliability of this large sample approximation for every possible significance level. The results for the two sample sizes we consider, which are available on request, confirm the high quality of the asymptotic approximation.

Finally, our results indicate a .94-.95 Gaussian rank correlation between our proposed test statistic and the LR across Monte Carlo simulations generated under the null, which is in line with our asymptotic equivalence results in Theorem 2. At the same time, they confirm that the LR test typically takes about 200 times as much CPU time to compute as the $\max\{supLM_{1n}, supLM_{2n}\}$ version of our test.

F Relationship to the previous literature

Davies (1987) proposed perhaps the most cited sup-type test, so it is illustrative to provide a link between Theorem 1 and his results. In view of the fact that $\|\boldsymbol{\theta}_r\|$ remains irrelevant regardless of q_r, without loss of generality we consider the reparametrization $\boldsymbol{\theta}_r = \eta\boldsymbol{\upsilon}$, with $\boldsymbol{\upsilon} \in R^{q_r}$, $\|\boldsymbol{\upsilon}\| = 1$ and $\eta \geq 0$, so that η and $\boldsymbol{\upsilon}$ represent the magnitude and direction of the parameter vector $\boldsymbol{\theta}_r$, respectively. Given that

$$\sup_{\boldsymbol{\phi},\boldsymbol{\theta}_1,\|\boldsymbol{\upsilon}\|=1,\eta\geq 0} L_n(\boldsymbol{\phi},\boldsymbol{\theta}_1,\eta\boldsymbol{\upsilon}) = \sup_{\boldsymbol{\phi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_r} L_n(\boldsymbol{\phi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_r),$$

we can rewrite the null hypothesis as $H_0 : \boldsymbol{\theta}_1 = \mathbf{0}, \eta = 0$, where $\boldsymbol{\upsilon}$ is a nuisance parameter that only appears under the alternative. If we considered the r^{th} derivative of $l_i(\boldsymbol{\rho})$ along a specific direction $\boldsymbol{\upsilon}$, which would effectively coincide with the r^{th} derivative with respect to η, then we could directly apply the Lee and Chesher (1986) approach to obtain the relationship between the LR and ET tests along that direction. We then look at the supremum of those tests over all possible directions, as suggested by Davies (1987), which would effectively yield GET_n.

Nevertheless, this intuitive explanation in terms of η and $\boldsymbol{\upsilon}$ has some limitations. First, Lee and Chesher (1986) would yield a pointwise result for a given $\boldsymbol{\upsilon}$, while Theorem 1 relies on uniform convergence. More importantly, Davies (1987) method is designed for models in which the log-likelihood function is absolutely flat for some parameters under the null, so regardless of its analytic nature, no higher order derivatives will provide moments to test. In contrast, we consider situations in which the log-likelihood function written in terms of $\boldsymbol{\theta}$ only has a finite number of zero derivatives, so a test statistic can be based on the first round of non-zero ones. In this respect, the underidentification of $\boldsymbol{\upsilon}$ is an artifact of the $\boldsymbol{\theta}_r = \eta\boldsymbol{\upsilon}$ reparametrization that would persist even if the information matrix had full rank, in which case the supremum over $\boldsymbol{\upsilon}$ of the test of $H_0 : \theta_1 = 0, \eta = 0$ will yield the usual LM test. In any event, in Theorems 2 and 3 we derive generalized extremum tests in more general contexts without resorting to any such reparametrization.

References

Amengual, D. and Sentana, E. (2015): "Is a normal copula the right copula?", CEMFI Working Paper 1504.

Amengual, D. and Sentana, E. (2020): "Is a normal copula the right copula?", *Journal of Business and Economic and Statistics* 38, 350-366.

Constantine, G.M. and Savits, T.H. (1996): "A multivariate Faà di Bruno formula with applications", *Transactions of the American Mathematical Society* 348, 503–520.

Davidson, R., and MacKinnon, J. G. (1998): "Graphical methods for investigating the size and power of hypothesis tests", *The Manchester School* 66, 1–26.

Davies, R.B. (1987): "Hypothesis testing when a nuisance parameter is present only under the alternatives", Biometrika 74, 33–43.

Fan, Y. and Patton, A. J. (2014): "Copulas in econometrics", *Annual Review of Economics* 6, 179–200.

Faà di Bruno, F. (1859): *Théorie générale de l'élimination*, De Leiber & Faraquet.

Rémillard, B. (2017): "Goodness-of-fit tests for copulas of multivariate time series", *Econometrics* 5(1), 13.

Spiegel, M. (2008): "Forecasting the equity premium: where we stand today", *Review of Financial Studies* 24, 1453–1454.

Table 3: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the Gaussian versus Hermite expansion copula test

	Null hypothesis			Alternative hypotheses					
				H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel A: $n = 400$									
GET	1.1	5.1	10.2	18.4	49.7	65.1	26.9	60.9	74.2
KS	0.9	4.7	9.3	0.9	4.7	9.9	1.1	5.4	10.6
KT-AS	1.2	5.3	10.3	18.9	39.2	52.0	31.7	55.4	68.0
GMM	1.1	5.2	10.2	3.8	38.4	57.0	6.3	49.7	67.2
Panel B: $n = 1600$									
GET	0.9	4.9	10.3	90.8	98.9	99.6	96.8	99.7	99.9
KS	0.9	4.7	9.8	1.9	7.7	14.5	3.1	10.4	18.6
KT-AS	0.9	5.3	10.6	60.9	82.8	90.1	87.1	95.9	98.2
GMM	1.1	5.0	9.9	44.0	95.5	99.0	68.2	98.8	99.7

Notes: Results based on 10,000 samples. Margins are assumed to be known. GET, KS, KT-AS and GMM are defined in Supplemental Appendix D. Finite sample critical values are computed using the parametric bootstrap. DGPs: The correlation parameter φ is set to 0.5 under both the null and alternative hypotheses. As for H_{a_1} and H_{a_2} correspond to pure, fourth-order Hermite expansion copulas with $\boldsymbol{\vartheta}' = (0.03, 0, 0, 0, 0)$ and $\boldsymbol{\vartheta}' = (0.02, 0, 0, 0, 0.02)$, respectively.

Table 4: Monte Carlo rejection rates (in %) under alternative hypotheses for white noise versus a purely nonlinear regression test

	Alternative hypotheses					
	H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%
Panel A: $n = 400$						
GET	19.5	41.3	54.4	18.5	39.7	52.4
LR	21.7	41.7	56.2	20.5	40.4	54.1
GMM	15.3	34.3	47.0	14.3	33.4	45.5
OLS_1	16.2	34.6	47.2	12.9	30.5	41.9
OLS_2	9.6	23.9	37.0	7.3	20.2	32.4
Panel B: $n = 1600$						
GET	65.5	83.9	90.2	61.3	80.5	87.6
LR	66.3	84.5	91.2	61.9	81.5	88.5
GMM	57.6	78.3	86.0	54.3	75.2	83.6
OLS_1	53.2	74.1	83.3	42.7	64.6	75.1
OLS_2	37.7	61.6	73.3	25.7	48.8	61.8

Notes: Results based on 10,000 samples. GET, LR, GMM, OLS_1 and OLS_2 are defined in Supplemental Appendix E. Finite sample critical values are computed by simulation. DGPs: $(y_1, y_2) \sim iid\ N(\mathbf{0}, \mathbf{I}_2)$ under both alternative hypotheses, with $\theta_1 = 0.25$ and $\theta_2 = 0.25$ (H_{a1}), and $\theta_1 = 0.3$ and $\theta_2 = 0.1$ (H_{a2}).