Hypothesis tests with a repeatedly singular information matrix*

Dante Amengual
CEMFI
<amengual@cemfi.es>

Xinyue Bei
Duke University
<xinyue.bei@duke.edu>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

Revised: May 2023

Abstract

We study score-type tests in likelihood contexts in which the nullity of the information matrix under the null is greater than one, thereby generalizing existing results in the literature. Examples include multivariate regressions with sample selectivity, semi-nonparametric distributions, Hermite expansions of Gaussian copulas, and purely non-linear predictive regressions among others. Our proposal, which involves higher-order derivatives, is asymptotically equivalent to the likelihood ratio test but only requires estimation under the null, a substantial advantage for resampling-based inference. We conduct extensive Monte Carlo exercises to study the finite sample size and power properties of our proposal, comparing it to alternative approaches.

Keywords: Generalized extremum tests, Higher-order identifiability, Likelihood ratio test, Non-Gaussian copulas, Predictive regressions, Selectivity, Semi non-parametric distributions.

JEL: C12, C34, C46, C58.

*We would like to thank Tincho Almuzara, Manuel Arellano, Eric Renault and Andrea Rotnitzky for useful comments and suggestions, as well as audiences at Banco Central de Chile, Boston College, Boston University, CEMFI, Duke University, Harbin Institute of Technology, Harvard, NYU, Universidad de la República, Università di Firenze, Université de Montréal, Universidad Torcuato Di Tella, the Econometric Society 2019 China Meeting (Guangzhou) and its 2020 World Congress, the 2019 Galatina Summer Meetings, the VIII Encuentro Anual de la SEU (Montevideo) and the 2023 TSE Financial Econometrics Conference. Of course, the usual caveat applies. Amengual and Sentana gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00.

1 Introduction

Rao's (1948) score test and Silvey's (1959) numerically equivalent Lagrange multiplier (LM) version completed the triad of classical hypothesis tests (see Bera and Bilias (2001) for a survey). Given that they only require estimation of the model parameters under the null, in the late 1970's and early 1980's they became the preferred choice for many specification tests which are nowadays routinely reported by econometric software packages (see the surveys by Breusch and Pagan (1980), Engle (1983), and Godfrey (1988)). In addition to computational considerations, which continue to be very relevant for resampling procedures, two other important advantages of LM tests are that (i) rejections provide a clear indication of the specific directions along which modelling efforts should focus, and (ii) they are often easy to interpret as moment tests, so they remain informative for alternatives they are not designed for. Furthermore, under standard regularity conditions, they are asymptotically equivalent to the Likelihood ratio (LR) and Wald tests under the null and sequences of local alternatives, and thus they share their optimality properties.

One of the crucial regularity conditions for a common asymptotic chi-square distribution for these three tests is a full rank information matrix of the unrestricted model parameters evaluated under the null. Nevertheless, there are empirically relevant situations in which this condition does not hold despite the fact that the model parameters are locally identified. In non-linear instrumental variable models, Sargan (1983) referred to those instances in which the expected Jacobian of the influence functions is singular but the expected Jacobian of the linear combinations of their derivatives that span its nullspace has full rank as second-order identified but first-order underidentified. In a likelihood context, a singular information matrix implies that there is a linear combination of the average scores which is identically 0, at least asymptotically. In their seminal paper, Lee and Chesher (1986) studied some popular examples of this situation in economics: i) univariate regression models with sample selectivity; ii) stochastic production frontier models; and iii) certain mixture models.[1]

Lee and Chesher (1986) proposed to replace the LM test by what they called an "extremum" test. Their suggestion was to study the restrictions that the null imposes on higher-order optimality conditions. Often, the second derivative will suffice, but sometimes it might be necessary to study the third or even higher-order ones. They proved the asymptotic equivalence between their extremum tests and the corresponding LR tests under the null and sequences of local alternatives in unrestricted contexts. Using earlier results by Cox and Hinkley (1974), this equivalence

[1]In all their examples, in fact, the average score with respect to one of the parameters of the alternative evaluated at the restricted parameter estimators that impose the null is identically 0 in finite samples.

intuitively follows from the fact that their tests can often be re-interpreted as standard LM tests of a suitable transformation of the parameter whose first derivative is 0 on average such that the new score is no longer so. In contrast, Wald tests are extremely sensitive to reparametrization under these circumstances. Bera et al (1998) provide some additional insights. In turn, Rotnitzky el al (2000) rigorously study the asymptotic distribution of the maximum likelihood (ML) estimators in those contexts. Finally, Bottai (2003) looks at the validity of confidence intervals obtained by inverting the three classical test statistics in this setup.

However, in all the existing literature the nullity of the information matrix, q_r say, is assumed to be 1. When the information matrix is repeatedly singular under the null, in the sense that q_r is two or more, the number of second-order derivatives exceeds the number of parameters effectively affected by the singularity by an order of magnitude. The unbalance gets worse when it becomes necessary to look at higher-order derivatives. Unfortunately, in general there is no reparametrization that leads to a regular information matrix.[2] In particular, transforming each of the parameters individually along the lines suggested by Lee and Chesher (1986) does not usually give rise to a test asymptotically equivalent to the LR. On the contrary, different reparametrizations will typically give rise to different test statistics.

The purpose of our paper is precisely to propose a feasible generalization of the Lee and Chesher (1986) approach in repeatedly singular contexts that leads to tests asymptotically equivalent to the LR, but which only require estimation under the null. Specifically, we propose a generalized extremum test (GET) which typically maximizes an easy to interpret statistic over a space of dimension $q_r - 1$ when all parameters show the same degree of underidentification, thereby simplifying to the Lee and Chesher (1986) proposal when the nullity is one. More generally, GET is an LR-type test that compares the log-likelihood function under the null to the maximum over q_r dimensions of its lowest-order expansion under the alternative capable of identifying the restricted parameters. In contrast, LR tests require the maximization over the entire parameter space of an unrestricted log-likelihood function which is extremely flat around its maximum when the null hypothesis is true.[3] These computational advantages are particularly pertinent for bootstrap-type inference, which is especially necessary in our context because the common sup-type asymptotic distribution of the GET and LR tests is often non-standard, and the sample sizes required for this distribution to be reliable unusually large.

Repeatedly singular information matrices are not a mere theoretical curiosity. In fact, we illustrate our proposed testing procedure in detail with several examples of interest that arise in

[2] An exception is the multiplicative seasonal ARMA model considered in Amengual, Bei and Sentana (2023).

[3] Obviously, both procedures require the estimation of the model under the null, but the restricted maximum likelihood estimator is typically available in closed form in many models subject to specification tests.

economic and finance applications when testing: 1) exogeneous sample selectivity in multivariate regressions; 2) normality against the flexible semi-nonparametric (SNP) family proposed by Gallant and Nychka (1987); 3) a Gaussian copula against another flexible Hermite expansion; and 4) unpredictability in a multiple regressor version of the purely non-linear model considered by Bottai (2003). Further, in Amengual, Bei and Sentana (2022, 2023) we discuss the application of the test proposed in this paper to two additional examples of substantial empirical interest: testing for multivariate normality against a skew normal distribution, and testing for neglected serial correlation in univariate time series models, respectively.

The structure of the rest of the paper is as follows. In section 2 we obtain our theoretical results first in the case in which all the underidentified parameters have the same degree of underidentification, and then when the degree of underidentification may be different for different parameters. Then, in section 3 we discuss the first two aforementioned examples in detail, assessing the finite sample size and power properties of our proposed tests by means of several extensive Monte Carlo exercises. Finally, we conclude in section 4, relegating proofs, the remaining two examples, and some additional results to the appendices.

2 Theoretical results

Consider the estimation of the parameter vector $\boldsymbol{\rho}$ characterizing the distribution of an $i.i.d.$ random vector $\mathbf{y}$. Let $l_i(\boldsymbol{\rho}) = \ln f(\mathbf{y}_i; \boldsymbol{\rho})$ denote the log-likelihood function contribution from observation i, so that the log-likelihood function of a sample of size n is $\mathcal{L}_n = \sum_{i=1}^{n} l_i(\boldsymbol{\rho})$.[4] In what follows,

$$s_{\rho_j i}(\boldsymbol{\rho}) = \frac{\partial l_i(\boldsymbol{\rho})}{\partial \rho_j}$$

will denote the contribution of observation i to the score with respect to the j^{th} element of $\boldsymbol{\rho}$ and $\mathbf{S}_{\rho_j n}(\boldsymbol{\rho}) = \sum_{i=1}^{n} \mathbf{s}_{\rho_j i}(\boldsymbol{\rho})$ their sum.

Let us partition $\boldsymbol{\rho}$ into two blocks: 1) $\boldsymbol{\phi}$, which contains the $p \times 1$ vector of parameters estimated under the null; and 2) $\boldsymbol{\theta}$, which is the $q \times 1$ vector of parameters such that the null hypothesis can be written in explicit form as $H_0 : \boldsymbol{\theta} = \mathbf{0}$. Let $\boldsymbol{\rho}^*$, $\hat{\boldsymbol{\rho}}$ and $\tilde{\boldsymbol{\rho}} = (\tilde{\boldsymbol{\phi}}', \mathbf{0}')'$ denote the true value of the parameter vector, its unrestricted ML estimator (UMLE), and the restricted one (RMLE), respectively, so that $\boldsymbol{\rho}^* = (\boldsymbol{\phi}^*, \mathbf{0})$ under H_0. As usual, $|.|$ and $||.||$ denote absolute value and Euclidean norm, respectively. Finally, we use $e_{\min}(\boldsymbol{A})$ and $e_{\max}(\boldsymbol{A})$ for the smallest and largest eigenvalues, respectively, of a symmetric square matrix $\boldsymbol{A}$.

Using this notation, we henceforth assume:

[4] Although we could easily generalize our results to explicitly deal with dependent data by using standard factorizations of the sample log-likelihood function, we maintain independence to simplify the expressions.

Assumption 1 *(Regularity conditions)*
(1.1) $\boldsymbol{\rho}$ *takes its value in a compact subset* $\mathbf{P}$ *of* $\mathbb{R}^{p+q}$ *that contains an open neighborhood* $\mathcal{N}$ *of the true value* $\boldsymbol{\rho}^*$ *which generates the observations.*
(1.2) Distinct values of $\boldsymbol{\rho}$ *in* $\mathbf{P}$ *correspond to distinct probability distributions.*
(1.3) $E[\sup_{\boldsymbol{\rho}\in\mathbf{P}} |l_i(\boldsymbol{\rho})|] < \infty$.
(1.4) $E[\partial l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\phi}\cdot\partial l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\phi}']$ *has full rank under the null for all* $(\boldsymbol{\phi},\mathbf{0}) \in \mathbf{P}$.

The compactness of $\mathbf{P}$ in Assumption 1.1 together with the continuity of $l_i(\boldsymbol{\rho})$ and Assumptions 1.2 and 1.3 guarantee the existence, uniqueness with probability tending to 1, and consistency of both the UMLE $\hat{\boldsymbol{\rho}}$ and the RMLE $\tilde{\boldsymbol{\rho}}$ (see Newey and McFadden 1994, Theorem 2.5). The "open neighborhood" part of Assumption 1.1 is just used to simplify the expressions and their derivation. Extensions to situations in which the true parameters lie at the boundary of the parameter space under the null are feasible, as we will show in Supplemental Appendix C, but at the expense of complicating the notation and blurring the message of the paper. Finally, Assumption 1.4 guarantees the convergence of the RMLE at the usual $n^{-\frac{1}{2}}$ rate.

2.1 Repeated singularity of the same order

We first consider the case in which q_1 elements of $\boldsymbol{\theta}$ are first-order identified, while the remaining q_r elements are r^{th}-order identified under the null, a concept that will become precisely defined after we introduce Assumption 3 below. Therefore, if we further partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_r')'$, where $q_1 = \dim(\boldsymbol{\theta}_1)$ and $q_r = \dim(\boldsymbol{\theta}_r)$, so that $q = q_1 + q_r$, then the information matrix under H_0 will be such that its top $(p+q_1)\times(p+q_1)$ block is regular and the rest contains zeros. Consequently, its nullity will be precisely q_r. Often, one needs to reparametrize the model to make sure it satisfies these conditions, an issue we discuss in detail in Supplemental Appendix B.1 in general terms, as well as in each of the examples that we consider.

Let $\mathbf{j} \in \mathbb{N}^{p+q}$ denote a $(p+q)\times 1$ vector of indices, $\mathbf{j}! = \prod_{i=1}^{p+q} j_i!$,

$$l_i^{[\mathbf{j}]}(\boldsymbol{\rho}) = \frac{1}{\mathbf{j}!}\frac{\partial^{\boldsymbol{\iota}'_{p+q}\mathbf{j}} l_i(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}^{\mathbf{j}}},$$

where $\boldsymbol{\iota}_m$ is a vector of m ones, and $L_n^{[\mathbf{j}]}(\boldsymbol{\rho}) = \sum_{i=1}^n l_i^{[\mathbf{j}]}(\boldsymbol{\rho})$. Throughout this subsection, we assume the following conditions hold:

Assumption 2 *(Regularity conditions on the derivatives of the log-likelihood function)*
(2.1) With probability 1, the derivatives $l_i^{[\mathbf{j}]}(\boldsymbol{\rho})$ *exist for all* $\boldsymbol{\rho}$ *in* $\mathcal{N}$ *and* $\boldsymbol{\iota}'_{p+q}\mathbf{j} \le 2r$, *and they satisfy* $E[\sup_{\boldsymbol{\rho}\in\mathcal{N}} |l_i^{[\mathbf{j}]}(\boldsymbol{\rho})|] < \infty$.
(2.2) For $r \le \boldsymbol{\iota}'_{p+q}\mathbf{j} \le 2r$, $E\{[l_i^{[\mathbf{j}]}(\boldsymbol{\rho})]^2\} < \infty$ *for all* $\boldsymbol{\rho}$ *in* $\mathcal{N}$.
(2.3) When $\boldsymbol{\iota}'_{p+q}\mathbf{j} = 2r$ *there is some function* $g(\mathbf{y})$ *satisfying* $E[g^2(\mathbf{y})] < \infty$ *such that with probability 1,* $|L_n^{[\mathbf{j}]}(\boldsymbol{\rho}) - L_n^{[\mathbf{j}]}(\boldsymbol{\rho}^\dagger)| \le ||\boldsymbol{\rho} - \boldsymbol{\rho}^\dagger|| \sum_i g(\mathbf{y}_i)$ *for all* $\boldsymbol{\rho}$ *and* $\boldsymbol{\rho}^\dagger$ *in* $\mathcal{N}$.

We borrow Assumptions 2.1–2.3 from Rotnitzky et al. (2000) with some modifications. The main difference is that they require $(2r+1)^{th}$ differentiability for the Taylor expansions they use to analyze the distribution of the MLE, while we only need $2r^{th}$ differentiability to study the asymptotic distribution of our tests. Assumptions 2.1 and 2.3 guarantee the existence of derivatives and the stochastic equicontinuity of the sample mean of $l_i^{[\mathbf{j}]}(\boldsymbol{\rho})$ with $\boldsymbol{\iota}'_{p+q}\mathbf{j} \leq 2r$. In turn, Assumption 2.2 allows us to apply a central limit theorem to $l_i^{[\mathbf{j}]}(\boldsymbol{\rho}^*)$.

Let $\boldsymbol{\theta}_r^{\otimes k} = \underbrace{\boldsymbol{\theta}_r \otimes \boldsymbol{\theta}_r \otimes \cdots \otimes \boldsymbol{\theta}_r}_{k \text{ times}}$ denote the k^{th}-order Kronecker power of the $q_r \times 1$ vector $\boldsymbol{\theta}_r$, and define

$$\frac{\partial^k L_n(\boldsymbol{\rho})}{\partial \boldsymbol{\theta}_r^{\otimes k}} = vec\left\{\frac{\partial}{\partial \boldsymbol{\theta}_r}\left[\frac{\partial^{k-1} L_n(\boldsymbol{\rho})}{\partial \boldsymbol{\theta}_r^{\otimes (k-1)}}\right]'\right\}.$$

Moreover, let

$$I(\boldsymbol{\phi}) = \begin{bmatrix} I_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\phi}\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\phi}\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} = \lim_{n\to\infty} Var\left\{\frac{1}{\sqrt{n}}\left[\begin{array}{c} \mathbf{S}_{\boldsymbol{\phi}n}(\boldsymbol{\phi},\mathbf{0}) \\ \mathbf{S}_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi},\mathbf{0}) \\ \partial^r L_n(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}_r^{\otimes r} \end{array}\right]\right| \boldsymbol{\phi},\mathbf{0}\right\}$$

denote the asymptotic covariance matrix of the relevant influence functions, which may be understood as a generalization of the information matrix.

In addition, let

$$V_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) = \begin{bmatrix} V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} - \begin{bmatrix} I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) \end{bmatrix} I_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi})\left[I_{\boldsymbol{\phi}\boldsymbol{\theta}_1}(\boldsymbol{\phi})\ I_{\boldsymbol{\phi}\boldsymbol{\theta}_r}(\boldsymbol{\phi})\right]$$

denote the asymptotic residual variance of $\mathbf{S}_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi},\mathbf{0})$ and $\partial^r L_n(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}_r^{\otimes r}$ after orthogonalizing these influence functions with respect to $\mathbf{s}_{\boldsymbol{\phi}}$.

Assumption 3 *(Rank conditions for $q_r \geq 1$)*
(3.1) For all $(\boldsymbol{\phi},\mathbf{0}) \in \mathbf{P}$,

$$\frac{\partial^{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r}} l_i(\boldsymbol{\phi},\mathbf{0})}{\partial \boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = \mathbf{0}$$

with probability 1 for all $\mathbf{j}_{\boldsymbol{\theta}_r} = (j_1, ..., j_{q_r})'$ such that $\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r} \leq r-1$.
(3.2) The asymptotic covariance matrix of the (scaled by $\sqrt{n}$) sample averages of

$$\left\{\mathbf{s}_{\boldsymbol{\phi}i}(\boldsymbol{\phi}^*,\mathbf{0}), \mathbf{s}_{\boldsymbol{\theta}_1 i}(\boldsymbol{\phi}^*,\mathbf{0}), \boldsymbol{\theta}_r^{\otimes r\prime}\frac{\partial^r l_i(\boldsymbol{\phi}^*,\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}}\right\}$$

has full rank for all possible non-zero values of $\boldsymbol{\theta}_r \in \mathbb{R}^{q_r}$ underlying the vector of coefficients $\boldsymbol{\theta}_r^{\otimes r}$ in the linear combination above.

Intuitively, the rationale for looking at

$$\boldsymbol{\theta}_r^{\otimes r\prime}\frac{\partial^r l_i}{\partial\boldsymbol{\theta}_r^{\otimes r}} = \sum_{\iota'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r}=r}\frac{r!}{\mathbf{j}_{\boldsymbol{\theta}_r}!}\left(\prod_{k=1}^{q_r}\theta_{rk}^{j_k}\right)\frac{\partial^r l_i(\boldsymbol{\phi},\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}}$$

is that it coincides with the r^{th}-order term in the expansion of the log-likelihood function. In that respect, note that although the higher order derivatives $\partial^r l_i/\partial\boldsymbol{\theta}_r^{\otimes r}$ will usually contain many repeated elements thanks to the Clairaut-Schwartz-Young's theorem, the rank deficiency condition in Assumption 3.2 applies to the inner product of $\boldsymbol{\theta}_r^{\otimes r}$ with those influence functions, so the requirement is that those linear combinations of the elements in $\partial^r l_i/\partial\boldsymbol{\theta}_r^{\otimes r}$ be linearly independent of $\mathbf{s}_{\boldsymbol{\phi}i}(\boldsymbol{\phi},\mathbf{0})$ and $\mathbf{s}_{\boldsymbol{\theta}_1 i}(\boldsymbol{\phi},\mathbf{0})$.

Finally, let

$$Q_n(\boldsymbol{\theta}_r,\boldsymbol{\phi})=\frac{\boldsymbol{\theta}_r^{\otimes r\prime}D_{rn}(\boldsymbol{\phi})D_{rn}'(\boldsymbol{\phi})\boldsymbol{\theta}_r^{\otimes r}}{\boldsymbol{\theta}_r^{\otimes r\prime}[V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi})-V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi})]\boldsymbol{\theta}_r^{\otimes r}},\tag{1}$$

where

$$D_{rn}(\boldsymbol{\phi})=\frac{\partial^r L_n(\boldsymbol{\phi},\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}}-V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}(\boldsymbol{\phi})S_{\boldsymbol{\theta}_1 n}(\boldsymbol{\phi},\mathbf{0})$$

is the residual in the least squares projection of $\partial^r L_n(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}_r^{\otimes r}$ onto the linear span of $S_{\boldsymbol{\theta}_{1n}}(\boldsymbol{\phi},\mathbf{0})$.[5] In this context, we can proof the following result:

Theorem 1 *If Assumptions 1, 2 and 3 hold, then under* $H_0:\boldsymbol{\theta}=\mathbf{0}$

$$LR_n=2\,[L_n(\hat{\boldsymbol{\rho}})-L_n(\tilde{\boldsymbol{\rho}})]=GET_n+O_p(n^{-\frac{1}{2r}}),$$

where

$$GET_n=\frac{1}{n}S_{\boldsymbol{\theta}_1 n}'(\tilde{\boldsymbol{\phi}},\mathbf{0})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}(\tilde{\boldsymbol{\phi}})S_{\boldsymbol{\theta}_{1n}}(\tilde{\boldsymbol{\phi}},\mathbf{0})+\frac{1}{n}\sup_{\boldsymbol{\theta}_r\neq 0}\begin{cases}Q_n(\boldsymbol{\theta}_r,\tilde{\boldsymbol{\phi}}) & \text{if } r \text{ is odd,}\\ Q_n(\boldsymbol{\theta}_r,\tilde{\boldsymbol{\phi}})\mathbf{1}[\boldsymbol{\theta}_r^{\otimes r\prime}D_{rn}(\tilde{\boldsymbol{\phi}})\geq 0] & \text{if } r \text{ is even.}\end{cases}$$

An important implication of Theorem 1 is that the rate of convergence of the difference between the LR and GET tests is inversely proportional to the order of identification, thereby generalizing the standard result for regular models.

Importantly, expression (1), which can be understood as a generalized Rayleigh quotient evaluated at the restricted $q_r^r\times 1$ vector $\boldsymbol{\theta}_r^{\otimes r}$, does not effectively depend on $\boldsymbol{\theta}_r$ when the nullity of the information matrix is 1, so Theorem 1 generalizes the results in Lee and Chesher (1986) and Rotnitzky et al. (2000) by allowing for the presence of multiple singularities under the null (see Supplemental Appendix E for further comparisons to the existing literature).

2.2 Repeated singularity of different orders

There are situations in which the degree of identification of the different elements of $\boldsymbol{\theta}$ under the null hypothesis is more heterogeneous than just either one or $r+1$. To characterise them in full, we need to generalize the conditions in Assumptions 2 and 3. Let $\boldsymbol{\varsigma}_{\boldsymbol{\phi}i}(\boldsymbol{\phi})$ and $\boldsymbol{\varsigma}_{\boldsymbol{\theta}i}(\boldsymbol{\phi})$ denote two measurable functions of dimensions $p\times 1$ and $m\times 1$, respectively, so that we can define the empirical process

$$\mathcal{S}_n(\boldsymbol{\phi})=\begin{bmatrix}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi})\\ \mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi})\end{bmatrix}=\sum_{i=1}^n\boldsymbol{\varsigma}_i(\boldsymbol{\phi}),\quad\text{where}\quad\boldsymbol{\varsigma}_i(\boldsymbol{\phi})=\begin{bmatrix}\boldsymbol{\varsigma}_{\boldsymbol{\phi}i}(\boldsymbol{\phi})\\ \boldsymbol{\varsigma}_{\boldsymbol{\theta}i}(\boldsymbol{\phi})\end{bmatrix}.$$

[5] Importantly, Assumption 3.2 guarantees that the denominator of $Q_n(\boldsymbol{\theta}_r,\boldsymbol{\phi})$ is positive because $V_{\boldsymbol{\theta\theta}}$ is the covariance matrix of the residuals from the least squares projection of $s_{\boldsymbol{\theta}_1}(\boldsymbol{\phi},\mathbf{0})$ and $\frac{\partial^r l(\boldsymbol{\phi},\mathbf{0})}{\partial\otimes^r\boldsymbol{\theta}_r}$ on the linear span of $s_{\boldsymbol{\phi}}(\boldsymbol{\phi},\mathbf{0})$, while $V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}-V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}$ is the residual covariance matrix of the projection of the second residual on the span of the first one, which by the Frisch-Waugh theorem coincides with the residual in the projection of $\frac{\partial^r l(\boldsymbol{\phi},\mathbf{0})}{\partial\otimes^r\boldsymbol{\theta}_r}$ onto the linear span of $s_{\boldsymbol{\phi}}(\boldsymbol{\phi},\mathbf{0})$ and $s_{\boldsymbol{\theta}_1}(\boldsymbol{\phi},\mathbf{0})$.

Typically, $\boldsymbol{\varsigma}_{\boldsymbol{\phi}i}(\boldsymbol{\phi})$ coincides with the scores with respect to $\boldsymbol{\phi}$, and $\boldsymbol{\varsigma}_{\boldsymbol{\theta}i}(\boldsymbol{\phi})$ with some higher-order derivatives with respect to the elements $\boldsymbol{\theta}$, so that $\mathcal{S}_n$ will serve as the analog to the sample score in regular models. In addition, let

$$\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) = \left[\begin{array}{c} (\boldsymbol{\phi}-\boldsymbol{\phi}^*)+\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \\ \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \end{array}\right],$$

where $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \in \mathbb{R}^p$ and $\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \in \mathbb{R}^m$ are non-random vector functions of the parameters that adequately capture their difference from the true values. Finally, let

$$\mathcal{I}(\boldsymbol{\phi}) = \left[\begin{array}{cc} \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}) \\ \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) \end{array}\right]$$

denote a non-random positive semidefinite symmetric $(p+m)\times(p+m)$ matrix, which once again will effectively play the role of an information matrix.

Using this notation, we state the following assumptions, many of which are simplified versions of the conditions in Assumption 5 in Meitz and Saikkonen (2021):

Assumption 4 *(LQ approximation) L_n has a "linear-quadratic" expansion given by*

$$L_n(\boldsymbol{\phi},\boldsymbol{\theta}) - L(\boldsymbol{\phi}^*,\mathbf{0}) = \mathcal{S}_n(\boldsymbol{\phi}^*)'\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) - \frac{1}{2}n\boldsymbol{\lambda}'(\boldsymbol{\phi},\boldsymbol{\theta})\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) + R_n(\boldsymbol{\phi},\boldsymbol{\theta}),$$

where $R_n(\boldsymbol{\phi},\boldsymbol{\theta})$ is a remainder term. In addition:
(4.1) $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$ is continuous in $\boldsymbol{\rho}$, and such that (i) $\boldsymbol{\lambda}(\boldsymbol{\phi}^*,\mathbf{0})=\mathbf{0}$ and (ii) for all $\epsilon>0$,

$$\inf_{\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\geq\epsilon} \|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\| \geq \delta_\epsilon \text{ for some } \delta_\epsilon>0.$$

(4.2) $n^{-\frac{1}{2}}\mathcal{S}_n \xrightarrow{d} \mathcal{S}$ for some zero-mean $\mathbb{R}^{p+m}$-valued Gaussian process with covariance kernel

$$E\left[\mathcal{S}(\boldsymbol{\phi}_1)\mathcal{S}'(\boldsymbol{\phi}_2)\right] = E\left[\boldsymbol{\varsigma}_i(\boldsymbol{\phi}_1)\boldsymbol{\varsigma}_i'(\boldsymbol{\phi}_2)\right] = \mathcal{K}(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2).$$

(4.3) $\mathcal{I}(\boldsymbol{\phi}^*)=\mathcal{K}(\boldsymbol{\phi}^*,\boldsymbol{\phi}^*)$ is Lipschitz continuous at a neighborhood of $\boldsymbol{\phi}^*$ and satisfies

$$0<e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]<e_{\max}[\mathcal{I}(\boldsymbol{\phi}^*)]<\infty.$$

(4.4) The remainder term $R_n(\boldsymbol{\phi},\boldsymbol{\theta})$ satisfies

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\boldsymbol{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n} \frac{|R_n(\boldsymbol{\phi},\boldsymbol{\theta})|}{1+n\|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\|^2} = o_p(1)$$

for all sequences of (non-random) positive scalars $\{\gamma_n : n\geq 1\}$ for which $\gamma_n\to 0$ as $n\to\infty$.
(4.5) There exists some function $\mathbf{g}(\mathbf{y})$ satisfying $E\left[(\mathbf{g}(\mathbf{y}_i))^2\right]<\infty$ such that

$$||\mathcal{S}(\boldsymbol{\phi}^\dagger)-\mathcal{S}(\boldsymbol{\phi}^*)|| \leq ||\boldsymbol{\phi}^\dagger-\boldsymbol{\phi}^*||\sum_{i=1}^n \mathbf{g}(\mathbf{y}_i) \tag{2}$$

with probability 1 for all $(\boldsymbol{\phi},\mathbf{0})\in\mathcal{N}$.
(4.6) If $n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi}_n,\boldsymbol{\theta}_n)=O(1)$, then $R_n(\boldsymbol{\phi},\boldsymbol{\theta})=O_p(n^{-a})$ for some a such that $\frac{1}{2}\geq a>0$.

Assumption 4 states that the likelihood ratio can be expressed as the sum of a linear-quadratic approximation and a residual term, R_n. The linear-quadratic part, though, represents a higher-order expansion of the likelihood ratio around $\boldsymbol{\theta} = \mathbf{0}$. Assumption 4.1 captures the local identification condition at the true parameter value. Assumption 4.2 is analogous to the information matrix equality, while Assumption 4.3 to the standard non-singular information matrix assumption. In turn, Assumption 4.4 ensures that the residual is dominated by the leading terms, and thus, negligible asymptotically, while Assumption 4.5 enables us to substitute the true parameter $\boldsymbol{\phi}^*$ with the restricted estimator $\tilde{\boldsymbol{\phi}}$ after an appropriate adjustment for sampling variability. Finally, Assumption 4.6 allows us to obtain the convergence rate of the linear-quadratic approximation, with a typically associated to the slowest rate of convergence of the parameter estimators under the null.

We can then prove the following result:

Theorem 2 *If Assumptions 1 and 4.1– 4.5 hold, then under* $H_0 : \boldsymbol{\theta} = \mathbf{0}$

$$LR = 2[L_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) - L_n(\tilde{\boldsymbol{\phi}}, \mathbf{0})] = GET_n + o_p(1),$$

where

$$\begin{aligned} GET_n &= \sup_{\boldsymbol{\theta}} \{2[\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &\qquad -n\boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\}. \end{aligned}$$

If, in addition, Assumption 4.6 holds, then

$$LR = 2[L_n(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) - L(\tilde{\boldsymbol{\phi}}, \mathbf{0})] = GET_n + O_p(n^{-a}).$$

As expected, we can easily show that our first theorem is a special case of this second theorem when the higher-order identification is of the same order for all the parameters involved regardless of the parity of r. More importantly, the proof of this theorem shows that we can interpret $L_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) + GET_n$ as a Taylor approximation of order $2r$ to the log-likelihood function around $\tilde{\boldsymbol{\rho}}$, which means that GET_n is effectively an LR-type test that compares the log-likelihood function under the null to the maximum of its lowest-order approximation under the alternative capable of identifying the restricted parameters.

Although GET cannot be directly understood as a moment test, a by-product of our most general theorem is a set of influence functions $S_n(\boldsymbol{\phi}, \mathbf{0})$ that can be used for that purpose after taking into account the sampling uncertainty in estimating $\boldsymbol{\phi}$ under the null. In fact, we can prove that this moment test, which converges in distribution to a χ^2_m under the null, where $m = dim[\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})]$, provides an upper bound to GET_n, albeit a rather loose one in many cases.

2.3 Distribution under local alternatives

Let us now consider the distribution of the test statistic under the following sequences of local alternatives:

$$H_{1n} : \sqrt{n} \begin{bmatrix} \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_n) \\ \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}_{\boldsymbol{\phi},\infty} \\ \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty} \end{bmatrix} = \boldsymbol{\lambda}_\infty \in \mathbb{R}^{\dim(\boldsymbol{\lambda}_{\boldsymbol{\theta}})}.$$

To do so, we need to assume that

Assumption 5 *(Cone cover) The sequence of sets*

$$\Lambda_n = \left\{ \sqrt{n}\boldsymbol{\lambda}_\theta(\boldsymbol{\theta}_n) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$$

covers a closed cone $\Lambda \subseteq \mathbb{R}^{dim(\boldsymbol{\lambda}_\theta)}$ *(with* $\Lambda_n \to \Lambda$*) so that there is a sequence of closed balls* B_{k_n} *of radius* $k_n \to \infty$ *centered at the origin such that* $\Lambda_n \cap B_{k_n} = \Lambda \cap B_{k_n}$*.*

Let $P_{\boldsymbol{\theta}_n}$ and P_0 denote the probability measures corresponding to H_{1n} and H_0, respectively. Then, we can prove the following result:

Theorem 3 *(Distribution under local alternatives)*
(3.1) $P_{\boldsymbol{\theta}_n}$ *is contiguous with respect to* P_0*.*
(3.2) Under H_{1n}*,*

$$\frac{1}{\sqrt{n}}\mathcal{S}_n(\boldsymbol{\phi}^*) \xrightarrow{d} N[\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty, \mathcal{I}(\boldsymbol{\phi}^*)].$$

(3.3) Under H_{1n} *and Assumption 5,*

$$GET_n \xrightarrow{d} \sup_{\boldsymbol{\lambda}\in\Lambda} \left\{ 2\left[S + \left(\mathcal{I}_{\theta\theta}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi}\phi}(\boldsymbol{\phi}^*)\mathcal{I}_{\phi\theta}(\boldsymbol{\phi}^*) \right) \boldsymbol{\lambda}_{\theta,\infty} \right]' \boldsymbol{\lambda} \right.$$
$$\left. -\boldsymbol{\lambda} \left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*) \right] \boldsymbol{\lambda} \right\}$$

where

$$S \sim \mathcal{N}[0, \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*)].$$

Therefore, the distribution of the empirical process underlying our tests converges to a Gaussian random element with a non-zero mean, and consequently, our test statistic to the supremum of a non-central χ^2-type process, despite the fact that our sequence of local alternatives written in terms of the model parameters converges at rates which are different from the usual ones. In fact, there may be different drifting sequences with the same limit, as we will see in section 3.2.3. In any event, we would like to emphasize that our proposed test is consistent against fixed alternatives because GET_n will diverge in those circumstances.

3 Examples

In this section, we discuss the application of our proposed tests to the first two examples of empirical interest that we mentioned in the introduction. Specifically, we derive a test for irrelevant sample selectivity in multivariate regression models, for which Theorem 1 suffices, and a test for normality against SNP alternatives, which requires our more general Theorem 2. In turn, in Supplemental Appendix C we obtain a test of a multivariate normal copula against its Hermite expansion, which is another example of Theorem 1 but with the added difficulty of inequality constraints on the parameters. Finally, in Supplemental Appendix D, we derive a test aimed at detecting non-linear predictability in a multiple regressor version of Bottai (2003), which again requires the use of Theorem 2 (see also Amengual, Bei and Sentana (2022, 2023) for another two empirically-relevant applications of Theorems 1 and 2, respectively).

3.1 Example 1: Testing for selectivity in multivariate regressions

Arguably, the study of the determinants and consequences of non-random sample selection that followed Heckman's (1974) seminal paper is one of the most important contributions of econometrics in the last fifty years. Nevertheless, the empirical analysis of a dataset would be much simpler if the sample from which it comes could be treated as if it were randomly generated even though it is not necessarily so. As is well known, this will happen when the unobserved determinants of the sample selection are independent of the unobserved determinants of the variables of interest conditional on the set of predetermined explanatory variables, or in simpler terms, when the selection is exogenous rather than endogenous. In the rest of this subsection, we shall develop a test of irrelevant sample selectivity in a multivariate regression context that highlights the hidden difficulties researchers often inadvertently encounter, but which can be easily overcome by the use of the GET procedures that we propose.

3.1.1 The model and its log-likelihood function

Consider the following multivariate version of the regression model with selectivity considered by Lee and Chesher (1986):

$$\mathbf{y} = \mathbf{y}^* d, \tag{3}$$

where d is a sample selection binary variable whose value is determined by an observed vector of exogenous regressors $\mathbf{w}$ and some unobserved determinant u_S according to the following equation written in terms of the usual indicator function

$$d = \mathbf{1}(\mathbf{w}'\boldsymbol{\varphi}^S + u_S \geq 0), \tag{4}$$

while the K partially observed variables $\mathbf{y}^* = (y_1^*, ..., y_K^*)'$ follow the multivariate regression

$$\begin{aligned} y_k^* &= \boldsymbol{\varphi}_k^{M\prime}\mathbf{x} + \varphi_k^D u_k, \quad k = 1, \ldots, K, & (5) \\ \begin{pmatrix} \mathbf{u} \\ u_S \end{pmatrix} |\mathbf{x}, \mathbf{w} &\sim N\left\{\mathbf{0}, \begin{bmatrix} \boldsymbol{R}\left(\boldsymbol{\varphi}^L\right) & \boldsymbol{\vartheta} \\ \boldsymbol{\vartheta}' & 1 \end{bmatrix}\right\}, & (6) \end{aligned}$$

with $\mathbf{x}$ being a vector of exogenous regressors that may partially overlap with $\mathbf{w}$, so that $\boldsymbol{\varphi}^D = (\varphi_1^D, \ldots, \varphi_K^D)'$ contains the standard deviations of the regression shocks, $\boldsymbol{\varphi}^L$ the correlations between them, and $\boldsymbol{\vartheta}$ the correlations between those shocks and the unobserved component of the selection equation, whose variance we normalize to 1 without loss of generality.

Therefore, the contribution of a single observation to the sample log-likelihood function will be given (up to a constant term) by

$$\begin{aligned} &(1-d)\ln\Phi\left(-\mathbf{w}'\boldsymbol{\varphi}^S\right) + d\ln\Phi\left[\frac{\mathbf{w}'\boldsymbol{\varphi}^S + \boldsymbol{\vartheta}'\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)}{\sqrt{1-\boldsymbol{\vartheta}'\mathbf{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\boldsymbol{\vartheta}}}\right] \quad (7) \\ &-\frac{d}{2}\left[2\textstyle\sum_{k=1}^K \ln\varphi_k^D + \ln\{\det[\boldsymbol{R}\left(\boldsymbol{\varphi}^L\right)]\} + \mathbf{u}'(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)\mathbf{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)\right], \end{aligned}$$

where $\boldsymbol{\varphi}^M = (\boldsymbol{\varphi}_1^{M\prime}, \ldots, \boldsymbol{\varphi}_K^{M\prime})'$, $\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D) = [u_1(\boldsymbol{\varphi}_1^M, \varphi_1^D), ..., u_K(\boldsymbol{\varphi}_K^D, \varphi_K^D)]'$, and

$$u_k(\boldsymbol{\varphi}_k^M, \varphi_k^D) = \frac{y_k - \boldsymbol{\varphi}_k^{M\prime}\mathbf{x}}{\varphi_k^D}.$$

3.1.2 The null hypothesis of lack of selectivity and the GET test statistic

Under the null that the unobserved selectivity determinants are uncorrelated with the regression residuals, one can efficiently estimate the multivariate regression coefficients $\boldsymbol{\varphi}^M$ together with the covariance matrix parameters $\boldsymbol{\varphi}^D$ and $\boldsymbol{\varphi}^L$ without selection bias from the non-zero values of $\mathbf{y}$ only using equation by equation OLS without the need to consider the model for d. However, under the alternative, those OLS estimators will be biased because of the sample selectivity, which justifies testing the null hypothesis $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$.

For simplicity, consider the case in which $w = 1$ and the regression equations contain a constant term. Straightforward algebra shows that if we evaluate the scores at $\boldsymbol{\vartheta} = \mathbf{0}$, then

$$s_{\vartheta_k} - M_1(\varphi^S)\varphi_k^D s_{\varphi_{k1}^M} = 0 \quad (8)$$

for $k = 1, \ldots, K$, where φ_{k1}^M contains the intercept in the conditional mean of y_k^*, and

$$M_1\left(\varphi^S\right) = \Phi^{-1}\left(\varphi^S\right)\phi\left(\varphi^S\right) \quad (9)$$

is the usual inverse Mills ratio. As Lee and Chesher (1986) explain in their univariate example, analogous singularities will arise for example when the observed selectivity determinants

$\mathbf{w}$ are given by a set of dummy variables and $\mathbf{x}$ contains those dummy variables too. In general, singularities will be present whenever Heckman's (1976) selectivity correction is perfectly collinear with the regressors that appear in the conditional means of the y^*'s even though the log-likelihood function in (7) is able to locally identify all the model parameters.

In addition to the K singularities in (8), there are $K(K+1)/2$ linear combinations of the scores and the elements of the Hessian corresponding to $\boldsymbol{\vartheta}$ that are 0 too, which effectively means that we need to look at third-order derivatives. To do so, it is convenient to reparametrize from $\boldsymbol{\varphi}$ and $\boldsymbol{\vartheta}$ to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ as we explain in the proof of Proposition 1 below, so that all the elements of the score and the Hessian matrix corresponding to $\boldsymbol{\theta}$ become identically 0 under the null. Fortunately, we can then show that the third-order derivatives with respect to $\boldsymbol{\theta}$, which are only zero on average under the null, will have a full-rank asymptotic covariance matrix, so that we can apply Theorem 1 in this context. Somewhat remarkably, we can show the following result:

Proposition 1 *The difference between LR test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ in model (4)-(6) based on a random sample of n observations on $(\mathbf{y},d)$ and the following test statistic*

$$GET_n = \sup_{\boldsymbol{\upsilon}\neq\mathbf{0}} \frac{1}{\sum_{i=1}^n d_i} \left[\sum_{i=1}^n d_i H_3 \left(\frac{\boldsymbol{\upsilon}' \mathbf{v}_i(\tilde{\boldsymbol{\varphi}}^M, \tilde{\boldsymbol{\varphi}}^D)}{\sqrt{\boldsymbol{\upsilon}'\boldsymbol{\upsilon}}} \right) \right]^2 \tag{10}$$

is $O_p(n^{-1/6})$, where $H_3(z) = (z^3 - 3z)/\sqrt{6}$ is the third-order normalized Hermite polynomial of a standardized variable z, $\boldsymbol{\upsilon}$ is a real vector of dimension K and $\mathbf{v}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$ denotes an affine transformation of the regression residuals $\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$ whose mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_K$, respectively, when evaluated at the restricted parameter estimators.

In simpler terms, our test statistics numerically coincides with the supremum of the moment tests for univariate skewness based on the third Hermite polynomial over all possible linear combinations of the OLS residuals that have 0 mean and unit variance in the sample of observations with $d = 1$. In fact, the standardization is unnecessary because the moment test for univariate skewness is numerically invariant to affine transformations of the observations, which in turn confirms that the test statistic (10) is homogeneous of degree 0 in $\boldsymbol{\upsilon}$. Thus, when $K = 1$ our proposed test reduces to the test for selectivity derived by Lee and Chesher (1986) in the univariate case, which simply assesses the symmetry of the regression residuals by looking at the sample mean of their third powers.

The rationale is also analogous in the multivariate case. Equations (3)-(6) imply that the OLS residuals should be approximately multivariate normally distributed when the unobserved component of the sample selection is independent of the shocks to the observed variables. Under the alternative, in contrast, asymmetry becomes a common feature, as in the multivariate skew normal distribution we discussed in Amengual, Bei and Sentana (2022). Intuitively, if we orthogonalize the regression residuals with respect to the unobserved component of the selectivity equation, we end up with $\boldsymbol{\vartheta} u_S$ as a common component, whose distribution conditional on $d = 1$ is asymmetric even though the unconditional distribution of u_S is symmetric.

3.1.3 Local power analysis

Although the null distribution of the test statistic (10) is non-standard, we can still say something about the determinants of its local power. Consider the following sequence of local alternatives:

$$\lim_{n \to \infty} n^{1/6} \boldsymbol{\theta}_n = \boldsymbol{\theta}_\infty$$

where the rate of convergence is 1/6 rather than 1/2 because of the need for a third-order expansion of the log-likelihood function. Then, we can show that

Proposition 2 *The local power of the test in Proposition 1 only depends on the magnitude of the quadratic form*

$$\boldsymbol{\vartheta}'_\infty \boldsymbol{R}^{-1}\left(\boldsymbol{\varphi}^L\right) \boldsymbol{\vartheta}_\infty.$$

Intuitively, once we orthogonalize the multivariate regression residuals $\mathbf{u}$ by premultiplying by the inverse square root matrix $\boldsymbol{R}^{-1/2}\left(\boldsymbol{\varphi}^L\right)$, the "direction" of the vector $\boldsymbol{R}^{-1/2}\left(\boldsymbol{\varphi}^L\right) \boldsymbol{\vartheta}$ is irrelevant, what matters is its magnitude. As a result, in our simulations we can choose $\boldsymbol{R}\left(\boldsymbol{\varphi}^L\right) = \mathbf{I}_K$ and $\boldsymbol{\vartheta}_\infty$ proportional to the first vector of the canonical basis without loss of generality.

3.1.4 Simulation evidence

For simplicity, we let $w = x_1 = 1$ and $x_2 \sim N(0,1)$. Given that the MLE of the multivariate regression coefficients is equation by equation OLS, and that we are studying the case in which all regressions contain an intercept, the sample mean of the multivariate regression residuals $\hat{\mathbf{u}}$ will be a vector of K zeros. Similarly, any orthogonalization of the $\hat{\mathbf{u}}'s$ based on the estimated covariance matrix will have the identity matrix as sample covariance matrix because the MLEs of the residual standard deviations $\boldsymbol{\varphi}^D$ and correlations $\boldsymbol{\varphi}^L$ match perfectly the sample variances and covariances of $\hat{\mathbf{u}}$ with denominator $\sum_{i=1}^n d_i$. Therefore, it is not surprising that the particular square root that orthonormalizes the OLS residuals in the sample is numerically irrelevant. For example, in the bivariate case, we could define v_1 as the standardized value of u_1 and v_2 as the standardized value of the residual in the OLS regression of u_2 on a constant and u_1. But we could also define them the other way round.

We can easily verify that the GET statistic is numerically invariant to the true values of $(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$, so if we set $K = 2$, we can choose $\boldsymbol{\varphi}_k^M = (0,1)$, $\boldsymbol{\varphi}^D = \boldsymbol{\iota}_2$ without loss of generality. In turn, we set the selection parameter φ^S to 1 and the correlation coefficient φ^L to 0.25.

If we exploited our knowledge of the values of these two parameters, we could compute exact critical values under the null for any sample size to any degree of accuracy by repeatedly simulating samples from the true distribution. In practice, though, we fix the selection parameter and the correlation coefficient to their estimated values in each sample in what is effectively a parametric bootstrap procedure (see Appendix D.1 in Amengual and Sentana (2015) for details), so that we can automatically compute size-adjusted rejection rates, as forcefully argued by Horowitz and Savin (2000).

As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.57, 0.57)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.80, 0)$ (H_{a2}). For each design, we generate 10,000 samples of size n and compute the parameter estimators and tests.

In Table 1 we compare the results of our tests with a bootstrap-based LR test. Panels A and B of Table 1 report the results for samples of length 400 and $1,600$, respectively. We can verify that the LR test statistic is also numerically invariant to the true values $(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$, which allows us to approximate its critical value using an analogous parametric bootstrap procedure. For comparison purposes, we also consider a J-test based on the influence functions underlying GET, which we label as GMM. The first three columns of Table 1 report rejection rates under the null at the 1%, 5% and 10% levels, confirming that our simulated critical values work remarkably well for both sample sizes.[6] In turn, the last six columns present the rejection rates at the 1%, 5% and 10% levels for the alternatives we consider. Our proposed test has similar power to the LR test for the two alternatives, and both these tests outperform the GMM one.

Finally, our results also indicate a Gaussian rank correlation[7] of 0.88 (0.95) between our proposed test statistic and the LR across Monte Carlo simulations of 400 (1,600) observations that satisfy the null, which is in line with the asymptotic equivalence result in Theorem 1. In addition, they indicate that the LR takes about 10 and 20 times as much CPU time to compute as GET does for $n = 400$ and $n = 1,600$, respectively, which makes a huge difference in the calculation of the bootstrap critical values.

3.2 Example 2: Testing for normality against SNP alternatives

Gram-Charlier expansions provide flexible and analytically tractable generalizations of the normal distribution. Unfortunately, their truncated versions lead to negative density values, and the parametric restrictions that Jondeau and Rockinger (2001) propose to guarantee positivity are not easy to implement even when the truncation order is low. In contrast, the SNP distributions introduced by Gallant and Nychka (1987) provide a Hermite expansion of the Gaussian

[6] Given the number of replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1, 5 and 10% levels.

[7] The Gaussian rank correlation between x_1 and x_2 is the Pearson correlation coefficient between $\Phi^{-1}(u_1)$ and $\Phi^{-1}(u_2)$, where u_1 and u_2 are the usual uniform ranks of the observations and $\Phi^{-1}(.)$ the quantile function of the standard normal (see Amengual, Sentana and Tian (2022) for details).

density that is positive by construction. Although these authors introduced those distributions for nonparametric estimation purposes, León, Mencía and Sentana (2009) treated them as parametric ones, studied their statistical properties, and used them in option valuation. Still, MLE under normality is much simpler than when the distribution of the shocks follows an SNP. For that reason, we shall derive a test of normality that will also highlight the hidden complications that researchers face in this context.

3.2.1 The model and its log-likelihood function

The model we consider is

$$y = \mu(\mathbf{x}, \boldsymbol{\alpha}) + \sigma(\mathbf{x}, \boldsymbol{\alpha})\, u \tag{11}$$

where μ and σ are known functions of $\mathbf{x}$ and a finite-dimensional unknown parameter $\boldsymbol{\alpha}$, and u is independent of the predetermined variables in $\mathbf{x}$ with finite mean and variance φ^M and φ^V, respectively. We want to test u is normal against the alternative that it follows an SNP density. Observations are given by $(\mathbf{x}_i, y_i)$, $i = 1, 2, \ldots, n$, where $\mathbf{x}_i$ could include the lagged value of y_i to allow for time-series models such as Ar and Garch. For simplicity, we assume that u_i conditional on $\mathbf{x}_i$ is *iid*. As we will show in section 3.2.5 below, estimation of $\boldsymbol{\alpha}$ does not affect the properties of the test, so we initially assume this parameter vector is known and focus on the case without conditioning variables, in which $\mu(\boldsymbol{\alpha})$ and $\sigma(\boldsymbol{\alpha})$ are 0 and 1 without loss of generality. As a result, researchers only need to estimate φ^M and φ^V under the null.

The probability density function (pdf) of an SNP random variable of order K is given by

$$f(y; \boldsymbol{\varrho}) = \frac{1}{\varphi_2}\phi\left(\frac{y-\varphi^M}{\sqrt{\varphi^V}}\right)\left[\epsilon + \frac{(1-\epsilon)\left\{P\left[\left(\frac{y-\varphi^M}{\sqrt{\varphi^V}}\right); \boldsymbol{\vartheta}\right]\right\}^2}{\int_{-\infty}^{\infty}\left\{P[u; \vartheta]\right\}^2\phi(u)du}\right], \tag{12}$$

with

$$P[u; \boldsymbol{\vartheta}] = 1 + \sum\nolimits_{i=1}^{K}\vartheta_i H_i(u), \tag{13}$$

where $\phi(\cdot)$ denotes the standard normal pdf, $H_i(u)$ is the normalized Hermite polynomial of order i, which can be defined recursively for $i \geq 2$ as

$$H_i(u) = \frac{uH_{i-1}(u) - \sqrt{i-1}H_{i-2}(u)}{\sqrt{i}}, \tag{14}$$

with initial conditions $H_0(u) = 1$ and $H_1(u) = u$, $\int_{-\infty}^{\infty}\{P[u;\vartheta]\}^2\phi(u)du = 1 + \sum_{i=1}^{K}\vartheta_i^2$ is a constant which guarantees that the density integrates to 1, and ϵ is an infinitesimal factor used to bound the density below from 0, which Gallant and Nychka (1987) introduced to simplify their proofs. Henceforth, we will set $\epsilon = 0$ for the purposes of developing our testing procedure, but the same method applies with $\epsilon > 0$. Intuitively, a non-negative density is automatically achieved by multiplying the Gaussian density by the square of a linear combination of Hermite polynomials. As explained by León, Mencía and Sentana (2009), the SNP distributions can have non-negligible positive and negative asymmetry and excess kurtosis even with $K = 2$.

3.2.2 The null hypothesis of normality and the GET test statistic

To simplify the notation, we focus on the case of $K = 2$. Normality is trivially obtained when $H_0 : \vartheta_1 = \vartheta_2 = 0$. The complication arises because

$$s_{\vartheta_1} - 2\sqrt{\varphi^V} s_{\varphi^M} = 0,$$
$$s_{\vartheta_2} - 2\sqrt{2}\varphi^V s_{\varphi^V} = 0,$$

under H_0, so that the nullity of the information matrix is 2. Hall (1990) highlighted this problem when he considered tests of normality against semi-nonparametric alternatives in which the ϑ coefficients were in turn functions of some exogenous variable. However, his proposed solution was to ignore the parameters involved in the singularity, focusing instead only on those which could be regularly estimated under the null. Unfortunately, his recipe would leave us with no test in the case of the unconditional model (12)-(13).

In fact, it is easy to prove that ϑ_1 and ϑ_2 have different orders of identification, which means that we need to resort to our Theorem 2. In this context, we can establish the following result after reparametrizing from $(\boldsymbol{\varphi}', \boldsymbol{\vartheta}') = (\varphi^M, \varphi^V, \vartheta_1, \vartheta_2)$ to $(\boldsymbol{\phi}', \boldsymbol{\theta}') = (\phi^M, \phi^V, \theta_1, \theta_2)$ as explained in its proof:

Proposition 3 *The difference between the LR test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ in model (12)-(13) based on a random sample of n observations on $\mathbf{y}$ and the following test statistic*

$$GET_n = n \left\{ \left[\frac{1}{n} \sum_{i=1}^{n} H_3(\tilde{u}_i) \right]^2 + \left[\frac{1}{n} \sum_{i=1}^{n} H_4(\tilde{u}_i) \right]^2 \right\} \tag{15}$$

is $O_p(n^{-1/6})$ when the null is true, where $H_3(\tilde{u}_i)$ and $H_4(\tilde{u}_i)$ are the third- and fourth-order normalized Hermite polynomials of the $\tilde{u}_i$'s, which are the values of the y_i's standardized so that their sample mean and variance are 0 and 1, respectively.

Remarkably, this means that the Jarque and Bera (1980) test is asymptotically equivalent to the LR test of normality against SNP densities, although they converge to each other at a much lower rate than in the case of the Pearson family of alternative distributions they considered.

3.2.3 Local power analysis

Let $\chi_k^2(\upsilon)$ denote a non-central chi-square random variable with k degrees of freedom and non-centrality parameter υ. We can show that

Proposition 4 *Consider a sequence of parameters* $\boldsymbol{\theta}_n$ *satisfying*

$$\lim_{n\to\infty} \sqrt{n} \begin{pmatrix} -2\sqrt{3}\theta_{1,n}\theta_{2,n} \\ \sqrt{6}(\frac{1}{9}\theta_{1,n}^4 - \theta_{2,n}^2) \end{pmatrix} = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty} \in \mathbb{R}^2. \tag{16}$$

Under the sequence of DGPs indexed by $\boldsymbol{\theta}_n$*,*

$$GET_n \xrightarrow{d} \chi_2^2(\boldsymbol{\lambda}'_{\boldsymbol{\theta},\infty}\boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}).$$

To understand this result, it is useful to note that

$$\begin{bmatrix} \sqrt{n}E\left\{H_3\left[\frac{y-E(y)}{\sqrt{Var(y)}}\right]\right\} \\ \sqrt{n}E\left\{H_4\left[\frac{y-E(y)}{\sqrt{Var(y)}}\right]\right\} \end{bmatrix} = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty} + o(1).$$

Unlike in the multivariate regression model with selectivity, though, we can have two different types of local alternatives compatible with (16):

$$H_{l1} : \theta_{1n} = n^{-\frac{1}{4}}h_1,\ \theta_{2n} = n^{-\frac{1}{4}}h_2,$$
$$H_{l2} : \theta_{1n} = n^{-\frac{1}{8}}h_1,\ \theta_{2n} = n^{-\frac{3}{8}}h_2.$$

Interestingly, $\sqrt{n}\theta_{2n}^2$ dominates $\sqrt{n}\theta_{1n}^4/9$ along H_{l1}, so that the SNP distributions under this sequence of local alternatives are platykurtic. In contrast, $\sqrt{n}\theta_{1n}^4/9$ dominates $\sqrt{n}\theta_{2n}^2$ along H_{l2}, so that the corresponding SNP distributions are leptokurtic.

3.2.4 Simulation evidence

Despite the fact that we estimate the sample mean and variance of each simulated sample, in this case there are effectively no nuisance parameters involved because both the GET and LR test statistics are numerically invariant to affine transformations of the observations. As a result, we can compute the exact finite sample distribution to any desired degree of accuracy for any sample size by simulating a large number of samples of the same size from a standard normal random variable. For that reason, we can focus directly on studying the power of the different tests.

As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.25, 0.10)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.75, 0.05)$ (H_{a2}), setting $\varphi^M = 0$ and $\varphi^V = 1$ without loss of generality. As in the previous example, for each design we generate 10,000 samples of size n. In Table 2 we compare the results of our tests with the LR test. Panels A and B of Table 2 report the results for samples of length 400 and $1,600$, respectively. Given that the LR test statistic is also numerically invariant to the true values (φ^M, φ^V), we once again obtain its exact critical values using an analogous parametric bootstrap procedure. The first three columns of Table 2 report rejection rates under H_{a1} at the same levels, while the last three columns present the rejection rates at the 1%, 5% and 10% levels for H_{a2}. As can be seen, our proposed test has similar power to the LR test for both alternatives.

Finally, our results also indicate that the LR takes around 160 and 100 times as much CPU time to compute as GET does for $n = 400$ and $n = 1,600$, respectively, which considerably slows down the calculation of the bootstrap critical values.

3.2.5 Robustness to the estimation of mean and variance parameters

We now extend our previous results to a situation in which the conditional mean and variance of y are parametric functions of the variable in $\mathbf{x}$, as in (11). In this context, the objective becomes to test whether the innovation u follows a normal distribution versus an SNP.

The conditional log-likelihood of the i^{th} observation is given by:

$$l_i(\boldsymbol{\alpha}, \vartheta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln \sigma_Y^2(x_i, \boldsymbol{\alpha}) - \frac{1}{2}u_i^2(\boldsymbol{\alpha}) + 2\ln\left(1 + \sum_{i=1}^{K} \vartheta_i H_i\left[u_i(\boldsymbol{\alpha})\right]\right) - \ln\left(1 + \sum_{i=1}^{K} \vartheta_i^2\right).$$

To be able to obtain the required higher-order log-likelihood expansions, we assume that the following regularity conditions hold:

Assumption 6 *(Smoothness of the conditional first two moments) The conditional mean and variance functions* $\mu_Y(\mathbf{x}_i, \boldsymbol{\alpha})$ *and* $\sigma_Y(\mathbf{x}_i, \boldsymbol{\alpha})$ *that appear in (11) are such that*
(6.1) They are eight times continuously differentiable with respect to $\boldsymbol{\alpha}$*.*
(6.2) For all $k \in N^{d_{\boldsymbol{\alpha}}}$ *and* $\iota' k = 1, \ldots, 8$*, it holds that*

$$E\left[\left(\frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\mu_Y(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}}\right)^2\right] < \infty,\; E\left[\left(\frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\sigma_Y^2(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}}\right)^2\right] < \infty,$$

where $k = (k_1, \ldots, k_{d_{\boldsymbol{\alpha}}})$*,*

$$\begin{aligned} \frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\mu_Y(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}} &= \frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\mu_Y(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_1^{k_1} \ldots \partial \alpha_{d_\alpha}^{k_{d_{\boldsymbol{\alpha}}}}}, \quad \textit{and} \\ \frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\sigma_Y^2(\mathbf{x}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\boldsymbol{k}}} &= \frac{\partial^{\boldsymbol{\iota}'\boldsymbol{k}}\sigma_Y^2(\mathbf{x}, \boldsymbol{\alpha})}{\partial \alpha_1^{k_1} \ldots \partial \alpha_{d_\alpha}^{k_{d_{\boldsymbol{\alpha}}}}}. \end{aligned}$$

Then, we can prove the following result, which is entirely analogous to Proposition 8 in Amengual, Bei, Carrasco and Sentana (2022):

Proposition 5 *Under Assumptions (6.1) and (6.2), replacing the true value of* $\boldsymbol{\alpha}$ *by* $\tilde{\boldsymbol{\alpha}}$*, its restricted maximum likelihood estimator under* H_0*, does not alter the expressions of the GET test in Proposition 3 or its asymptotic distribution under the null or sequences of local alternatives.*

4 Conclusions

We propose a generalization of the extremum-type tests in Lee and Chesher (1986) to models in which the nullity of the information matrix under the null hypothesis is larger than one. In the case of a single singularity, our results are consistent with theirs, as well as with those in Rotnitzky et al. (2000). However, when the information matrix is repeatedly singular, we provide a computationally convenient alternative to the LR test, which is particularly useful for resampling-based calculations of p-values. Specifically, our proposed test statistic is a sup-type test over a space whose dimension is at most the nullity of the information matrix, and often less, while the maximization of the original log-likelihood function is over a space of the same dimension as the vector of parameters, which is usually much larger. In addition, the fact that several log-likelihood derivatives of various orders are 0 under the null implies that the LR requires the estimation of all the parameters that appear under the alternative in a model whose log-likelihood function is extremely flat around its maximum. Intuitively, the substantial computational gains that we find arise because GET is a LR-type test that compares the log-likelihood function under the null to the maximum of its lowest-order approximation under the alternative capable of identifying the restricted parameters.

Despite having many features in common, our results cannot be directly applied to testing normality against finite Gaussian mixtures. Nevertheless, we used them as a powerful lever to derive such tests in Amengual, Bei, Carrasco and Sentana (2022).

Interestingly, the asymptotic distribution of our test statistic is similar to the asymptotic distribution of the usual overidentification test statistic in a GMM model in which the expected Jacobian of the moment conditions is of reduced rank but the parameters are second-order identified (see Supplemental Appendix E of Amengual, Bei and Sentana (2020) for a formal link to the results in Dovonon and Renault (2013)). An application of our approach to GMM contexts in which not only the expected Jacobian matrix is singular but some higher order Jacobian matrices are singular too would constitute a very interesting extension.

Finally, the tests developed in this paper allowed us to provide some new insights about the cross-section distribution of city sizes and their growth rates in Amengual, Bei and Sentana (2022). Their use in some of the other empirically relevant situations discussed in this paper would also provide a particularly valuable complement to our theoretical results.

References

Amengual, D., Bei, X., Carrasco, M. and Sentana, E. (2022): "Score-type tests for normal mixture models", CEMFI Working Paper 2213.

Amengual, D., Bei, X. and Sentana, E. (2020): "Hypothesis tests with a repeatedly singular information matrix", CEMFI Working Paper 2002.

Amengual, D., Bei, X. and Sentana, E. (2022): "Normal but skewed?", *Journal of Applied Econometrics* 37, 1295–1313.

Amengual, D., Bei, X. and Sentana, E. (2023): "Highly irregular serial correlation tests", CEMFI Working Paper 2302.

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

Amengual, D., Sentana, E. and Tian, Z. (2022): "Gaussian rank correlation and regression", in A. Chudik, C. Hsiao and A. Timmermann (eds.) *Essays in honor of M. Hashem Pesaran: panel modeling, micro applications and econometric methodology*, *Advances in Econometrics* 43B, 269–306, Emerald.

Bera, A.K. and Bilias, Y. (2001): "Rao's Score, Neyman's C(α) and Silvey's LM tests: an essay on historical developments and some new results", *Journal of Statistical Planning and Inference* 97, 9–44.

Bera, A., Ra, S. and Sarkar, N. (1998): "Hypothesis testing for some nonregular cases in econometrics", *Econometrics: theory and practice*, Chakravarty, Coondoo and Mukherjee (eds.), 319–351, Allied Publishers.

Bottai, M. (2003): "Confidence regions when the Fisher information is zero", *Biometrika* 90, 73–84.

Breusch, T.S. and Pagan, A.R. (1980): "The Lagrange multiplier test and its applications to model specification in econometrics", *Review of Economic Studies* 47, 239–253.

Cox, D. and Hinkley, D. (1974): *Theoretical statistics*, Chapman and Hall.

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

Davidson, J. (1994): *Stochastic limit theory: an introduction for econometricians*, Oxford University Press.

Dovonon, P. and Renault, E. (2013): "Testing for common conditionally heteroskedastic factors", *Econometrica* 81, 2561–2586.

Engle, R.F. (1983): "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics", in Intriligator, M. D.; Griliches, Z., eds., *Handbook of Econometrics*, 796–801, Elsevier.

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

Gallant, A.R. and Nychka, D.W. (1987): "Semi-nonparametric maximum likelihood estimation", *Econometrica* 55, 363–390.

Godfrey, L.G. (1988): *Misspecification tests in econometrics.* Cambridge University Press.

Hall, A. (1990): "Lagrange Multiplier tests for normality against seminonparametric alternatives", *Journal of Business and Economic Statistics* 8, 417–426.

Heckman, J. (1974): "Shadow prices, market wages, and labor supply", *Econometrica* 42, 679–694.

Heckman, J. (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models", *Annals of Economic and Social Measurement* 5, 475–492.

Horowitz, J. and Savin, N.E. (2000): "Empirically relevant critical values for hypothesis tests: a bootstrap approach", *Journal of Econometrics* 95, 375–389.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, homoscedasticity and serial independence of regression residuals", *Economics Letters* 6, 255–259.

Jondeau, E. and M. Rockinger (2001): "Gram-Charlier densities", *Journal of Economic Dynamics and Control* 25, 1457–1483.

Lee, L. F. and A. Chesher (1986): "Specification testing when score test statistics are identically zero", *Journal of Econometrics* 31, 121–149.

León, A., Mencía, J. and Sentana, E. (2009): "Parametric properties of semi-nonparametric distributions, with applications to option valuation", *Journal of Business and Economic Statistics* 27, 176-192.

Meitz M. and Saikkonen, P. (2021): "Testing for observation-dependent regime switching in mixture autoregressive models", *Journal of Econometrics* 222, 601–624.

Newey, W. and McFadden, D. (1994): "Large sample estimation and hypothesis testing", in Engle, R. and McFadden, D., eds., *Handbook of Econometrics*, 2111–2245, Elsevier.

O'Hagan, A. and Leonard, T. (1976): "Bayes estimation subject to uncertainty about parameter constraints", *Biometrika* 63, 201–203.

Rao, C.R. (1948): "Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation", *Mathematical Proceedings of the Cambridge Philosophical Society* 44, 50–57.

Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000): "Likelihood-based inference with singular information matrix", *Bernoulli* 6, 243–284.

Sargan, J.D. (1983): "Identification and lack of identification", *Econometrica* 51, 1605–1633.

Silvey, S. D. (1959): "The Lagrangian multiplier test", *Annals of Mathematical Statistics* 30, 389–407.

van der Vaart, A.W. (1998): *Asymptotic statistics*, Cambridge.

Appendices

A Proofs

We first state and prove several lemmas that we will use in the proofs of our main theorems. But before doing so, let us introduce some definitions. Let

$$LM_n(\boldsymbol{\rho}) = 2\mathcal{S}_n'(\boldsymbol{\phi}^*)\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) - n\boldsymbol{\lambda}'(\boldsymbol{\phi},\boldsymbol{\theta})\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$$

and define $\boldsymbol{\rho}^{LM} = (\boldsymbol{\phi}^{LM}, \boldsymbol{\theta}^{LM})$ such that

$$LM_n(\boldsymbol{\phi}^{LM}, \boldsymbol{\theta}^{LM}) = \sup_{\boldsymbol{\rho}\in \mathrm{P}} LM_n(\boldsymbol{\rho}).$$

Lemmata

Lemma 1 *If Assumptions 1 and 4.1, 4.2, 4.3 hold, then (i)* $\boldsymbol{\rho}^{LM} \xrightarrow{p} \mathbf{0}$ *and (ii)* $n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM}) = O_p(1)$.

Proof. Let us start by Lemma 1.(ii). Fix $\epsilon > 0$. By Assumption 4.2, we have that $n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*) = O_p(1)$, which means that there exists an M_1 such that for all $n \geq N$,

$$\Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1) \leq \epsilon. \tag{A1}$$

Next, let $M = (2M_1+1)/e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]$, which is a positive real number because of Assumption 4.3. We can then prove that

$$\Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\}) = 0. \tag{A2}$$

In addition, noticing that if $||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M$ and $||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1$, we will have that

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

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 $LM_n\left(\boldsymbol{\theta}^*, \mathbf{0}\right) = 0$, which contradicts $\boldsymbol{\rho}^{LM}$ being the minimizer. Thus (A2) holds.

Therefore,

$$\begin{aligned}\Pr(||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M) =& \Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\}) \\ &+ \Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1\}) \\ \leq& \Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}^{LM}\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\}) \quad \text{(A3)} \\ &+ \Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| > M_1) \\ \leq& \epsilon, \quad \text{(A4)}\end{aligned}$$

where to go from (A3) to (A4) we have used (A1) and (A2). As a consequence, (A4) trivially implies that Lemma 1.(ii) 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 4.1, while the convergence follows from Lemma 1.(ii), as desired. □

Lemma 2 *If Assumptions 1 and 4.1–4 hold, then* $n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}) = O_p(1)$.

Proof. Fix $\epsilon > 0$. Assumption 1 implies the consistency of $\hat{\boldsymbol{\rho}}$, while Assumption 4.4 implies that

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

Thus, there exists an N such that for all $n > N$,

$$\Pr\left(A_n\right) \geq 1 - \frac{\epsilon}{2}, \quad \text{(A5)}$$

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,

$$\Pr(||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \geq M_1) < \frac{\epsilon}{2}. \quad \text{(A6)}$$

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

$$\Pr(\{||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| \geq M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1\} \cap A_n) = 0. \quad \text{(A7)}$$

Further, if we notice that

$$||n^{\frac{1}{2}}\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)|| \geq M, \;\; ||n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)|| \leq M_1 \;\;\text{and}\;\; \left|\frac{R_n(\hat{\boldsymbol{\rho}})}{1 + n\left\|\boldsymbol{\lambda}\left(\hat{\boldsymbol{\rho}}\right)\right\|^2}\right| \leq \frac{1}{6}e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)], \quad \text{(A8)}$$

we can show that

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

where the first equality follows from Assumption 4, the first inequality from (A8), the next three lines are straightforward, the subsequent inequality follows from $||n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}})|| \geq M \geq (6M_1 + 3)/e_{\min}[\mathcal{I}(\boldsymbol{\phi}^*)]$, and the last equality from $LR(\boldsymbol{\phi}^*, \mathbf{0}) = 0$. Therefore,

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

for all $n > N$, where the inequalities follow from (A5), (A6) and (A7). □

Lemma 3 *If Assumptions 1 and 4.1–4 hold, then* $LR_n(\hat{\boldsymbol{\rho}}) = LM_n(\boldsymbol{\rho}^{LM}) + o_p(1)$.

Proof. We will show that for all $\epsilon > 0$, there exists an N such that for all $n > N$,

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

To do so, we know that $\max\{n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}), n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM})\} = O_p(1)$, so that for all $\epsilon > 0$, there exists an M such that for all n,

$$\Pr(\max\{n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}), n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}^{LM})\} \leq M) > 1 - \frac{\epsilon}{2}. \tag{A9}$$

Next, letting $P_n = \{\boldsymbol{\rho} \in \mathbf{P} : n^{\frac{1}{2}}\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \leq M\}$, we can use Assumption 4.1 to choose a sequence of $\gamma_n \to 0$ satisfying

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

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

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

where the first line follows from Assumption 4, the second one from the definition of P_n, the third one from $A_n = \{\boldsymbol{\rho} \in \mathbf{P}$: $n^{\frac{1}{2}} \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \leq M\} \subset \{\boldsymbol{\rho} \in \mathrm{P} : \|\boldsymbol{\rho} - (\boldsymbol{\phi}^*, \mathbf{0})\| \leq \gamma_n\}$, and the last equality from $\gamma_n \to 0$ and Assumption 4.4. Thus, there exists an N such that for all $n > N$,

$$
\Pr\left(\sup_{\boldsymbol{\rho} \in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon\right) > 1 - \frac{\epsilon}{2}. \tag{A10}
$$

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

$$
\begin{aligned}
& \Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < \epsilon\right) \\
\geq & \Pr\left(\left\{\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < \epsilon\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) & \text{(A11)} \\
\geq & \Pr\left(\left\{\sup_{\boldsymbol{\rho} \in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) & \text{(A12)} \\
\geq & \Pr\left(\sup_{\boldsymbol{\rho} \in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon\right) + P\left(\{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) - 1 & \text{(A13)} \\
\geq & 1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 = 1 - \epsilon, & \text{(A14)}
\end{aligned}
$$

where to go from (A11) to (A12) we have used

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

from (A12) to (A13) the fact that $\Pr(E_1 \cap E_2) \geq \Pr(E_1) + \Pr(E_2) - 1$, while from (A13) to (A14) we relied on (A9) and (A10). □

Lemma 4 *If Assumptions 1, 4.1–4 and 4.6 hold, then* $LR_n(\hat{\boldsymbol{\rho}}) = LM_n(\boldsymbol{\rho}^{LM}) + O_p(n^{-a})$.

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

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

The proof is almost the same as the one of Lemma 3. Let M and P_n be as the ones in that lemma. Then, by Assumption 4.6,

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

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

$$\Pr\left(\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-a}\right) > 1 - \frac{\epsilon}{2}. \tag{A15}$$

Thus,

$$\begin{aligned}
&\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < K_\epsilon n^{-a}\right)\\
\geq &\Pr\left(\left\{\left|LR_n(\hat{\boldsymbol{\rho}}) - LM_n(\boldsymbol{\rho}^{LM})\right| < K_\epsilon n^{-a}\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right)\\
\geq &\Pr\left(\left\{\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-a}\right\} \cap \{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) && \text{(A16)}\\
\geq &\Pr\left(\sup_{\boldsymbol{\rho}\in P_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-a}\right) + \Pr\left(\{\hat{\boldsymbol{\rho}} \in P_n\} \cap \left\{\boldsymbol{\rho}^{LM} \in P_n\right\}\right) - 1 && \text{(A17)}\\
\geq &1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 = 1 - \epsilon, && \text{(A18)}
\end{aligned}$$

where the last inequality follows from (A9) and (A15). □

Lemma 5 *If Assumptions 1 and 4.1–4 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 in addition Assumption 4.6 holds, then* $LR_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) = \sup_{(\boldsymbol{\phi},\mathbf{0})\in \boldsymbol{P}} LM_n(\boldsymbol{\phi}, \mathbf{0}) + O_p(n^{-a})$.

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 $\mathbf{P}$ to $\{\boldsymbol{\phi} : (\boldsymbol{\phi}, \mathbf{0}) \in \mathbf{P}\}$. □

Proof of Theorem 2

By virtue of Lemma 1, we have that $\boldsymbol{\rho}^{LM} \in \Phi \times \Theta$ with probability approaching 1 (w.p.a. 1 henceforth), with Θ and Φ defined as $\Phi \times \Theta = \mathbf{P}$, that is, if $\boldsymbol{\rho} \in \mathbf{P}$, then $\boldsymbol{\phi} \in \Phi$ and $\boldsymbol{\theta} \in \Theta$. It is then easy to verify that

$$\begin{aligned}
&\sup_{\boldsymbol{\rho}\in \mathrm{P}} 2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})] - [n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]\\
&= \sup_{\theta\in\Theta}\sup_{\phi\in\Phi}\Big\{2n^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)'n^{\frac{1}{2}}[\boldsymbol{\phi} - \boldsymbol{\phi}^* + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta})] - n[\boldsymbol{\phi} - \boldsymbol{\phi}^* + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta})]'\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)[\boldsymbol{\phi} - \boldsymbol{\phi}^* + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta})]\\
&- 2n^{\frac{1}{2}}[\boldsymbol{\phi} - \boldsymbol{\phi}^* + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta})]'\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})] + 2n^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*)'[n^{\frac{1}{2}}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})]\\
&-[n^{\frac{1}{2}}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})]'\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})]\Big\} \text{ w.p.a. 1}\\
&= \sup_{\boldsymbol{\theta}\in\Theta}\Big\{2[\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)S_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\\
&-n\boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta})[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\Big\} + n^{-1}\mathcal{S}'_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)
\end{aligned}$$

w.p.a. 1, where the first equality follows from $\boldsymbol{\rho}^{LM} \in \Phi \times \Theta$ w.p.a. 1, and the second one from $\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)[n^{-1}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^{LM})] - \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}^{LM}) \in \{\boldsymbol{\phi} - \boldsymbol{\phi}^* : \boldsymbol{\phi} \in \Phi\}$ w.p.a. 1.

Similarly, we have that

$$\sup_{(\phi,\mathbf{0})\in\mathbf{P}} 2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]-[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]=\frac{1}{n}\mathcal{S}'_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)$$

w.p.a. 1. As a result,

$$\begin{aligned}
LR &= 2[L_n(\hat{\boldsymbol{\phi}},\boldsymbol{\theta})-L_n(\tilde{\boldsymbol{\phi}},\mathbf{0})]\\
&= 2[L_n(\hat{\boldsymbol{\phi}}_n,\boldsymbol{\theta})-L_n(\boldsymbol{\phi}^*,\mathbf{0})]-2[L_n(\tilde{\boldsymbol{\phi}},\mathbf{0})-L_n(\boldsymbol{\phi}^*,\mathbf{0})]\\
&= \sup_{\boldsymbol{\rho}\in\mathrm{P}}\Big\{2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]-[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]\Big\}\\
&- \sup_{(\boldsymbol{\phi},\mathbf{0})\in\mathbf{P}}\Big\{2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]-[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]\Big\}+o_p(1)\\
&= \sup_{\boldsymbol{\theta}\in\Theta}\Big\{2[\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*)-\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \qquad \text{(A19)}\\
&-n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*)-\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*)]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\Big\}+o_p(1),
\end{aligned}$$

where the first two equalities are trivial, while the third one follows from Lemmas 3 and 5.

The last step is to evaluate (A19) at $\tilde{\boldsymbol{\phi}}$ instead of $\boldsymbol{\phi}^*$. Specifically, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}||\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})-\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})-\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*)+\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)||\\
&= \left\|\begin{bmatrix} -\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*) & I\end{bmatrix} n^{-\frac{1}{2}}[\mathcal{S}_n(\tilde{\boldsymbol{\phi}})-\mathcal{S}_n(\boldsymbol{\phi}^*)]\right\|\\
&= \left\| -\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*) \quad I \right\| \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{g}(\mathbf{y}_i)\right\| ||\tilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^*|| = O_p(n^{-\frac{1}{2}}), \qquad \text{(A20)}
\end{aligned}$$

where the first equality is straightforward, the second one follows from (2), and the last equality from $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{g}(y_i)=O_p(1)$ and $||\tilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^*||=O_p(n^{-\frac{1}{2}})$, with $\mathbf{g}(.)$ defined in Assumption (4.5). Moreover, Assumption 4.3 means that $\mathcal{I}(\boldsymbol{\phi})$ is Lipschitz, so that

$$||\mathcal{I}(\tilde{\boldsymbol{\phi}})-\mathcal{I}(\boldsymbol{\phi}^*)||=O_p(n^{-\frac{1}{2}}). \qquad \text{(A21)}$$

Combining (A20) and (A21), we get

$$\begin{aligned}
&\sup_{\theta\in\Theta}\Big\{2(\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*)-\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*))'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\\
&\quad -n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*)-\mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*)]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\Big\}\\
&= \sup_{\theta\in\Theta}\Big\{2[\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})-\mathcal{I}_{\boldsymbol{\theta\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\\
&\quad -n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'[\mathcal{I}_{\boldsymbol{\theta\theta}}(\tilde{\boldsymbol{\phi}})-\mathcal{I}_{\boldsymbol{\theta\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\theta}}(\tilde{\boldsymbol{\phi}})]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\Big\}+O_p(n^{-\frac{1}{2}}), \qquad \text{(A22)}
\end{aligned}$$

which, together with (A19) and (A22), complete the proof of the first part of the theorem.

Using the same argument, we will have that

$$\begin{aligned} LR &= \sup_{\boldsymbol{\theta}\in\Theta}\Big\{2[\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &\quad -n\boldsymbol{\lambda}_{\boldsymbol{\theta}}'(\boldsymbol{\theta})[\mathcal{I}_{\boldsymbol{\theta\theta}}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\theta}}(\tilde{\boldsymbol{\phi}})]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\Big\} + O_p(n^{-a}) \\ &= \sup_{\boldsymbol{\theta}}\Big\{2[\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \\ &\quad -n\boldsymbol{\lambda}_{\boldsymbol{\theta}}'(\boldsymbol{\theta})[\mathcal{I}_{\boldsymbol{\theta\theta}}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi\theta}}(\tilde{\boldsymbol{\phi}})]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\Big\} + O_p(n^{-a}), \end{aligned}$$

when Assumption 4.5 also holds, where the second equality holds because $\mathbf{0}$ is an interior point of Θ and the maximizer is $o_p(1)$, which proves the second part of the theorem. □

Proof of Theorem 1

We will use Theorem 2 to prove Theorem 1. The first step is to verify Assumption 4. To do so, define $\varsigma_{\boldsymbol{\theta}r}(\boldsymbol{\phi}^*) = \mathbf{B}\mathbf{H}_{rn}(\boldsymbol{\phi}^*)$, where

$$\mathbf{H}_{rn}(\boldsymbol{\phi}) = \frac{\partial^r l_n(\boldsymbol{\phi},\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}} - \begin{bmatrix} I_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) \end{bmatrix} \begin{bmatrix} I_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\phi\theta}_1}(\boldsymbol{\phi}) \\ I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\theta_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial l_n}{\partial\boldsymbol{\phi}}(\boldsymbol{\phi},\mathbf{0}) \\ \frac{\partial l_n}{\partial\boldsymbol{\theta}_1}(\boldsymbol{\phi},\mathbf{0}) \end{bmatrix},$$

and $\mathbf{B}$ is a matrix with elements equal to 0 or 1 such that $\varsigma_{\boldsymbol{\theta}r}(\boldsymbol{\phi}^*)$ contains the elements in $\mathbf{H}_{rn}(\boldsymbol{\phi}^*)$ that are not linearly dependent. Notice that $\mathbf{B}$ and $\varsigma_{\boldsymbol{\theta}r}(\boldsymbol{\phi}^*)$ always exist even though they are not necessarily unique. But then,

$$\frac{\partial^r l}{\partial\boldsymbol{\theta}_r^{\otimes r}}(\boldsymbol{\phi}^*,\mathbf{0}) = \mathbf{A}_1\frac{\partial l_n}{\partial\boldsymbol{\phi}}(\boldsymbol{\phi}^*,\mathbf{0}) + \mathbf{A}_2\frac{\partial l_n}{\partial\boldsymbol{\theta}_1}(\boldsymbol{\phi}^*,\mathbf{0}) + \mathbf{A}_3\varsigma_{\boldsymbol{\theta}r}(\boldsymbol{\phi}^*),$$

where $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ are $r^2\times(p-q)$, $r^2\times q$ and $r^2\times\dim(\varsigma_{\boldsymbol{\theta}r})$ matrices, respectively. As a consequence, we will have that

$$\frac{1}{r!}\boldsymbol{\theta}_r^{\otimes r\prime}\frac{\partial^r l}{\partial\boldsymbol{\theta}_r^{\otimes r}}(\boldsymbol{\phi}^*,\mathbf{0}) = \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r)\frac{\partial l}{\partial\boldsymbol{\phi}}(\boldsymbol{\phi}^*,\mathbf{0}) + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)\frac{\partial l}{\partial\boldsymbol{\theta}_1}(\boldsymbol{\phi}^*,\mathbf{0}) + \boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)\varsigma_{\boldsymbol{\theta}r}(\boldsymbol{\phi}^*),$$

with $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r) = \frac{1}{r!}v_r^{\otimes r\prime}\mathbf{A}_1$, $\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r) = \frac{1}{r!}\boldsymbol{\theta}_r^{\otimes r\prime}\mathbf{A}_2$, $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) = \frac{1}{r!}\boldsymbol{\theta}_r^{\otimes r\prime}\mathbf{A}_3$. It is then easy to see that $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r)$, $\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)$ and $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)$ are continuous and $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\eta\boldsymbol{v}) = \eta_{\boldsymbol{\phi}}^r\boldsymbol{\lambda}(\boldsymbol{v})$ for all $\eta\in R$ and $\boldsymbol{v}\in\mathbb{R}^{q_r}$, and the same applies to $\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)$ and $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)$.

Next, let $\mathcal{S}_n = (\mathbf{S}'_{\boldsymbol{\phi}n}, \mathbf{S}'_{\boldsymbol{\theta}_1n}, \mathcal{S}'_{\boldsymbol{\theta}_rn})'$, with

$$\mathbf{S}_{\boldsymbol{\phi}n}(\boldsymbol{\phi}) = \sum_{i=1}^n \mathbf{s}_{\boldsymbol{\phi},i}(\boldsymbol{\phi}) = \sum_{i=1}^n \frac{\partial l_i}{\partial\boldsymbol{\phi}}(\boldsymbol{\phi},\mathbf{0}),\ \ \mathbf{S}_{\boldsymbol{\theta}_1}(\boldsymbol{\phi}) = \sum_{i=1}^n \mathbf{s}_{\boldsymbol{\theta}_1,i}(\boldsymbol{\phi}) = \sum_{i=1}^n \frac{\partial l_i(\boldsymbol{\phi},\mathbf{0})}{\partial\boldsymbol{\theta}_1},$$

$$\mathcal{S}_{\boldsymbol{\theta}_rn}(\boldsymbol{\phi}) = \sum_{i=1}^n \varsigma_{\boldsymbol{\theta}r,i}(\boldsymbol{\phi}).$$

Further, let

$$\mathcal{I}(\boldsymbol{\phi}) = \begin{bmatrix} I_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\phi\theta}_1}(\boldsymbol{\phi}) & \mathbf{0} \\ I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) & I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix}$$

denote the asymptotic variance of $n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi})$, which is block diagonal by construction. Let us also define

$$LM_n(\boldsymbol{\rho}) = 2n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})] - [n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})],$$

where

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) &= [\boldsymbol{\phi}-\boldsymbol{\phi}^*+\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r), \boldsymbol{\theta}_1+\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r), \boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)]', \qquad \text{(A23)}\\ LR(\boldsymbol{\rho}) &= 2\left[L_n(\boldsymbol{\rho}) - L_n(\boldsymbol{\phi}^*,\mathbf{0})\right],\end{aligned}$$

and

$$R(\boldsymbol{\phi},\boldsymbol{\theta}) = \frac{1}{2}[LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})].$$

We next verify Assumption 4.1 for $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$, whose definition is given in (A23). The continuity of $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$ means that we only need to verify that the unique solution to $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) = \mathbf{0}$ is $(\boldsymbol{\phi}^*,\mathbf{0})$ because it is trivial to see that $\boldsymbol{\lambda}(\boldsymbol{\phi}^*,\mathbf{0}) = \mathbf{0}$. First, if $\boldsymbol{\theta}_r = \mathbf{0}$, then it immediately follows that we must have $\boldsymbol{\phi} = \boldsymbol{\phi}^*$ and $\boldsymbol{\theta}_1 = \mathbf{0}$. Consider the case when $\boldsymbol{\theta}_r \neq \mathbf{0}$. By Assumption 3.2, for all $\boldsymbol{\theta}_r \neq \mathbf{0}$, $\boldsymbol{\theta}_r^{\otimes r\prime}\frac{\partial^r l(\boldsymbol{\phi}^*,\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}}$ is linearly independent of $[\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\phi}^*), \mathbf{s}_{\boldsymbol{\theta}_1}(\boldsymbol{\phi}^*)]'$, which implies that $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) \neq \mathbf{0}$ because

$$\boldsymbol{\theta}_r^{\otimes r\prime}\frac{\partial^r l(\boldsymbol{\phi}^*,\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}} = \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r)\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\phi}^*) + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)\mathbf{s}_{\boldsymbol{\theta}_1}(\boldsymbol{\phi}^*) + \boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)\mathbf{s}_{\boldsymbol{\theta}_r}(\boldsymbol{\phi}^*).$$

To verify Assumptions 4.2 and 4.3, notice that the covariance kernel of $\mathcal{S}$ is finite by Assumption 2.2, which implies that Assumption 4.2 will hold by the uniform central limit theorem. Next, $(n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\phi}n}, n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\theta}_1 n})'$ has a full rank asymptotic variance because of Assumption 3.2, so $\frac{1}{\sqrt{n}}\mathbf{S}_{\boldsymbol{\theta}_r n}$ does not belong to the linear span of $(n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\phi}n}, n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\theta}_1 n})'$ by construction. If we combine this result with 4.2, we will have $0 < e_{\min}(\boldsymbol{\phi}^*) < e_{\max}(\boldsymbol{\phi}^*) < \infty$, as desired.

The verification of Assumption 4.4 contains two parts. In the first part, we show that

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\le\gamma_n}\frac{|R_n(\boldsymbol{\phi},\boldsymbol{\theta})|}{h_n(\boldsymbol{\phi},\boldsymbol{\theta})} = o_p(1),$$

where

$$h_n(\boldsymbol{\phi},\boldsymbol{\theta}) = \max\{1, n\|\boldsymbol{\phi}-\boldsymbol{\phi}^*\|^2, n\|\boldsymbol{\theta}_1\|^2, n\|\boldsymbol{\theta}_r\|^{2r}\}.$$

Then, in the second part, we show that

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\le\gamma_n}\frac{h(\boldsymbol{\phi},\boldsymbol{\theta})}{1+n\|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\|^2} = O(1). \qquad \text{(A24)}$$

Combining the two parts, we will get

$$\begin{aligned}\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\le\gamma_n}\frac{|R_n(\boldsymbol{\phi},\boldsymbol{\theta})|}{1+n\|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\|^2} &\le \sup_{\boldsymbol{\rho}\in\mathbf{P}:\|\boldsymbol{\rho}-(\boldsymbol{\phi}^*,\mathbf{0})\|\le\gamma_n}\frac{|R_n(\boldsymbol{\phi},\boldsymbol{\theta})|}{h_n(\boldsymbol{\phi},\boldsymbol{\theta})}\\ &\quad\times\sup_{\boldsymbol{\rho}\in\mathbf{P}:\|\boldsymbol{\rho}-(\boldsymbol{\phi}^*,\mathbf{0})\|\le\gamma_n}\frac{h_n(\boldsymbol{\phi},\boldsymbol{\theta})}{1+n\|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\|^2}\\ &= o_p(1)O(1) = o_p(1).\end{aligned}$$

Let us now prove those two parts in detail. Regarding the first one, a $2r^{th}$-order Taylor expansion of $L_n\left(\boldsymbol{\phi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_r\right)$ around the $(\boldsymbol{\phi}^*,\mathbf{0})$ yields

$$L_n(\boldsymbol{\phi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_r)-L_n(\boldsymbol{\phi}^*,\boldsymbol{\theta})=\sum_{j=1}^{9}A_j+\sum_{j=1}^{17}B_j,$$

where

$$\begin{aligned}
A_1 &= (\boldsymbol{\phi}-\boldsymbol{\phi}^*)'\frac{\partial L_n}{\partial\boldsymbol{\phi}}=(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'S_{\boldsymbol{\phi}n},\\
A_2 &= \frac{1}{2}n\left(\boldsymbol{\phi}-\boldsymbol{\phi}^*\right)^{\otimes 2\prime}E\left[\frac{\partial l}{\partial\boldsymbol{\phi}^{\otimes 2}}\right]=-\frac{1}{2}n\left(\boldsymbol{\phi}-\boldsymbol{\phi}^*\right)'\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}\left(\boldsymbol{\phi}-\boldsymbol{\phi}^*\right),
\end{aligned}$$

$$A_3=\boldsymbol{\theta}_1'\frac{\partial L_n}{\partial\boldsymbol{\theta}_1}=\boldsymbol{\theta}_1'\mathbf{S}_{\boldsymbol{\theta}_1 n},\;\; A_4=\frac{1}{2}n\left(\boldsymbol{\theta}_1^{\otimes 2}\right)'E\left[\frac{\partial l}{\partial\boldsymbol{\theta}_1^{\otimes 2}}\right]=-\frac{1}{2}n\boldsymbol{\theta}_1'\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}\boldsymbol{\theta}_1,$$

$$A_5=\frac{1}{r!}\left(\boldsymbol{\theta}_r^{\otimes r}\right)'\frac{\partial L_n}{\partial\boldsymbol{\theta}_r^{\otimes r}}=\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r)\mathbf{S}_{\boldsymbol{\phi}n}+\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)\mathbf{S}_{\boldsymbol{\theta}_1 n}+\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)\mathcal{S}_{\boldsymbol{\theta}_r n},$$

$$A_6=\frac{1}{(2r)!}n\left(\boldsymbol{\theta}_r^{\otimes 2r}\right)'E\left[\frac{\partial l}{\partial\boldsymbol{\theta}_r^{\otimes j}}\right]=-\frac{1}{2}n[\boldsymbol{\lambda}_{\boldsymbol{\phi}}'(\boldsymbol{\theta}_r),\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}'(\boldsymbol{\theta}_r),\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}'(\boldsymbol{\theta}_r)]\mathcal{I}[\boldsymbol{\lambda}_{\boldsymbol{\phi}}'(\boldsymbol{\theta}_r),\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}'(\boldsymbol{\theta}_r),\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}'(\boldsymbol{\theta}_r)]',$$

$$A_7=n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'E\left[\frac{\partial^2 l}{\partial\boldsymbol{\phi}\partial\boldsymbol{\theta}_1'}\right]\boldsymbol{\theta}_1=-n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}_1}\boldsymbol{\theta}_1,$$

$$A_8=\frac{1}{r!}n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'E\left[\frac{\partial^{1+r}l}{\partial\boldsymbol{\phi}\partial\boldsymbol{\theta}_r^{\otimes r\prime}}\right]\boldsymbol{\theta}_r^{\otimes r}=-n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'[\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r)+\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}_1}\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)],$$

$$A_9=\frac{1}{r!}n\boldsymbol{\theta}_1'E\left[\frac{\partial^{1+r}l}{\partial\boldsymbol{\phi}\partial\boldsymbol{\theta}_r^{\otimes r\prime}}\right]\boldsymbol{\theta}_r^{\otimes r}=-n\boldsymbol{\theta}_1'[\mathcal{I}_{\boldsymbol{\theta}_1\phi}\boldsymbol{\lambda}_{\phi}(\boldsymbol{\theta}_r)+\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)],$$

$$B_1=\frac{1}{2}n\left(\boldsymbol{\phi}-\boldsymbol{\phi}^*\right)^{\otimes 2\prime}\left(\frac{1}{n}\frac{\partial L_n}{\partial\boldsymbol{\phi}^{\otimes 2}}-E\left[\frac{\partial l}{\partial\boldsymbol{\phi}^{\otimes 2}}\right]\right),\;\; B_2=\sum_{j=3}^{2r}\frac{1}{j!}n\left(\boldsymbol{\phi}-\boldsymbol{\phi}^*\right)^{\otimes j\prime}\left\{\frac{1}{n}\frac{\partial^j L_n}{\partial\boldsymbol{\theta}_1^{\otimes j}}\right\},$$

$$B_3=\frac{1}{2}n\left(\boldsymbol{\theta}_1^{\otimes 2}\right)'\left(\frac{1}{n}\frac{\partial L_n}{\partial\boldsymbol{\theta}_1^{\otimes 2}}-E\left[\frac{\partial l}{\partial\boldsymbol{\theta}_1^{\otimes 2}}\right]\right),\;\; B_4=\sum_{j=3}^{2r}\frac{1}{j!}n\left(\boldsymbol{\theta}_1^{\otimes j}\right)'\left\{\frac{1}{n}\frac{\partial^j L_n}{\partial\boldsymbol{\theta}_1^{\otimes j}}\right\},$$

$$B_5=\sum_{j=r+1}^{2r-1}\frac{1}{j!}\sqrt{n}\left(\boldsymbol{\theta}_r^{\otimes j}\right)'\left\{\frac{1}{\sqrt{n}}\frac{\partial L_n}{\partial\boldsymbol{\theta}_r^{\otimes j}}\right\},\;\; B_6=\frac{1}{(2r)!}n\left(\boldsymbol{\theta}_r^{\otimes 2r}\right)'\left(\frac{1}{n}\frac{\partial^{2r}L_n}{\partial\boldsymbol{\theta}_r^{\otimes 2r}}-E\left[\frac{\partial l}{\partial\boldsymbol{\theta}_r^{\otimes j}}\right]\right),$$

$$B_7=\sum_{j_1+j_2=3,j_1,j_2\geq 1}^{8}\frac{1}{j_1!j_2!}n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)^{\otimes j_1\prime}\left\{\frac{1}{n}\frac{\partial^{j_1+j_2}L_n}{\partial\boldsymbol{\phi}^{\otimes j_1}\partial\boldsymbol{\theta}_1^{\otimes j_2\prime}}\right\}\boldsymbol{\theta}_1^{\otimes j_2},$$

$$B_8=n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'\left(\frac{1}{n}\frac{\partial^2 L_n}{\partial\boldsymbol{\phi}\partial\boldsymbol{\theta}_1'}-E\left[\frac{\partial^2 l}{\partial\boldsymbol{\phi}\partial\boldsymbol{\theta}_1'}\right]\right)\boldsymbol{\theta}_1,$$

$$\begin{aligned}
B_9 &= \frac{1}{r!}n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'\left(\frac{1}{n}\frac{\partial^{1+r}L_n}{\partial\boldsymbol{\phi}^1\partial\boldsymbol{\theta}_r^{\otimes r\prime}}-E\left[\frac{\partial^{1+r}l}{\partial\boldsymbol{\phi}^1\partial\theta_r^{\otimes r\prime}}\right]\right)\boldsymbol{\theta}_r^{\otimes r},\\
B_{10} &= \sum_{j=r+1}^{2r}\frac{1}{j!}n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)'\left\{\frac{1}{n}\frac{\partial^{1+j}L_n}{\partial\boldsymbol{\phi}\partial\boldsymbol{\theta}_r^{\otimes j\prime}}\right\}\boldsymbol{\theta}_r^{\otimes j},
\end{aligned}$$

$$B_{11} = \sum_{j_1+j_2=3, j_1\geq 2, j_2\geq 1}^{8} \frac{1}{j_1!j_2!} n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)^{\otimes j_1 \prime} \left\{ \frac{1}{n} \frac{\partial^{j_1+j_2} L_n}{\partial \boldsymbol{\phi}^{\otimes j_1} \partial \boldsymbol{\theta}_1^{\otimes j_2 \prime}} \right\} \boldsymbol{\theta}_r^{\otimes j_2},$$

$$B_{12} = \sum_{j=1}^{r-1} \frac{1}{j!} \sqrt{n} \boldsymbol{\theta}_1' \left\{ \frac{1}{\sqrt{n}} \frac{\partial^{1+j} L_n}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes j \prime}} \right\} \boldsymbol{\theta}_r^{\otimes j}, \quad B_{13} = \frac{1}{r!} n \boldsymbol{\theta}_1' \left(\frac{1}{n} \frac{\partial^{1+r} L_n}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes r \prime}} - E\left[\frac{\partial^{1+r} l}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes r \prime}} \right] \right) \boldsymbol{\theta}_r^{\otimes r},$$

$$B_{14} = \sum_{j=r+1}^{2r} \frac{1}{j!} n \boldsymbol{\theta}_1' \left\{ \frac{1}{n} \frac{\partial^{1+j} L_n}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes j \prime}} \right\} \boldsymbol{\theta}_r^{\otimes j},$$

$$B_{15} = \sum_{j_1+j_2=3, j_1\geq 2, j_2\geq 1}^{8} \frac{1}{j_1!j_2!} n \boldsymbol{\theta}_1^{\otimes j_1 \prime} \left\{ \frac{1}{n} \frac{\partial^{j_1+j_2} L_n}{\partial \boldsymbol{\phi}^{\otimes j_1} \partial \boldsymbol{\theta}_1^{\otimes j_2 \prime}} \right\} \boldsymbol{\theta}_r^{\otimes j_2},$$

$$B_{16} = \sum_{j_1+j_2+j_3=3, j_{1,2,3}\geq 1}^{8} \left\{ \frac{1}{n} L_n^{[j_1,j_2,j_3]} \right\} n \boldsymbol{\phi}_1^{j_1} \boldsymbol{\theta}^{j_2} \boldsymbol{\theta}_r^{j_3}, \text{ and}$$

$$B_{17} = \sum_{j_1+j_2+j_3=3, j_{1,2,3}\geq 1}^{8} \left(\frac{1}{n} L_n^{[j_1,j_2,j_3]}(\bar{\boldsymbol{\rho}}) - \frac{1}{n} L_n^{[j_1,j_2,j_3]} \right) n(\boldsymbol{\phi}-\boldsymbol{\phi}^*)^{j_1} \boldsymbol{\theta}_1^{j_2} \boldsymbol{\theta}_r^{j_3},$$

with the omitted argument above being either $\boldsymbol{\phi}^*$ or $(\boldsymbol{\phi}^*, \mathbf{0})$. The simplification of A_2, A_4 and A_7 is based on the information matrix equality, while we have used Corollary 1 in Rotnitzky et al (2000) to obtain A_6, A_8, and A_9. It is also easy to see that $\sum_{j=1}^{9} A_j = \frac{1}{2} LM_n(\boldsymbol{\theta})$ because of the definition $R_n(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum B_j$. We can then verify that

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in \mathbf{P}: \|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\| \leq \gamma_n} \frac{|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})|}{h_n(\boldsymbol{\phi}, \boldsymbol{\theta})} = o_p(1)$$

by noting that the expressions in curly brackets in the B_j terms are $O_p(1)$, those inside parentheses are $o_p(1)$, and $(\boldsymbol{\phi}-\boldsymbol{\phi}^*, \boldsymbol{\theta}_1, \boldsymbol{\theta}_r) = o(1)$.

Further, note that if $h_n(\boldsymbol{\phi}, \boldsymbol{\theta}) = O(1)$, then

$$\frac{|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})|}{h_n(\boldsymbol{\phi}, \boldsymbol{\theta})} = O_p(n^{-\frac{1}{2r}}) \tag{A25}$$

because $(\boldsymbol{\phi}-\boldsymbol{\phi}^*, \boldsymbol{\theta}_1) = O(n^{-\frac{1}{2}})$ and $\boldsymbol{\theta}_r = O(n^{-\frac{1}{2r}})$.

To verify the second part, let

$$\pi_{\boldsymbol{\phi}} = \max_{||\boldsymbol{v}||=1} \|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v})\|, \quad \pi_{\boldsymbol{\theta}_1} = \max_{||\boldsymbol{v}||=1} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{v})\| \quad \text{and} \quad \pi_r = \min_{||\boldsymbol{v}||=1} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{v})\| > 0, \tag{A26}$$

where the last inequality follows from (i) $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{v})$ is a continuous function, and (ii) $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{v}) \neq \mathbf{0}$ for all $\boldsymbol{v} \neq \mathbf{0}$. In this context, to verify (A24) it suffices to check that

$$\max_{(\boldsymbol{\phi},\boldsymbol{\theta})\in \mathbf{P}: \|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\| \leq \gamma_n} \frac{h_n^{\pi}(\boldsymbol{\phi}, \boldsymbol{\theta})}{1 + ||n^{\frac{1}{2}} \boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})||^2} = O(1), \tag{A27}$$

with

$$h_n^{\pi}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \max\left\{1, \pi_1 n \|\boldsymbol{\phi}-\boldsymbol{\phi}^*\|^2, \pi_2 n \|\boldsymbol{\theta}_1\|^2, n \|\boldsymbol{\theta}_r\|^{2r}\right\},$$

where the coefficients

$$\pi_1 = \frac{1}{2\pi_{\boldsymbol{\phi}} + 1} > 0 \quad \text{and} \quad \pi_2 = \frac{1}{2\pi_{\boldsymbol{\theta}_1} + 1} > 0$$

are only used to simplify the expressions. Thus, for n large enough, we will have that

$$\{(\boldsymbol{\phi}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_r) : \|(\boldsymbol{\phi}, \boldsymbol{\theta}) - (\boldsymbol{\phi}^*, \mathbf{0})\| \leq \gamma_n\} \subset \mathbf{P}.$$

The compactness of $\{(\boldsymbol{\phi}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_r) : \|(\boldsymbol{\phi}, \boldsymbol{\theta}) - (\boldsymbol{\phi}^*, \mathbf{0})\| \leq \gamma_n\}$ and the continuity of $\frac{h_n^\pi(\boldsymbol{\phi},\boldsymbol{\theta})}{1+n\|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\|^2}$ implies that there exists $(\boldsymbol{\phi}, \boldsymbol{\theta})$ such that

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n} \frac{h_n^\pi(\boldsymbol{\phi}, \boldsymbol{\theta})}{1 + n\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\|^2} = \frac{h_n^\pi(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)}{1 + n\|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\|^2} \tag{A28}$$

for all large enough n. Consequently, there will exist a subsequence $\{w_n\}$ of $\{n\}$ such that

$$\begin{aligned}\limsup_{n\to\infty} \sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi},\mathbf{0})\|\leq\gamma_n} \frac{h_n^\pi(\boldsymbol{\phi}, \boldsymbol{\theta})}{1 + n\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\|^2} &= \lim_{n\to\infty} \sup \frac{h_n^\pi(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)}{1 + n\|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\|^2} \\ &= \lim_{w_n\to\infty} \frac{h_{w_n}^\pi(\boldsymbol{\phi}_{w_n}, \boldsymbol{\theta}_{w_n})}{1 + w_n\|\boldsymbol{\lambda}(\boldsymbol{\phi}_{w_n}, \boldsymbol{\theta}_{w_n})\|^2},\end{aligned}$$

where the first equality follows directly from (A28) and the second one by the properties of lim sup. Consequently, it is easy to see that if $h_{w_n}^\pi(\boldsymbol{\phi}_{w_n}, \boldsymbol{\theta}_{w_n}) = O(1)$, then (A24) holds trivially. In turn, if $h_{w_n}^\pi(\boldsymbol{\phi}_{w_n}, \boldsymbol{\theta}_{w_n}) \neq O(1)$, then we can find a further subsequence $\{u_n\}$ of $\{w_n\}$ such that at least one of the following conditions holds:

$$h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) = u_n \|\boldsymbol{\theta}_{r,u_n}\|^{2r} \to \infty, \tag{A29}$$

$$h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) = \pi_1^2 u_n \|\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*\|^2 \to \infty, \quad \text{or} \tag{A30}$$

$$h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) = \pi_2^2 u_n \|\boldsymbol{\theta}_{1,u_n}\|^2 \to \infty. \tag{A31}$$

Let $\boldsymbol{\theta}_{r,n} = \eta_n \boldsymbol{v}_n$ with $\|\boldsymbol{v}_n\| = 1$ and η_n a scalar. If (A29) holds, then

$$\frac{h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})}{1 + u_n\|\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})\|^2} \leq \frac{u_n \|\boldsymbol{\theta}_{r,n}\|^{2r}}{u_n \|\boldsymbol{\lambda}_{\boldsymbol{\theta}^r}(\boldsymbol{\theta}_{r,n})\|^2} = \frac{u_n \eta_n^{2r}}{u_n \|\eta_n^r \boldsymbol{\lambda}_{\boldsymbol{\theta}^r}(\boldsymbol{v}_n)\|^2} = \frac{1}{\|\boldsymbol{\lambda}_{\boldsymbol{\theta}^r}(\boldsymbol{v}_n)\|^2} \leq \frac{1}{\pi_r^2},$$

where the first inequality follows from

$$h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) = u_n \|\boldsymbol{\theta}_{r,n}\|^{2r} \quad \text{and} \quad u_n\|\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})\|^2 \geq u_n \|\boldsymbol{\lambda}_{\boldsymbol{\theta}^r}(\boldsymbol{\theta}_{r,n})\|^2,$$

the second equality follows from the definition of $\boldsymbol{\lambda}_{\boldsymbol{\theta}^r}$, and the last inequality follows from the characterization of π_r in (A26).

If (A30) holds, then

$$h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) = \max\{1, \pi_1 u_n\|\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*\|^2, \pi_2 u_n \|\boldsymbol{\theta}_{1u_n}\|^2, u_n \|\boldsymbol{\theta}_{ru_n}\|^{2r}\} = u_n \pi_1^2 \|\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*\|^2$$

will imply that

$$\pi_1^2||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*||^2 \geq \|\boldsymbol{\theta}_{ru_n}\|^{2r} = \eta_{u_n}^{2r} \Rightarrow \pi_1||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*|| \geq \eta_{u_n}^r, \tag{A32}$$

which in turn yields

$$\begin{aligned}\|\boldsymbol{\phi}_n - \boldsymbol{\phi}^* + \eta_n^r \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_n)\| &\geq |\|\boldsymbol{\phi}_n - \boldsymbol{\phi}^*\| - \eta_n^r \|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_n)\|| = \|\boldsymbol{\phi}_n - \boldsymbol{\phi}^*\| \left|1 - \frac{\eta_n^r}{\|\boldsymbol{\phi}_n - \boldsymbol{\phi}^*\|} \|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_n)\|\right| \\ &\geq \|\boldsymbol{\phi}_n - \boldsymbol{\phi}^*\| \, |1 - \pi_1 \pi_{\boldsymbol{\phi}}| > \frac{1}{2} \|\boldsymbol{\phi}_n - \boldsymbol{\phi}^*\|, \end{aligned} \tag{A33}$$

where the first line follows from triangle inequality and the second one from $\frac{\eta_n^r}{\|\boldsymbol{\phi}_n - \boldsymbol{\phi}^*\|} \leq \pi_1$ in view of (A32) and $\|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_n)\| \leq \pi_{\boldsymbol{\phi}}$ because of (A26). Then, we will have that

$$\frac{h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})}{1 + u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})||^2} < \frac{\pi_1||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*||^2}{||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^* + \eta_{u_n}^r \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_{u_n})||^2} \leq \frac{\pi_1||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*||^2}{\frac{1}{2}||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*||^2} = 2\pi_1,$$

where the first inequality follows from $u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})||^2 > u_n||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^* + \eta_{u_n}^r \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_{u_n})||^2$ and $h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) \leq \pi_1||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*||^2$, while the second one from (A33).

Similarly, if (A31) holds, then we will have that

$$\pi_2^2 \|\boldsymbol{\theta}_{1u_n}\|^2 \geq \|\boldsymbol{\theta}_{ru_n}\|^{2r} = \eta_{u_n}^{2r} \quad \text{implies} \quad \pi_2 \|\boldsymbol{\theta}_{1u_n}\| \geq \eta_{u_n}^r, \tag{A34}$$

whence

$$\begin{aligned}\|\boldsymbol{\theta}_{1u_n} + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_{ru_n})\| &\geq \left|\|\boldsymbol{\theta}_{1u_n}\| - \eta_{u_n}^r \|\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\upsilon}_{u_n})\|\right| \\ &= \|\boldsymbol{\theta}_{1u_n}\| \left|1 - \frac{\eta_{u_n}^r}{\|\boldsymbol{\theta}_{1u_n}\|} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\upsilon}_{u_n})\|\right| \\ &\geq \|\boldsymbol{\theta}_{1u_n}\| \, |1 - \pi_2 \pi_{\boldsymbol{\theta}_1}| > \frac{1}{2} \|\boldsymbol{\theta}_{1u_n}\|, \end{aligned} \tag{A35}$$

where the first two inequalities are straightforward, and the third one follows from (A26) and (A34). In addition, we can show that

$$\frac{h_{u_n}^\pi(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})}{1 + u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})||^2} < \frac{\pi_2^2 \|\boldsymbol{\theta}_{1u_n}\|^2}{\|\boldsymbol{\theta}_{1u_n} + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_{ru_n})\|^2} \leq \frac{\pi_2^2 \|\boldsymbol{\theta}_{1u_n}\|^2}{\frac{1}{2} \|\boldsymbol{\theta}_{1u_n}\|^2} = 2\pi_2^2, \tag{A36}$$

where the first inequality follows from (A31) and the second one from (A35).

The previous argument also implies that if $h_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) \to \infty$, then $h_n^\pi(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) \to \infty$ and $n \|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\| \to \infty$. Consequently,

$$n^{\frac{1}{2}} \|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\| = O(1) \Rightarrow h_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1). \tag{A37}$$

Regarding Assumption 4.5, if $n^{\frac{1}{2}} \boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$, then we have $h_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$ in view of (A37), which in turn implies $|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})| = O_p(n^{-\frac{1}{2r}})$ thanks to (A25).

But then, Theorem 2 implies that

$$
\begin{aligned}
LR &= 2\left[L_n(\boldsymbol{\phi}_n,\boldsymbol{\theta}_n) - L_n(\boldsymbol{\phi}_n,\mathbf{0})\right] \\
&= \sup_{\boldsymbol{\theta}} \left\{ 2[\mathbf{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}\mathbf{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'\mathcal{V}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \right\} + O_p(n^{-\frac{1}{2r}}),
\end{aligned}
$$

where

$$
\mathcal{V}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}} = \begin{bmatrix} I_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} - I_{\boldsymbol{\theta}_1\boldsymbol{\phi}}I_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}I_{\boldsymbol{\phi}\boldsymbol{\theta}_1} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r} \end{bmatrix} = \begin{bmatrix} V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r} \end{bmatrix}
$$

and $\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = [\boldsymbol{\theta}_1' + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}'(\boldsymbol{\theta}_r), \boldsymbol{\lambda}_{\boldsymbol{\theta}_r}'(\boldsymbol{\theta}_r)]'$. Hence, it is not difficult to see that $\mathbf{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}}) = \mathbf{0}$.

Next, rearranging terms we get

$$
\begin{aligned}
2\mathbf{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})_{\boldsymbol{\theta}}'\boldsymbol{\lambda}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'\mathcal{V}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) =& 2\mathbf{S}_{\boldsymbol{\theta}_1,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta})'V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}) \\
&+ 2\mathbf{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})_{\boldsymbol{\theta}_r}'\boldsymbol{\lambda}(\boldsymbol{\theta}_r) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)'\mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r),
\end{aligned}
$$

where $\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}) = \boldsymbol{\theta}_1 + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)$. Thus, we will have

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}} \left\{ 2\mathbf{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'\mathcal{V}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \right\} \\
=&\sup_{\boldsymbol{\theta}_r}\sup_{\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta})} \left\{ 2\mathbf{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'\mathcal{V}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \right\} \quad \text{w.p.a. 1} \\
=&\frac{1}{n}\mathbf{S}_{\boldsymbol{\theta}_1,n}(\tilde{\boldsymbol{\phi}})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}(\tilde{\boldsymbol{\phi}})\mathbf{S}_{\boldsymbol{\theta}_1,n}(\tilde{\boldsymbol{\phi}}) \\
&+ \sup_{\boldsymbol{\theta}_r} \left\{ 2\mathbf{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)'\mathcal{V}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) \right\} \quad \text{w.p.a. 1.}
\end{aligned}
$$

To further simplify the last sup, let $\boldsymbol{\theta}_r = \eta\boldsymbol{\upsilon}$ with $\eta \geq 0$ and $\|\boldsymbol{\upsilon}\| = 1$. Then,

$$
\begin{aligned}
&\sup_{\eta,\boldsymbol{\upsilon}} \left\{ 2\mathbf{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)'V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) \right\} \\
=& \sup_{||\boldsymbol{\upsilon}||=1}\sup_{\eta\geq 0} \left\{ 2\mathbf{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})\eta^r - n\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})'V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})\eta^{2r} \right\} \quad \text{w.p.a. 1} \\
=& \begin{cases} \frac{1}{n}\sup_{||\boldsymbol{\upsilon}||=1} \frac{[\mathcal{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})]^2}{\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})'\mathcal{V}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})} & \text{if } r \text{ is odd} \\ \frac{1}{n}\sup_{||\boldsymbol{\upsilon}||=1} \frac{\lfloor \mathcal{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon}) \rfloor_+^2}{\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})'\mathcal{V}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})} & \text{if } r \text{ is even} \end{cases}
\end{aligned}
$$

Finally, noticing that

$$
\mathcal{S}_{\boldsymbol{\theta}_r,n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon}) = r!\boldsymbol{\theta}_r^{\otimes r\prime}D_{rn}(\tilde{\boldsymbol{\phi}})
$$

and

$$
\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon})'\mathcal{V}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\upsilon}) = (r!)^2\,\boldsymbol{\theta}_r^{\otimes r\prime}[V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) - V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi})]\boldsymbol{\theta}_r^{\otimes r},
$$

we can finally see that Theorem 1 holds. □

Proof of Theorem 3

By Le Cam's first Lemma (see Lemma 6.4 of van der Vaart (1998)), contiguity holds if under P_0, $dP_{\boldsymbol{\phi},\boldsymbol{\theta}_n}/dP_0 \xrightarrow{d} U$ with $E(U) = 1$. Let $L_n(\boldsymbol{\phi}^*, \boldsymbol{\theta}_n)$ denote the log of the joint likelihood of the observations. Given Assumption 4, we can write

$$\begin{aligned} L_n(\boldsymbol{\phi}^*, \boldsymbol{\theta}_n) - L_n(\boldsymbol{\phi}^*, \mathbf{0}) &= \frac{1}{\sqrt{n}}\mathcal{S}_n'(\boldsymbol{\phi}^*)\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\phi}^*, \boldsymbol{\theta}_n) - \frac{1}{2}\sqrt{n}\boldsymbol{\lambda}'(\boldsymbol{\phi}^*, \boldsymbol{\theta}_n)\mathcal{I}(\boldsymbol{\phi}^*)\sqrt{n}\boldsymbol{\lambda}(\boldsymbol{\phi}^*, \boldsymbol{\theta}_n) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\mathcal{S}_n'(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty - \frac{1}{2}\boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty + o_p(1). \end{aligned}$$

Therefore, under H_0,

$$\frac{dP_{\boldsymbol{\theta}_n}}{dP_0} = \exp\left\{\frac{1}{\sqrt{n}}\mathcal{S}_n'(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty - \frac{1}{2}\boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty\right\} + o_p(1) \xrightarrow{d} U = \exp\left\{S - \frac{1}{2}\boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty\right\},$$

where $S \sim \mathcal{N}[0, \boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty]$. Using the expression of the moment generating function of a normal distribution, we have that $E(U) = 1$. The joint distribution of $\mathcal{S}_n$ and $\ln\left(\frac{dP_{\boldsymbol{\theta}_n}}{dP_0}\right)$ converges under H_0 to the Gaussian process:

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\mathcal{S}_n(\boldsymbol{\phi}^*) \\ \ln\left(\frac{dP_{\boldsymbol{\theta}_n}}{dP_0}\right) \end{bmatrix} \xrightarrow{d} N\left\{\begin{bmatrix} \mathbf{0} \\ -\frac{1}{2}\boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty \end{bmatrix}, \begin{pmatrix} \mathcal{I}(\boldsymbol{\phi}^*) & \mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty \\ \boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*) & \boldsymbol{\lambda}_\infty'\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty \end{pmatrix}\right\}.$$

In addition, it follows from Le Cam's third lemma (see van der Vaart (1998)) that

$$\frac{1}{\sqrt{n}}\mathcal{S}_n(\boldsymbol{\phi}^*) \xrightarrow{d} N\left[\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_\infty, \mathcal{I}(\boldsymbol{\phi}^*)\right]$$

under $P_{\boldsymbol{\theta}_n}$.

Finally, given Assumption 5, we can then prove that under $P_{\boldsymbol{\theta}_n}$,

$$\begin{aligned} GET_n = &\sup_{\boldsymbol{\theta}}\left\{2\left[\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}\left(\tilde{\phi}_n\right) - \frac{1}{\sqrt{n}}\mathcal{I}_{\boldsymbol{\theta}\phi}\left(\tilde{\phi}_n\right)\mathcal{I}_{\boldsymbol{\phi}\phi}^{-1}\left(\tilde{\phi}_n\right)\mathcal{S}_{\phi,n}\left(\tilde{\phi}_n\right)\right]'\sqrt{n}\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right. \\ &\left. -n\boldsymbol{\lambda}_{\boldsymbol{\theta}}'(\boldsymbol{\theta})\left[\mathcal{I}_{\boldsymbol{\theta\theta}}\left(\tilde{\phi}_n\right) - \mathcal{I}_{\boldsymbol{\theta}\phi}\left(\tilde{\phi}_n\right)\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}\left(\tilde{\phi}_n\right)\mathcal{I}_{\phi\boldsymbol{\theta}}\left(\tilde{\phi}_n\right)\right]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\} \\ \xrightarrow{d} &\sup_{\boldsymbol{\lambda}\in\Lambda}\left\{2\left[S + \left(\mathcal{I}_{\theta\theta}\left(\boldsymbol{\phi}^*\right) - \mathcal{I}_{\boldsymbol{\theta}\phi}\left(\boldsymbol{\phi}^*\right)\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}\left(\boldsymbol{\phi}^*\right)\mathcal{I}_{\phi\theta}\left(\boldsymbol{\phi}^*\right)\right)\boldsymbol{\lambda}_{\theta,\infty}\right]'\boldsymbol{\lambda}\right. \\ &\left. -\boldsymbol{\lambda}\left[\mathcal{I}_{\boldsymbol{\theta\theta}}\left(\boldsymbol{\phi}^*\right) - \mathcal{I}_{\boldsymbol{\theta}\phi}\left(\boldsymbol{\phi}^*\right)\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}\left(\boldsymbol{\phi}^*\right)\mathcal{I}_{\boldsymbol{\phi\theta}}\left(\boldsymbol{\phi}^*\right)\right]\boldsymbol{\lambda}\right\} \end{aligned}$$

where

$$S \sim \mathcal{N}\left[\mathbf{0}, \mathcal{I}_{\boldsymbol{\theta\theta}}\left(\boldsymbol{\phi}^*\right) - \mathcal{I}_{\boldsymbol{\theta\phi}}\left(\boldsymbol{\phi}^*\right)\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}\left(\boldsymbol{\phi}^*\right)\mathcal{I}_{\boldsymbol{\phi\theta}}\right],$$

as desired. □

Proof of Proposition 1

We first reparametrize the model as follows:

$$
\begin{aligned}
\varphi^S &= \phi^S, \\
\varphi_k^M &= \boldsymbol{\phi}_k^M - \boldsymbol{\phi}_k^D M_1(\phi^S)\theta_k \mathbf{e}_1, \\
(\varphi_k^D)^2 &= (\phi_k^D)^2\{1 + M_1(\phi^S)[M_1(\phi^S) + \phi^S]\theta_k^2\}, \\
\varphi_{kj}^L &= \phi_{kj}^L - \frac{1}{2}M_1(\phi^S)\left(\theta_k^2\theta_j + \theta_k\theta_j^2 - 2\theta_k\theta_j\right)[M_1(\phi^S) + \phi^S], \quad \text{and} \\
\boldsymbol{\vartheta} &= \boldsymbol{\theta},
\end{aligned}
$$

where $\mathbf{e}_1 = (1, \mathbf{0}')'$ is the first vector of the canonical basis of $\mathbb{R}^{\dim(\boldsymbol{\varphi}_k)}$, φ_{kh}^L the correlation coefficient between u_k and u_h and $M_1(\phi^S)$ the Mills ratio defined in (9).

Next, letting

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

where $u_k(\boldsymbol{\varphi}_k^M, \varphi_k^D) = (y_k - \boldsymbol{\varphi}_k^{M\prime}\mathbf{x})/\varphi_k^D$ and $\mathbf{r}_{(k)}(\boldsymbol{\phi}^L)$ 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

$$
\begin{aligned}
\frac{\partial l}{\partial \phi^S} &= \mathbf{w}\phi_N(\boldsymbol{\phi}^{S\prime}\mathbf{w})\left[\frac{d}{\Phi_N(\boldsymbol{\phi}^{S\prime}\mathbf{w})} - \frac{1-d}{\Phi_N(-\boldsymbol{\phi}^{S\prime}\mathbf{w})}\right] \\
\frac{\partial l}{\partial \phi_k^M} &= \frac{\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]}{\phi_k^D \det\left[\mathbf{R}(\boldsymbol{\phi}^L)\right]} dx\varepsilon_k \\
\frac{\partial l}{\partial \phi_k^D} &= d\left[a_{kk}\left(u_k^2 - 1\right) + \sum_{h \neq k} a_{kh}\left(u_k u_h - \varphi_{kh}^L\right)\right] \\
\frac{\partial l}{\partial \phi_{kj}^L} &= d\left[\sum_h b_{kj,h}\left(u_h^2 - 1\right) + \sum_{h \neq i} b_{kj,ih}\left(u_i u_h - \varphi_{ih}^L\right)\right] \\
\frac{\partial l}{\partial \theta_k} &= 0 \\
\frac{\partial^2 l}{\partial \theta_k \partial \theta_j} &= 0 \\
\frac{\partial^3 l}{\partial \theta_k^3} &= Cd\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]^3 \varepsilon_k^3 + A_k\frac{\partial l}{\partial \phi} \\
\frac{\partial^3 l}{\partial \theta_k^2 \partial \theta_j} &= Cd\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]^2 \det\left[\mathbf{R}_{(j)}(\boldsymbol{\phi}^L)\right]\varepsilon_k^2\varepsilon_j + A_{kj}\frac{\partial l}{\partial \phi} \\
\frac{\partial^3 l}{\partial \theta_k \partial \theta_j \partial \theta_h} &= Cd\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^L)\right]\det\left[\mathbf{R}_{(j)}(\boldsymbol{\phi}^L)\right]\det\left[\mathbf{R}_{(h)}(\boldsymbol{\phi}^L)\right]\varepsilon_k\varepsilon_j\varepsilon_h + A_{kjh}\frac{\partial l}{\partial \phi},
\end{aligned}
$$

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

$$a_{kh},\ b_{kj,ih}A_{k},\ A_{kj},\ A_{kjh}\ \ \text{for } k,j,h=1,...,K$$

are some terms whose detailed expressions, which are available on request, are irrelevant for the proof.

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

$$\begin{aligned}&\sum_{k}\frac{1}{6}\upsilon_{k}^{\dagger 3}\frac{\partial^{3}l}{\partial\theta_{k}^{3}}+\sum_{j\neq k}\frac{1}{2}\upsilon_{k}^{\dagger 2}\upsilon_{j}^{\dagger}\frac{\partial^{3}l}{\partial\theta_{k}^{2}\partial\theta_{j}}+\sum_{h\neq j\neq k}\upsilon_{k}^{\dagger}\upsilon_{h}^{\dagger}\upsilon_{j}^{\dagger}\frac{\partial^{3}l}{\partial\theta_{k}\partial\theta_{j}\partial\theta_{h}}\\&\propto\left\{\sum_{k}d\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^{L})\right]w_{k}\upsilon_{k}^{\dagger}\right\}^{3}+A^{\dagger}\frac{\partial l}{\partial\boldsymbol{\phi}}\\&\propto dH_{3}\left(\frac{\boldsymbol{\upsilon}'\mathbf{v}}{\sqrt{\boldsymbol{\upsilon}\boldsymbol{\upsilon}}}\right)+A\frac{\partial l}{\partial\boldsymbol{\phi}}\end{aligned}$$

Finally, by suitably choosing $\boldsymbol{\upsilon}$ in the last expression so that

$$\sum_{k}d\det\left[\mathbf{R}_{(k)}(\boldsymbol{\phi}^{L})\right]w_{k}\upsilon_{k}^{\dagger}\propto d\boldsymbol{\upsilon}'\mathbf{v},$$

we can show that the test has form in (10). □

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

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

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

$$\left\{\left[\frac{\boldsymbol{\upsilon}^{\dagger\prime}}{\boldsymbol{\upsilon}^{\dagger\prime}\left[\mathbf{R}\left(\boldsymbol{\phi}^{L}\right)-\boldsymbol{\vartheta}\boldsymbol{\vartheta}'\right]^{-1/2}\boldsymbol{\vartheta}}\mathbf{z}^{\dagger}\right]_{\boldsymbol{\upsilon}^{\dagger}\neq\mathbf{0}},u_{S}\right\}. \tag{A38}$$

Letting

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

we will show that the joint distribution depends only on ν. To do so, first note that $\boldsymbol{\ell}'\boldsymbol{\ell} = 1$, which means that $\mathbf{I} - \boldsymbol{\ell}\boldsymbol{\ell}'$ has rank $K-1$. Therefore, the singular value decomposition implies the existence of a $(K-1)\times K$ matrix $\mathbf{A}$ with full row rank such that

$$\mathbf{A}'\mathbf{A} = \mathbf{I} - \boldsymbol{\ell}'\boldsymbol{\ell}.$$

Letting $\boldsymbol{\upsilon}' = \boldsymbol{\upsilon}^{\dagger\prime}[\ \boldsymbol{\ell} \quad \mathbf{A}\ ]^{-1}$, 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

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

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

If we reparametrize from $(\boldsymbol{\varphi}, \boldsymbol{\vartheta})$ to $(\boldsymbol{\phi}, \boldsymbol{\theta})$ as follows:

$$\begin{aligned}\varphi_1 &= \phi_1 - 2\sqrt{\phi_2}\theta_1 + 2\theta_1^3,\\ \varphi_2 &= \left(1 - 2\sqrt{2}\theta_2 + \frac{2}{3}\theta_1^2\right)\phi_2,\\ \vartheta_1 &= \theta_1, \quad \text{and}\\ \vartheta_2 &= \theta_2 + \frac{\sqrt{2}}{3}\theta_1^2,\end{aligned}$$

then we can show that

$$\frac{\partial l}{\partial \phi_1} = \frac{1}{\sqrt{\varphi_2}} H_1(u), \quad \frac{\partial l}{\partial \phi_2} = \frac{1}{\sqrt{2}\varphi_2} H_2(u),$$
$$\frac{\partial l}{\partial \theta_1} = \frac{\partial l}{\partial \theta_2} = \frac{\partial^2 l}{\partial \theta_1^2} = \frac{\partial^3 l}{\partial \theta_1^3} = 0,$$
$$\frac{1}{2}\frac{\partial^2 l}{\partial \theta_2^2} = -\sqrt{6} H_4(u) - 2\sqrt{2}\frac{\partial l}{\partial \phi_2},$$
$$\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} = -2\sqrt{3} H_3(u), \quad \text{and}$$
$$\frac{1}{4!}\frac{\partial^4 l}{\partial \theta_1^4} = \frac{\sqrt{6}}{9} H_4(u) - \frac{\sqrt{2}}{9}\frac{\partial l}{\partial \phi_2}$$

hold at $(\boldsymbol{\phi}, \mathbf{0})$, where $u = (y - \phi_1)/\sqrt{\phi_2}$. Next, letting $\tilde{u}_i = (y_i - \tilde{\phi}_L)/\sqrt{\tilde{\phi}_V}$ and $\tilde{\mathrm{H}}_j = \sum_{i=1}^{n} H_j(\tilde{u}_i)$, it is easy to see that

$$\tilde{\mathrm{H}}_1 = \tilde{\mathrm{H}}_2 = 0.$$

Then, it holds that

$$LR_n = \sup_{\boldsymbol{\theta} \in \Theta} \left\{ 2S'_{\boldsymbol{\theta},n} \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \right\} + o_p(1) \tag{A39}$$

by virtue of Theorem 2, 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 $\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathbf{I}_2$. Finally, after some tedious calculations available on request, we can verify that the conditions for Theorem 2 are satisfied in this example.

Moreover, in this special case we can further simplify the right-hand side of (A39) as follows. First, it is easy to see that an upper bound will be given by

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

Second, we can construct θ_1 and θ_2 such that

$$\begin{cases} -2\sqrt{3}\sqrt{n}\theta_1\theta_2 = n^{-1/2}\mathrm{H}_3 + o_p(1) \\ -\sqrt{6}\sqrt{n}\theta_2^2 + \frac{\sqrt{6}}{9}\sqrt{n}\theta_1^4 = n^{-1/2}\mathrm{H}_4 + o_p(1) \end{cases} \tag{A40}$$

which implies that a lower bound will be

$$\frac{1}{n}\tilde{\mathrm{H}}_3^2 + \frac{1}{n}\tilde{\mathrm{H}}_4^2 + o_p(1).$$

Therefore, we end up with

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

as desired. □

Proof of Proposition 4

In this example,

$$\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \begin{pmatrix} -2\sqrt{3}\theta_1\theta_2 \\ \sqrt{6}(\frac{1}{9}\theta_1^4 - \theta_2^2) \end{pmatrix}, \quad \boldsymbol{\Lambda} = \mathbb{R}^2,$$

and

$$\mathcal{I}_{\boldsymbol{\theta\theta}}(\tilde{\phi}) - \mathcal{I}_{\boldsymbol{\theta}\phi}(\tilde{\phi})\mathcal{I}^{-1}_{\phi\phi}(\tilde{\phi})\mathcal{I}'_{\boldsymbol{\theta}\phi}(\tilde{\phi}) = \mathbf{I}_2.$$

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

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

as claimed. □

Proof of Proposition 5

The proof is entirely analogous to the proof of Proposition 8 in Amengual, Bei, Carrasco and Sentana (2022), so we omit it for the sake of brevity. □

Table 1: Monte Carlo rejection rates (in %) under null and alternative hypotheses for testing for selectivity in multivariate regression

	Null hypothesis			Alternative hypotheses					
				H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel A: $n = 400$									
GET	1.0	5.0	10.2	8.5	23.2	35.1	8.6	23.9	35.9
LR	0.9	4.9	10.4	9.1	25.2	37.1	9.1	25.2	36.9
GMM	1.0	5.1	10.1	7.6	22.0	32.5	7.8	22.4	33.3
Panel B: $n = 1,600$									
GET	0.8	5.1	9.7	62.2	82.7	88.8	62.7	83.1	89.5
LR	0.9	4.8	9.6	68.0	86.6	91.6	68.9	86.4	91.7
GMM	1.0	5.2	10.0	57.9	79.3	87.5	58.5	79.2	87.6

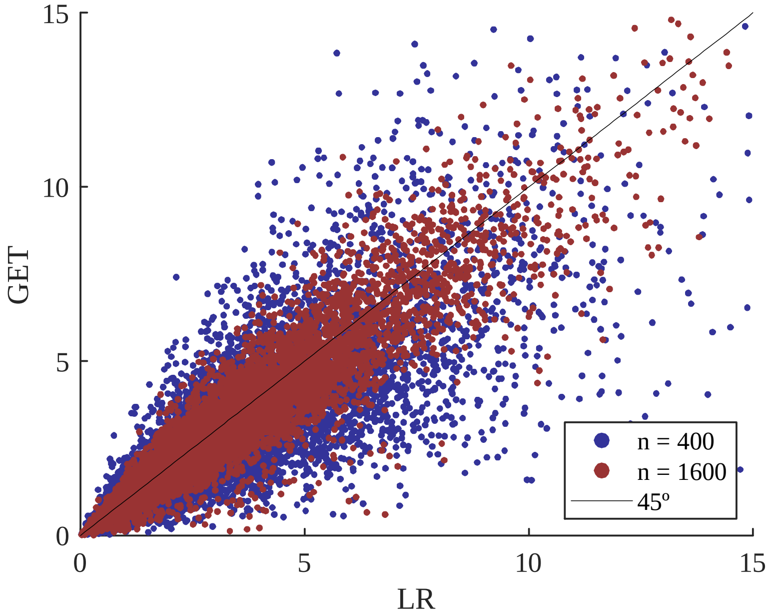
Notes: Results based on 10,000 samples. GET and LR are defined in section 3.1. GMM refers to the J-test based on the influence functions underlying GET. Finite sample critical values are computed by simulation. DGPs: $w = x_1 = 1$ and $x_2 \sim N(0,1)$, $\boldsymbol{\varphi}_k^M = (0,1)$, $\boldsymbol{\varphi}^D = \boldsymbol{\iota}_2$, $\varphi^S = 1$ and $\varphi^L = 0.25$. As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.57, 0.57)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.80, 0)$ (H_{a2}); see section 3.1 for the parametrization.

Table 2: Monte Carlo rejection rates (in %) under alternative hypotheses for testing normality versus SNP

	Alternative hypotheses					
	H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%
Panel A: $n = 400$						
GET	8.8	27.7	39.5	30.2	40.4	46.6
LR	10.6	26.8	39.4	25.0	37.5	45.2
Panel B: $n = 1,600$						
GET	59.5	83.5	89.7	67.8	78.2	82.3
LR	64.3	83.1	89.7	64.7	76.4	82.2

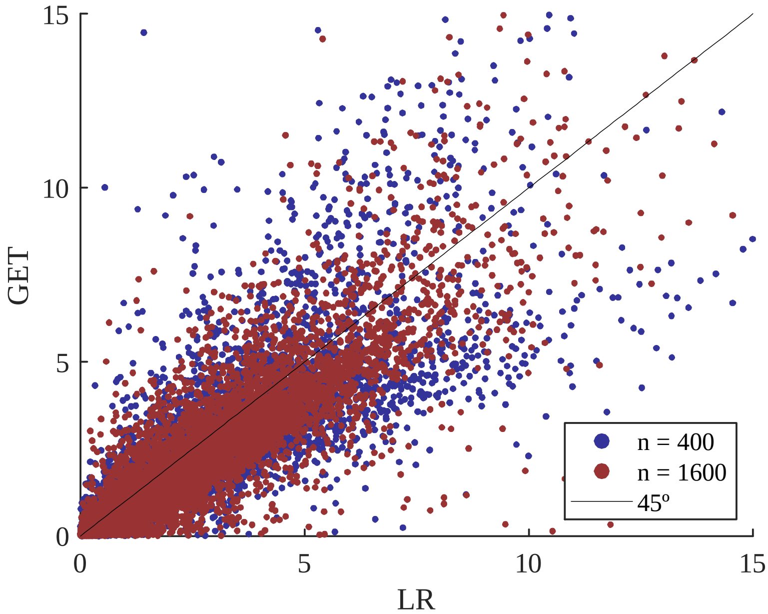
Notes: Results based on 10,000 samples. GET and LR are defined in section 3.2. Finite sample critical values are computed by simulation. DGPs: $\varphi^M = 0$, $\varphi^V = 1$, $\boldsymbol{\vartheta}' = (0.25, 0.10)$ for H_{a1}, and $\boldsymbol{\vartheta}' = (0.75, 0.05)$ for H_{a2}.

Figure 1: Alignment of GET and LR under the null hypothesis when testing for selectivity in multivariate regression



Notes: Results based on 10,000 samples. GET and LR are defined in section 3.1. DGPs: $w = x_1 = 1$ and $x_2 \sim N(0,1)$, $\boldsymbol{\varphi}_k^M = (0,1)$, $\boldsymbol{\varphi}^D = \mathbf{I}_2$, $\varphi^S = 1$ and $\varphi^L = 0.25$

Figure 2: Alignment of GET and LR under the null hypothesis for normality versus SNP test



Notes: Results based on 10,000 samples. GET and LR are defined in section 3.2. DGPs: $\varphi^M = 0$, $\varphi^V = 1$.

Supplemental Appendices for

Hypothesis tests with a repeatedly singular information matrix

Dante Amengual
CEMFI
<amengual@cemfi.es>

Xinyue Bei
Duke University
<xinyue.bei@duke.edu>

Enrique Sentana
CEMFI
<sentana@cemfi.es>

Revised: May 2023

B Reparametrizations

B.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 7 *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) $\left.\frac{\partial^{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}}l}{\partial \boldsymbol{\vartheta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}}\right|_{(\boldsymbol{\varphi},\mathbf{0})} = 0$*, for all index vectors such that* $\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r} < r-1$*.*
3) There exists a set of coefficients $\{m_k^{\mathbf{j}_{\boldsymbol{\theta}_r}}\}_{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}=r-1, k=1,\dots,p-q_r}$ *which may be functions of* $\boldsymbol{\varphi}$ *such that*

$$m_1^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\varphi_1} + ... + m_{p-q}^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\varphi_{p-q}} + m_{p-q+1}^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\vartheta_{11}} + ... + m_{p-q_r}^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\vartheta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}}l}{\partial \boldsymbol{\vartheta}^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = 0$$

for all $\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r} = r-1$*, 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:

$$\varphi_1 = \phi_1 + \sum_{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}=r-1} \frac{m_1^{\mathbf{j}_{\boldsymbol{\theta}_r}}}{\mathbf{j}_{\boldsymbol{\theta}_r}!}\boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}, ..., \varphi_{p-q} = \phi_{p-q} + \sum_{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}=r-1} \frac{m_{p-q}^{\mathbf{j}_{\boldsymbol{\theta}_r}}}{\mathbf{j}_{\boldsymbol{\theta}_r}!}\boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}},$$

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

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

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

$$\frac{\partial^{r-1}l}{\partial \boldsymbol{\theta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = m_1^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\varphi_1} + ... + m_{p-q}^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\varphi_{p-q}} + m_{p-q+1}^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\vartheta_{11}} + ... + m_{p-q_r}^{\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\vartheta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}}l}{\partial \boldsymbol{\vartheta}^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = 0$$

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

Finally, we need to check whether $\sum_{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}=r} \frac{\boldsymbol{\lambda}^{\mathbf{j}_{\boldsymbol{\theta}_r}}}{\mathbf{j}_{\boldsymbol{\theta}_r}!}\frac{\partial^r l}{\partial \boldsymbol{\theta}^{\mathbf{j}_{\boldsymbol{\theta}_r}}}$ evaluated at $(\boldsymbol{\phi}, \mathbf{0})$ is linearly independent of $(\mathbf{s}_{\boldsymbol{\phi}}, \mathbf{s}_{\boldsymbol{\theta}_1})$ for all $\boldsymbol{\lambda}_1^2 + \cdots + \boldsymbol{\lambda}_{q_r}^2 = 1$. If so, Theorem 1 applies. Otherwise, we should check whether either:
1) there exists a new set of coefficients $\{m_k^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}}\}_{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}=r, k=1,\dots,p-q_r}$ which may be functions of $\boldsymbol{\phi}$ such that

$$m_1^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\phi_1} + \ldots + m_{p-q}^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\phi_{p-q}} + m_{p-q+1}^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\theta_{11}} + \ldots + m_{p-r}^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\theta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}'_{qr}\mathbf{j}_{\boldsymbol{\theta}_r}}l}{\partial \theta^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = 0 \qquad \text{(B1)}$$

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 2, which covers far more general cases.

B.2 Numerical invariance to reparametrization

Let us now prove that the GET statistic that we proposed in Theorem 1 is invariant to reparametrization, exactly like the LR test or the usual LM tests that rely on the information matrix rather than the sample average of the Hessian. For simplicity of notation, we will do so in a simple case in which $r = 2$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_2$, so that we can omit the subscript 2 from $\boldsymbol{\theta}$ henceforth. Additionally, we 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 $\boldsymbol{\varphi}$ is $p \times 1$ and $\boldsymbol{\vartheta}$ a $q \times 1$ vector. 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 $\mathbf{s}_{\boldsymbol{\varphi}}$ has full rank, 2) there is a $q \times p$ matrix $\mathbf{M}$ of possible functions of $\boldsymbol{\varphi}$ such that

$$\mathbf{M}\mathbf{s}_{\boldsymbol{\varphi}i}(\boldsymbol{\varphi}, \mathbf{0}) + \mathbf{s}_{\boldsymbol{\vartheta}i}(\boldsymbol{\varphi}, \mathbf{0}) = \mathbf{0} \tag{B2}$$

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

$$\frac{\partial l}{\partial \boldsymbol{\phi}} = \frac{\partial \mathrm{l}}{\partial \boldsymbol{\varphi}}, \tag{B3}$$

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \mathbf{M}\frac{\partial \mathrm{l}}{\partial \boldsymbol{\varphi}} + \frac{\partial \mathrm{l}}{\partial \boldsymbol{\vartheta}} = \mathbf{M}\mathbf{s}_{\boldsymbol{\varphi}i} + \mathbf{s}_{\boldsymbol{\vartheta}i} = \mathbf{0}, \tag{B4}$$

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

In addition, (B3) and (B4) 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. Consequently, we can apply Theorem 1, which yields $\mathrm{GET}_n^{\boldsymbol{\rho}} = \sup_{||\boldsymbol{\upsilon}|| \neq 0} ET_n^{\boldsymbol{\rho}}(\boldsymbol{\upsilon})$, where

$$\begin{aligned} ET_n^{\boldsymbol{\rho}}(\boldsymbol{\upsilon}) &= \frac{[\boldsymbol{\upsilon}'\mathbb{H}(\tilde{\boldsymbol{\varphi}})\boldsymbol{\upsilon}]^2 \, \mathbf{1}\,[\boldsymbol{\upsilon}'\mathbb{H}(\tilde{\boldsymbol{\varphi}})\boldsymbol{\upsilon} \geq \mathbf{0}]}{\mathcal{V}(\boldsymbol{\upsilon}, \tilde{\boldsymbol{\varphi}})}, \\ \mathbb{H}(\boldsymbol{\varphi}) &= \begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}' \left.\frac{\partial^2 \mathrm{l}(\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'}\right|_{(\boldsymbol{\varphi}, \mathbf{0})} \begin{pmatrix} \mathbf{M}' \\ \mathbf{I}_q \end{pmatrix}, \end{aligned} \tag{B5}$$

and

$$\mathcal{V}_\eta(\boldsymbol{\upsilon},\boldsymbol{\varphi}) = V[\boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}] - Cov[\boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}, \mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\varphi})]V^{-1}[\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\varphi})]Cov[\mathbf{s}_{\boldsymbol{\phi}}(\boldsymbol{\varphi}), \boldsymbol{\upsilon}'\mathbb{H}(\boldsymbol{\varphi})\boldsymbol{\upsilon}]$$

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 $\mathbf{g}(\cdot)$ is some second-order continuously differentiable vector of functions which 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 $\mathbf{0}$ at $H_0 : \boldsymbol{\theta}^\dagger = \mathbf{0}$, 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 (B2) in the second one, we need to assume that

$$\det\left(\frac{\partial\mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial\boldsymbol{\phi}^\dagger} - \frac{\partial\mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial\boldsymbol{\phi}^\dagger}\mathbf{M}\right) \neq 0 \tag{B6}$$

for $\partial l/\partial\boldsymbol{\phi}^\dagger$ to have full rank. Similarly, given that (B2) 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

$$\frac{\partial\mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial\boldsymbol{\theta}^\dagger} = \frac{\partial\mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial\boldsymbol{\theta}^\dagger}\mathbf{M} \tag{B7}$$

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, (B7) and (B2) 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, (B5) implies that

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

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

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

In this context, if we further assume that

$$\det\left(\frac{\partial\mathbf{g}^{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}^{\dagger\prime}}\right)\neq 0, \tag{B8}$$

then it is easy to see that $\boldsymbol{\upsilon}'\frac{\partial^2 l}{\partial\boldsymbol{\theta}^\dagger\partial\boldsymbol{\theta}^\dagger}\boldsymbol{\upsilon}$ will be linearly independent of $\mathbf{s}_{\boldsymbol{\phi}^\dagger}$ 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 $\mathbf{s}_{\boldsymbol{\varphi}}$.

In sum, once we guarantee that (B6), (B7) and (B8) 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

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

Then, we will have that

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

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.

C Example 3: Testing Gaussian vs Hermite copulas

C.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 p, which are defined as differentials of the multivariate normal density:

$$H_{\mathbf{j}}(\mathbf{x},\boldsymbol{\varphi}) = f_{NK}(\mathbf{x};\mathbf{R})^{-1}\left(\frac{-\partial}{\partial\mathbf{x}}\right)^{\mathbf{j}} f_{NK}(\mathbf{x};\mathbf{R}), \tag{C9}$$

where $\boldsymbol{\iota}'_K\mathbf{j} = p$ with $\mathbf{j}\in\mathbb{N}^K$, $\boldsymbol{\varphi} = vecl(\mathbf{R})$, 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 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

$$f_H(x_1,x_2;\varphi,\boldsymbol{\vartheta}) = f_{N2}\left[\left(\begin{array}{c} x_1 \\ x_2 \end{array}\right);\left(\begin{array}{cc} 1 & \varphi \\ \varphi & 1 \end{array}\right)\right] P(x_1,x_2;\varphi,\boldsymbol{\vartheta}), \tag{C10}$$

where

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

φ is 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 (C10) 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 $\boldsymbol{\vartheta}=\mathbf{0}$.

We can easily show that the corresponding marginal distributions are given by

$$\left.\begin{array}{c} f_H(x_1;\vartheta_1) = \phi(x_1)[1+\vartheta_1 H_{40}(x_1,x_2)] \\ f_H(x_2;\vartheta_5) = \phi(x_2)[1+\vartheta_5 H_{04}(x_1,x_2)] \end{array}\right\}, \tag{C11}$$

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. Specifically, if $\mathbf{y}=(y_1,y_2)$ denotes the original data, we can define $\mathbf{u}=(u_1,u_2)=[F_1(y_1),F_2(y_2)]$ 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 (C11). When the copula is Gaussian, x_i coincides with the Gaussian rank $\Phi^{-1}(u_i)$.

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

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

C.2 The null hypothesis and the GET test statistic

Straightforward calculations show that in this case

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

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}. In this case, 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, with coefficients that depend on the correlation coefficient, as we explain the next section in detail.

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

$$\frac{1}{n} S_{1n}' V_{11}^{-1} S_{1n} + \frac{1}{n} \sup_{\|\boldsymbol{\upsilon}\|=1} \mathcal{D}_n' (\mathcal{V}_{\eta\eta} - \mathcal{V}_{\eta 1} V_{11}^{-1} \mathcal{V}_{1\eta})^{-1} \mathcal{D}_n \mathbf{1} \left[\mathcal{D}_n > 0\right], \tag{C12}$$

where

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

and the omitted arguments are $(\tilde{\boldsymbol{\phi}}, 0, \boldsymbol{\upsilon})$ for $\mathcal{D}_n$, $(\tilde{\boldsymbol{\phi}}, \boldsymbol{\upsilon})$ for $\mathcal{V}_{\eta\eta}$, $\mathcal{V}_{\eta 1}$ and $\mathcal{V}_{1\eta}$, $(\tilde{\boldsymbol{\phi}}, \mathbf{0})$ for $S_{1,n}$ and $\tilde{\boldsymbol{\phi}}$ for V_{11}.

In this case, the asymptotic distribution of GET_n is bounded above by a χ^2_6 distribution because of the six influence functions. In addition, it is bounded below by a 50:50 mixture of χ^2_3 and χ^2_4 because θ_{11}, θ_{12} and θ_{13} are first-order identified parameters and an even-order derivative of η is involved.

C.3 Computational details

C.3.1 Influence functions

In practice, the calculation of the GET statistic requires explicit expressions for all the different ingredients that appear in (C12). Tedious but straightforward algebra implies that

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

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

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

$$\frac{\partial l}{\partial \theta_{13}} = H_{13}(x_1, x_2; \phi),$$

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

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

and

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

where

$$\mathbf{H}_p(x_1, x_2; \phi) = \left[H_{p0}(x_1, x_2; \phi), H_{p-1,1}(x_1, x_2; \phi), ..., H_{0,p}(x_1, x_2; \phi)\right]'.$$

C.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 $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})'$. Therefore, Assumption 2.1 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, in which case 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 $\boldsymbol{\upsilon} = (\upsilon_1, \upsilon_2)'$. As we show below, 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 (C15). 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

$$\frac{1}{n} S_{1n}' V_{11}^{-1} S_{1n} + \frac{1}{n} \sup_{\omega \in (\omega_l, \omega_u)} \mathcal{D}_n' \left(\mathcal{V}_{\eta\eta} - \mathcal{V}_{\eta 1} \mathcal{V}_{11}^{-1} \mathcal{V}_{1\eta} \right)^{-1} \mathcal{D}_n \mathbf{1} \left[\mathcal{D}_n > 0 \right]. \tag{C13}$$

This test is asymptotically equivalent to the LR test, which implicitly imposes positivity because a zero density gives rise to an infinitely penalized log-likelihood. Nevertheless, 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, after reparametrization the marginal distributions only depend on θ_{21} or θ_{22}. For that reason, it is convenient to consider two groups of parameters, namely $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12}, \theta_{13})$ and $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})$. In addition, the positivity constraint depends mainly on $\boldsymbol{\theta}_2$ because $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$ are $O_p(n^{-\frac{1}{4}})$ under the null while $\hat{\theta}_{11}$, $\hat{\theta}_{12}$ and $\hat{\theta}_{13}$ are $O_p(n^{-\frac{1}{2}})$. Therefore, $\boldsymbol{\theta}_1$ 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 $x_2 = t x_1$, $\theta_{22} = k\theta_{21}$, $k \geq 0$ so that the polynomial that multiplies the Gaussian pdf simplifies to

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

where

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

It is easy to see that the minimum of $\tilde{P}(x, \phi, k, t, \theta_{21})$ 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 $\min_x \tilde{P}(x, \phi, k, t, \theta_{21})$ is greater than 0. Thus, we need to find a set $K(\phi)$ such that for all $\phi \neq 0$, for all $k \in K(\phi) \subseteq [0, +\infty)$ and for all $t \in \mathbb{R}$, we have either (1) $C_4(k, t, \phi) > 0$ or (2) $C_4(k, t, \phi) = 0$ and $C_2(k, t, \phi) \geq 0$. In other words, we need $C_4(k, t, \phi) = kt^4 - k\phi t^3 - \phi t + 1 \geq 0$ for all 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. Finally, if the discriminant is zero, then at least two roots must be equal. Therefore, we want the discriminant of $C_4(k, t, \phi)$ to be non-negative. Indeed, we can 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.

We can easily 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

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

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 contrast, 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 $C_4(k, t, \phi) > 0$. In summary, $k \in [lb(\phi), ub(\phi)]$ is sufficient for $C_4(k, t, \phi) \geq 0$ and the same is true for $\phi < 0$.

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.

In sum, we have shown that when $\boldsymbol{\theta}_1 = \mathbf{0}$, the asymptotes of the feasible set near 0 are $\theta_{22} = lb(\phi)\theta_{21}$ and $\theta_{22} = ub(\phi)\theta_{21}$.

Next, we know from Theorem 1 that

$$LR = ET(\boldsymbol{\theta}^{ET}) + O_p(n^{-\frac{1}{2r}}), \tag{C14}$$

where

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

$$\boldsymbol{\theta}^{ET} = \text{argmax}_{\boldsymbol{\theta}\in\Theta} ET_n(\boldsymbol{\theta}),$$

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

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

where

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

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

$$\omega_l = \arctan[lb(\tilde{\phi})], \quad \omega_u = \arctan[ub(\tilde{\phi})]. \tag{C15}$$

Let $\theta_{21} = \upsilon_1\eta$ and $\theta_{22} = \upsilon_2\eta$, then

$$ET_n(\boldsymbol{\theta}_1,\eta,\boldsymbol{\upsilon}) = 2\begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}\begin{pmatrix} S_{\boldsymbol{\theta}_1}(\tilde{\phi},\mathbf{0}) \\ \mathcal{S}_{\boldsymbol{\theta}_2}(\tilde{\phi},0,\boldsymbol{\upsilon}) \end{pmatrix} - n\begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}\begin{bmatrix} V_{11}(\tilde{\phi}) & \mathcal{V}_{12}(\tilde{\phi},\boldsymbol{\upsilon}) \\ \mathcal{V}_{21}(\tilde{\phi},\boldsymbol{\upsilon}) & \mathcal{V}_{22}(\tilde{\phi},\boldsymbol{\upsilon}) \end{bmatrix}\begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}, \tag{C16}$$

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, let $\tilde{\eta} = \max\{\eta^{ET}, n^{-k}\}$ with $\frac{1}{4} < k < \frac{1}{2}$. Then it is easy to see that

$$ET_n(\boldsymbol{\theta}_1^{ET},\tilde{\eta},\boldsymbol{\upsilon}^{ET}) = ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) + o_p(1). \tag{C17}$$

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

$$ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) \geq ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*) \geq ET_n(\boldsymbol{\theta}_1^{ET},\tilde{\eta},\boldsymbol{\upsilon}^{ET}) \tag{C18}$$

because $(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) = \text{argmax}_{pc} ET_n(\boldsymbol{\theta}_1,\eta,\boldsymbol{\upsilon})$ is chosen from a larger feasible set, and the event $(\boldsymbol{\theta}_1^{ET},\tilde{\eta},\boldsymbol{\upsilon}^{ET}) \in pc$ and $\{\tilde{\eta}\geq n^{-k}\}$ happens w.p.a. 1. Combining (C17) and (C18), we have

$$ET_n(\boldsymbol{\theta}_1^*,\eta^*,\boldsymbol{\upsilon}^*) = ET_n(\boldsymbol{\theta}_1^{ET},\eta^{ET},\boldsymbol{\upsilon}^{ET}) + o_p(1), \tag{C19}$$

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

In this context, note that there exists a $k' \in (k, \frac{1}{2})$ such that

$$\lim_n P(\|\boldsymbol{\theta}_1^*\| < n^{-k'} < n^{-k} \leq \eta^*) = 1. \tag{C20}$$

Therefore, this confirms that $\boldsymbol{\theta}_1^*$ is asymptotically irrelevant for the positivity constraints because it is effectively unrestricted. Consequently, (C20) implies that the only relevant restriction will affect the direction of $\boldsymbol{\theta}_2$.

In view of (C16), the first order condition for $\boldsymbol{\theta}_1^*$ for given η^* and $\boldsymbol{v}^*$ implies that

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

Hence, if we substitute $\boldsymbol{\theta}_1^*(\eta^*, \boldsymbol{v}^*)$ in the expression for $ET(\boldsymbol{\theta}_1, \eta, \boldsymbol{v})$, we end up with

$$\begin{aligned} ET_n(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{v}^*) = {} & \frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0}) \\ & - n^{\frac{1}{2}}\eta^{*2}[\mathcal{V}_{22}(\tilde{\phi}, \boldsymbol{v}^*) - \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{v}^*)V_{11}^{-1}(\tilde{\phi})\mathcal{V}_{12}(\tilde{\phi}, \boldsymbol{v}^*)]n^{\frac{1}{2}}\eta^{*2} \\ & + 2n^{\frac{1}{2}}\eta^{*2}[n^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{\theta}_2}(\tilde{\phi}, \mathbf{0}, \boldsymbol{v}^*) - \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{v}^*)V_{11}^{-1}(\tilde{\phi})n^{-\frac{1}{2}}S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0})]. \end{aligned} \tag{C21}$$

Given that (C21) is quadratic in η^{*2}, if take into account the restriction $\eta^* \geq n^{-k}$, we obtain

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

where $\mathcal{D}(\phi, \boldsymbol{v}) = \mathcal{S}_{\boldsymbol{\theta}_2}(\phi, \mathbf{0}, \boldsymbol{v}^*) - \mathcal{V}_{21}(\phi, \boldsymbol{v})V_{11}^{-1}(\phi)S_{\boldsymbol{\theta}_1}(\phi, \mathbf{0})$.

Thus, if we replace the previous expression for $\eta^*(\boldsymbol{v}^*)$ into (C21), we end up with

$$\begin{aligned} ET_n(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{v}^*) &= \frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0}) \\ & \quad + \underbrace{\frac{1}{n}\frac{\mathcal{D}^2(\tilde{\phi}, \boldsymbol{v}^*)\mathbf{1}[\mathcal{D}(\tilde{\phi}, \boldsymbol{v}^*) \geq 0]}{\mathcal{V}_{22}(\tilde{\phi}, \boldsymbol{v}^*) - \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{v}^*)V_{11}^{-1}(\tilde{\phi})\mathcal{V}_{12}(\tilde{\phi}, \boldsymbol{v}^*)}}_{\text{part 2}} + o_p(1). \end{aligned} \tag{C22}$$

But since part 2 in (C22) is a function of $\boldsymbol{v}^*$, which by definition is a maximizer of ET_n, we will finally end up with

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

which confirms that

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

in view of (C19).

C.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) for further discussion of this topic). 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^{o} 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 $i.i.d.$ bivariate normals with correlation φ. In practice, though, we fix the correlation coefficient to its estimated value in each sample in what is effectively a parametric bootstrap procedure (see Appendix D.1 in Amengual and Sentana (2015) for 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 same structure as in Table 1, the first three columns of Table 3 report rejection rates under the null at the 1%, 5% and 10% levels for $n = 400$ (top) and $n = 1,600$ (bottom). 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. In particular, when the sample size is large our proposal is the most powerful given that it is designed to direct power against alternatives in which the copula follows a Hermite expansion of the Gaussian one. In contrast, its non-parametric competitor has close to trivial power in samples of 400 observations, a situation that improves marginally when $n = 1,600$. 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 the alternatives we consider. Importantly, GET does a better job than the moment test based on the influence functions $\mathcal{S}_n$ 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

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: $n = 1,600$									
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. The correlation parameter φ is estimated under the null using the Gaussian rank correlation estimator described in Amengual, Sentana and Tian (2019). KS denotes the Kolmogorov–Smirnov test for copula models (see Rémillard (2017) for details) while KT–AS is the Kuhn-Tucker test based on the score of the symmetric Student t copula (see Amengual and Sentana (2020) for details). GMM refers to the J-test based on the influence functions underlying GET. 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 the alternative hypotheses, 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.

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.

D Example 4: Purely non-linear predictive regression

D.1 The model and its log-likelihood function

Consider the following extension of the nonlinear regression model in Bottai (2003), in which the data consist of n observations $\mathbf{y} = (y_1, y_2, y_3)$ drawn from a joint distribution characterized by

$$f(\mathbf{y};\boldsymbol{\theta}) = f(y_3|y_1, y_2; \boldsymbol{\theta}) f(y_1, y_2),$$

where $f(y_1, y_2)$ is fixed and known, while

$$f(y_3|y_1, y_2; \boldsymbol{\theta}) = \phi\left[y_3 - \exp(\theta_1 y_1 + \theta_2 y_2) + \theta_1 y_1 + \theta_2 y_2 + \frac{1}{2}\theta_2^2 y_2^2\right], \tag{D23}$$

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 (D23) 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).

D.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 log-likelihood function with respect to θ_1 and θ_2 evaluated at the null hypothesis are

$$\frac{\partial l}{\partial \theta_1} = 0, \quad \frac{\partial l}{\partial \theta_2} = 0,$$

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

and

$$\frac{\partial^3 l}{\partial \theta_2^3} = y_2^3(y_3 - 1).$$

Therefore, we have a situation in which the degree of underidentification is different for the two regression coefficients. But since Assumption 4 is satisfied with $C = \{(2,0), (1,1), (0,3)\}$, a straightforward application of Theorem 2 implies that

$$LR_n = \mathrm{GET}_n + O_p(n^{-\frac{1}{6}})$$
$$= \sup_{\theta_1,\theta_2} 2(\theta_1^2, \theta_1\theta_2, \theta_2^3)\begin{pmatrix} L_n^{[2,0]} \\ L_n^{[1,1]} \\ L_n^{[0,3]} \end{pmatrix} - n(\theta_1^2, \theta_1\theta_2, \theta_2^3)\begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}\begin{pmatrix} \theta_1^2 \\ \theta_1\theta_2 \\ \theta_2^3 \end{pmatrix} + O_p(n^{-\frac{1}{6}}), \quad \text{(D24)}$$

where

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

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 A. If p_1 is negligible, then (D24) is asymptotically equivalent to

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

If instead p_3 is negligible, then (D24) becomes asymptotically equivalent to

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

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

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

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

where

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

and

$$r_{13}=\frac{I_{13}-I_{12}I_{22}^{-1}I_{23}}{\sqrt{I_{11}-I_{12}I_{22}^{-1}I_{21}}\sqrt{I_{33}-I_{32}I_{22}^{-1}I_{23}}}.$$

Then, $supET_{1n}=Z_{2n}^2+Z_{3n}^2$ and $supET_{2n}=Z_{2n}^2+Z_{1n}^2\mathbf{1}\left[Z_{1n}\geq 0\right]$. As a consequence,

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

In other words, 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 Z_1, Z_2 and Z_3 will be three independent $N(0,1)$ random variables.

D.3 Simulation evidence

As alternative hypotheses, we consider $\theta_1=0.3$, $\theta_2=0$ (H_{a1}) and $\theta_1=0$, $\theta_2=0.5$ (H_{a2}) in specification (D23). And like in the normal versus SNP example, by maintaining 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 *i.i.d.* spherical normal vectors (y_1,y_2,y_3), which effectively imposes the null hypothesis.

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 $n=400$ (top) and $n=1,600$ (bottom) 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

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
LM_3	17.6	39.8	52.9	18.2	38.8	50.9
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: $n = 1,600$						
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
LM_3	57.7	79.1	87.4	53.1	75.3	84.2
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 and LR are defined in Supplemental Appendix D. GMM refers to the J-test based on the influence functions underlying GET. OLS_1 denotes a standard LM test 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 while OLS_2 is the LM test 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. Finite sample critical values are computed by simulation. DGPs: $(y_1y_2) \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}_2)$ under both alternative hypotheses. In turn, $y_3|y_2, y_1$ is $i.i.d.$ standard normal under the alternatives $\theta_1 = 0.25$ and $\theta_2 = 0.25$ (H_{a1}), and $\theta_1 = 0.3$ and $\theta_2 = 0.1$ (H_{a2}).

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 $\max\{supET_{1n}, supET_{2n}\}$ version of our test.

E 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 can consider the reparametrization $\boldsymbol{\theta}_r = \eta\boldsymbol{\upsilon}$, with $\boldsymbol{\upsilon} \in \mathbb{R}^{q_r}$, $\|\boldsymbol{\upsilon}\| = 1$ 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 could rewrite the null hypothesis as $H_0 : \boldsymbol{\theta}_1 = 0, \eta = 0$, 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. Next, we could 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 $H_0 : \boldsymbol{\theta}_1 = 0, \eta = 0$ will yield the usual LM test. In any event, in Theorem 2 we derive a generalized extremum test in a more general context without resorting to any such reparametrization.

References

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

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.

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

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