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>

March 2021

Abstract

We study score-type tests in likelihood contexts in which the nullity of the information matrix under the null is larger than one, thereby generalizing earlier results in the literature. Examples include multivariate skew normal distributions, Hermite expansions of Gaussian copulas, purely non-linear predictive regressions, multiplicative seasonal time series models and multivariate regression models with selectivity. Our proposal, which involves higher order derivatives, is asymptotically equivalent to the likelihood ratio but only requires estimation under the null. We conduct extensive Monte Carlo exercises that study the finite sample size and power properties of our proposal and compare it to alternative approaches.

Keywords: Generalized extremum tests, Higher-order identifiability, Likelihood ratio test, Non-Gaussian copulas, Predictive regressions.

JEL: C12, C46, C58, C22, C34

*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, CEMFI, Duke University, Harbin Institute of Technology, Universidad de la República, Université de Montréal, Universidad Torcuato Di Tella, the 2019 China Meeting of the Econometric Society (Guangzhou), the 2019 Galatina Summer Meetings and the VIII Encuentro Anual de la SEU (Montevideo). Of course, the usual caveat applies. Financial support from the Santander - CEMFI Research Chair is also gratefully acknowledged.

1 Introduction

Rao's (1948) score test and Silvey's (1959) numerically equivalent Lagrange multiplier (LM) version completed the classic triad of classical hypothesis tests (see Bera and Bilias (2001) for a survey). Under standard regularity conditions, Likelihood ratio (LR), Wald and LM tests are asymptotically equivalent under the null and sequences of local alternatives, and thus they share their optimality properties.

A standard regularity condition is a full rank information matrix of the unrestricted model parameters evaluated under the null. Nevertheless, there are 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 situations in which the expected Jacobian of the influence functions is singular but the expected Jacobian of their derivatives 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) provided several examples of this situation: i) univariate type II Tobit models with selectivity, ii) stochastic production frontier models, and iii) mixture models. In all their examples, in fact, the average score with respect to one of the parameters of the alternative evaluated under the null is identically 0 in finite samples.

Lee and Chesher (1986) proposed to replace the LM test by what they called an "extremum test". Their suggestion is 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. Lee and Chesher (1986) 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 intuitively follows from the fact that their extremum 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 the existing literature the nullity of the information matrix is assumed to be 1. When the information matrix is repeatedly singular, in the sense that its nullity 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. 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 generalization of the Lee and Chesher (1986) approach which leads to extremum-type tests asymptotically equivalent to the corresponding LR test.

To illustrate our proposal, consider the estimation of the parameter vector $\boldsymbol{\rho}$ characterizing the probability density function (pdf) of the *i.i.d.* random vector $\mathbf{y}$, $f(\mathbf{y};\boldsymbol{\rho})$.[1] To keep the notation to a minimum, we begin by considering the simplest possible case. Let us partition $\boldsymbol{\rho}$ into two blocks: 1) $\boldsymbol{\phi}$, which contains the $p\times 1$ vector of parameters estimated under H_0; 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 $H_0 : \boldsymbol{\theta} = \mathbf{0}$. In what follows,

$$s_{\rho_j i}(\boldsymbol{\rho}) = \frac{\partial l_i(\boldsymbol{\rho})}{\partial \rho_j} = \frac{\partial \log f(\mathbf{y}_i;\boldsymbol{\rho})}{\partial \rho_j}$$

denotes the contribution of observation i to the score with respect to ρ_j, $1 \leq j \leq p+q$. We maintain throughout the assumption that the first p scores, $\mathbf{s}_{\boldsymbol{\phi}i}(\boldsymbol{\phi},\mathbf{0})$, are linearly independent under the null. In contrast, we initially assume that the remaining ones are zero.

Assuming that the variance of $\{\mathbf{s}_{\boldsymbol{\phi}i}(\boldsymbol{\phi},\mathbf{0}), vech[\partial^2 l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']\}$ has full rank under the null, the number of different elements of $\partial^2 l_i/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ is $\binom{q+1}{2} = q(q+1)/2 > q$ for $q > 1$ even if the Clairaut-Schwartz-Young theorem holds.

Let $V_{\boldsymbol{\theta\theta}}$ denote the asymptotic residual variance of $vec(\partial^2 l_i/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')$ after orthogonalizing these influence functions with respect to $\mathbf{s}_{\boldsymbol{\phi}i}$. In this context, we can define the extremum statistic for a given value of $\boldsymbol{\theta}$ as

$$ET_n(\boldsymbol{\theta}) = \frac{1}{n}\frac{[\boldsymbol{\theta}'(\partial^2 L_n/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')\boldsymbol{\theta}]^2\mathbf{1}[\boldsymbol{\theta}'(\partial^2 L_n/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')\boldsymbol{\theta} > 0]}{(\boldsymbol{\theta}\otimes\boldsymbol{\theta})' V_{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}\otimes\boldsymbol{\theta})},$$

where n denotes the sample size, $L_n = \sum_{i=1}^n l_i$ and $\mathbf{1}[A]$ the usual indicator function that takes the value 1 if the event A happens, and 0 otherwise. Importantly, the expected value of $\boldsymbol{\theta}'(\partial^2 L_n/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')\boldsymbol{\theta}$, which is proportional to the second-order term in the expansion of the log-likelihood function, is zero under the null rather than negative, as it happens in the regular case.

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

By analogy to the LR test, our proposed test statistic is simply the supremum of $ET_n(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$. In fact, under suitable regularity conditions, we show in Theorem 1 below that

$$LR_n = 2\left[L_n(\hat{\boldsymbol{\rho}}) - L_n(\tilde{\boldsymbol{\rho}})\right] = \sup_{\|\boldsymbol{\theta}\|\neq 0} ET_n(\boldsymbol{\theta}) + O_p(n^{-\frac{1}{4}}),$$

where $\hat{\boldsymbol{\rho}}$ denotes the unrestricted ML estimator (UMLE) and $\tilde{\boldsymbol{\rho}}$ the restricted one (RMLE). In what follows, we shall refer to the sup statistic above as the generalized extremum test (GET).

In the case of a single parameter, Theorem 1 collapses to the results obtained by Lee and Chesher (1986) and Rotnitzky et al (2000). However, when the information matrix is repeatedly singular, our result provides an asymptotically equivalent but computationally convenient alternative to the LR test, which requires the estimation under the alternative of a model whose log-likelihood function is extremely flat under the null. In addition, the maximization of $ET_n(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$ takes place on a space of dimension $q-1$ because we can alter the norm of $\boldsymbol{\theta}$ without changing the value of this statistic, while the maximization of the unrestricted log-likelihood function of the sample $L_n(\boldsymbol{\rho})$ is over a space of dimension $p+q$, which is usually much larger.[2] Importantly, although the common asymptotic distribution of the GET and LR test is often non-standard, there are examples, such as the multiplicative seasonal Arma model in Supplemental Appendix D.1, in which it will be χ^2-like.[3]

The structure of the paper is as follows. In section 2 we obtain our theoretical results first in the baseline case in which all the underidentified parameters have the same degree of underidentification $r > 1$, and then when the degree of underidentification may be different for different parameters. Then, in section 3 we illustrate our testing procedure in detail through three examples of interest in econometrics: 1) a bivariate generalization of the Tobit II model with selectivity in Lee and Chesher (1986), 2)

, and 3) testing for predictability in a purely non-linear regression model. We assess the finite sample performance of our proposals in those examples through an extensive Monte Carlo analysis in section 4. Finally, we conclude in section 5, relegating proofs and additional results to the appendices.

2 Theoretical results

2.1 Notation and regularity conditions

Consider the estimation of the parameter vector $\boldsymbol{\rho}$ characterizing the distribution of an *i.i.d.* random vector $\mathbf{y}$, where $\boldsymbol{\rho} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$, where $p = \dim(\boldsymbol{\phi})$ and $q = \dim(\boldsymbol{\theta})$. This parameter

[2] Obviously, both procedures require the estimation of the model under the null, but the RMLE $\tilde{\boldsymbol{\rho}}$ is often available in closed form.

[3] A standard asymptotic distribution is usually associated to the existence of some regular reparametrization.

vector is such that $\boldsymbol{\phi}$ contains those parameters estimated under the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$, so that $\boldsymbol{\theta}$ only appears under the alternative. We assume $\boldsymbol{\phi}$ is always first-order identified. Further, we assume that some elements of $\boldsymbol{\theta}$ concentrate the singularity of the information matrix. More specifically, we denote the log-likelihood function contribution from observation i, $l_i(\boldsymbol{\rho}) = \log f(\mathbf{y}_i; \boldsymbol{\rho})$ –which sometimes we denote as $l_i(\boldsymbol{\phi}, \boldsymbol{\theta})$ or simply l_i– so that the log-likelihood function of a sample of size n is $L_n = \sum_{i=1}^{n} l_i$.

As we mentioned in the introduction, we denote by $\boldsymbol{\rho}^*$ the true value of the parameter vector, while $\hat{\boldsymbol{\rho}}$ denotes the unrestricted ML estimator (UMLE) and $\tilde{\boldsymbol{\rho}}$, sometimes $(\tilde{\boldsymbol{\phi}}', \mathbf{0}')'$, the restricted one (RMLE). As usual, let $|.|$ and $||.||$ denote absolute value and Euclidean norm, respectively. Finally, we denote by $e_{\min}(\boldsymbol{A})$ and $e_{\max}(\boldsymbol{A})$ the smallest and largest eigenvalues, respectively, of a square matrix $\boldsymbol{A}$.

In what follows, we assume:

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 ball $\mathcal{N}$ 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.

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 the UMLE of $\boldsymbol{\rho}_0$, $\hat{\boldsymbol{\rho}}$ (Newey and McFadden 1994, Theorem 2.5). The "open ball" part of Assumption 1.1 is just used to simplify the expressions and their derivation. Extensions to situations in which the parameters lie at the boundary of the parameter space are feasible, but at the expense of complicating the notation and blurring the message of the paper. Finally, Assumption 1.4 guarantees convergence of the RMLE at the usual $n^{-\frac{1}{2}}$ rate.

2.2 Repeated singularity of the same order

We first consider the case in which q_1 elements $\boldsymbol{\theta}$, $\boldsymbol{\theta}_1$ say, are first-order identified, while the remaining q_r elements $\boldsymbol{\theta}_r$ are r^{th}-order identified, a definition that will become precise after we introduce Assumption 3 below. Hence, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_r')'$, where $q_1 = \dim(\boldsymbol{\theta}_1)$ and $q_r = \dim(\boldsymbol{\theta}_r)$, so that $q = q_1 + q_r$, and the information matrix under H_0 is such that its top $(p + q_1) \times (p + q_1)$ block is regular and the rest contains zeros, so that its nullity is precisely q_r.[4]

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

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

[4] One often needs to reparametrize the model to make sure it satisfies these conditions, an issue mentioned in the introduction that we discuss in detail in Supplemental Appendix B.1

where $\boldsymbol{\iota}_m$ is a vector of m ones,

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

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* $\boldsymbol{\iota}'_{p+q}\mathbf{j} \leq 2r$ *and satisfy* $E[\sup_{\boldsymbol{\rho}\in\mathcal{N}} |l_i^{[\mathbf{j}]}(\boldsymbol{\rho})|] < \infty$*. Furthermore, with probability 1,* $f(\mathbf{y}_i, \boldsymbol{\rho}) > 0$ *for all* $\boldsymbol{\rho} \in \mathcal{N}$*.*
(2.2) For $r \leq \boldsymbol{\iota}'_{p+q}\mathbf{j} \leq 2r$, $E\{[l_i^{[\mathbf{j}]}(\boldsymbol{\rho}_0)]^2\} < \infty$*.*
(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,* $|L_n^{[\mathbf{j}]}(\boldsymbol{\rho}) - L_n^{[\mathbf{j}]}(\boldsymbol{\rho}')| \leq ||\boldsymbol{\rho} - \boldsymbol{\rho}'|| \sum_i g(\mathbf{y}_i)$ *for all* $\boldsymbol{\rho}$ *and* $\boldsymbol{\rho}'$ *in* $\mathcal{N}$*.*
(2.4) For all $\mathbf{j}_1, \mathbf{j}_2 \in \left\{(\mathbf{e}, \mathbf{0}_{q_r}), (\mathbf{0}_{p+q_1}, \mathbf{j}_{\boldsymbol{\theta}_r}) \,\middle|\, \boldsymbol{\iota}'_{p+q_1}\mathbf{e} = 1, \boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r} = r, \mathbf{e} \in \mathbb{R}^{p+q_1}, \mathbf{j}_{\boldsymbol{\theta}_r} \in \mathbb{R}^{q_r}\right\}$*, we have*

$$\sup_{(\boldsymbol{\phi},\mathbf{0})\in\mathcal{N}} E\left[\frac{\partial}{\partial\boldsymbol{\phi}}[l_i^{[\mathbf{j}_1]}(\boldsymbol{\phi},\mathbf{0})] \cdot l_i^{[\mathbf{j}_2]}(\boldsymbol{\phi},\mathbf{0})] \,\middle|\, \boldsymbol{\phi},\mathbf{0}\right] < \infty.$$

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 $l_i^{[\mathbf{j}]}(\boldsymbol{\rho}_0)$, while we use Assumption 2.4 to prove that the estimated covariance matrix of the influence functions under the null converges to the true value at the usual $n^{-\frac{1}{2}}$ rate. This last assumption is not in Rotnitzky et al (2000) because they were interested in estimation, not testing.

Let $\boldsymbol{\theta}_r^{\otimes k} = \underbrace{\boldsymbol{\theta}_r \otimes \boldsymbol{\theta}_r \otimes ... \otimes \boldsymbol{\theta}_r}_{k \text{ times}}$ 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 denote the asymptotic covariance matrix of the relevant influence functions, which may be understood as a generalized information matrix. Specifically,

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

so that

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

coincides with 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) With probability 1

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

for all $\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r} \leq r-1$ such that $\mathbf{j}_{\boldsymbol{\theta}_r} = (j_1, ..., j_{q_r})'$.
(3.2) For all $\boldsymbol{\theta}_r \in \mathbb{R}^{q_r} : \boldsymbol{\theta}_r \neq \mathbf{0}$, 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.

Intuitively, the rationale for looking at

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

is that it coincides with the r^{th}-order term in the expansion of the log-likelihood function. In that regard, note that although the higher order derivatives $\partial^r l_i / \partial \boldsymbol{\theta}_r^{\otimes r}$ will usually contain many repeated elements because of Clairaut'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

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

where

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

is the residual in the least squares projection of $\partial^r L_n(\boldsymbol{\phi}, \mathbf{0}) / \partial \boldsymbol{\theta}_r^{\otimes r}$ on $S_{\boldsymbol{\theta}_{1n}}(\boldsymbol{\phi}, \mathbf{0})$.[5]

Theorem 1 *If Assumptions 1, 2 and 3 hold, then:*

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

where

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

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

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.

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 the "nuisance" parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}_1$ that can be estimated at standard rates.

Since $\|\boldsymbol{\theta}_r\|$ is irrelevant, we can without loss of generality set $\boldsymbol{\theta}_r$ to lie on the unit circle. This allows us to intuitively link Theorem 1 to those earlier results when $q_r > 1$. Specifically, consider the reparametrization $\boldsymbol{\theta}_r = \eta\boldsymbol{\lambda}$, with $\boldsymbol{\lambda} \in \mathbb{R}^{q_r}$, $\|\boldsymbol{\lambda}\| = 1$ and $\eta \geq 0$, so that η and $\boldsymbol{\lambda}$ represent the magnitude and direction of the parameter vector $\boldsymbol{\theta}_r$, respectively. Given that

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

we could rewrite the null hypothesis as $H_0 : \boldsymbol{\theta}_1 = 0, \eta = 0$, where $\boldsymbol{\lambda}$ 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 $\boldsymbol{\lambda}$, 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 suprema 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 $\boldsymbol{\lambda}$ has some limitations. First, Lee and Chesher (1986) would yield a pointwise result for a given $\boldsymbol{\lambda}$, 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 regard, the underidentification of $\boldsymbol{\lambda}$ is an artifact of the $\boldsymbol{\theta}_r = \eta\boldsymbol{\lambda}$ reparametrization that would persist even if the information matrix had full rank, in which case the supremum over $\boldsymbol{\lambda}$ of the test of $H_0 : \boldsymbol{\theta}_1 = 0, \eta = 0$ will yield the usual LM test. In any event, in the next section we shall derive GET_n in a more general context without resorting to any such reparametrization.

2.3 Repeated singularity of different orders

Theorem 1 provides a substantive generalization over existing results. Specifically, it covers situations in which all the partial (cross) derivatives up to a given order are identically 0. It also says that tests will be one-sided for even ordered derivatives and two-sided for odd ordered ones. However, there are situations in which the degree of underidentification of the different elements of $\boldsymbol{\theta}$ is heterogeneous.

In this section we present the most general case, which is characterized in terms of less primitive conditions than the ones presented in Assumptions 2 and 3. To do so, we need to give a more general interpretation to some of the objects defined in section 2.2.

Specifically, letting $\varsigma_{\boldsymbol{\phi}i}(\boldsymbol{\phi})$ and $\varsigma_{\boldsymbol{\theta}i}(\boldsymbol{\phi})$ be $p \times 1$ and $m \times 1$ (with $m \geq q$)[6] sequences of measurable stochastic processes –not necessarily the score–, respectively, define

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

Also let

$$\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \begin{bmatrix} (\boldsymbol{\phi} - \boldsymbol{\phi}^*) + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \\ \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \end{bmatrix},$$

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, and

$$\mathcal{I}(\boldsymbol{\phi}) = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}) \\ \mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) \end{bmatrix}$$

denote a non-random symmetric $(p+m) \times (p+m)$ matrix (with $\mathcal{I}_{\boldsymbol{\phi\phi}}$ being a $p \times p$ matrix). Finally, in what follows, if we do not specify the argument of either $\mathcal{S}_n$, $\mathcal{S}$ and $\mathcal{I}$, the default argument is $\boldsymbol{\phi}^*$.

With these notational conventions, we state the following high-level assumption:

Assumption 4 *(Quadratic approximation) Let*

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

where $R_n(\boldsymbol{\phi}, \boldsymbol{\theta})$ is a remainder term. Then, assume:
(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) $n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*) \xrightarrow{d} \mathcal{S}(\boldsymbol{\phi}^*)$ for some zero-mean $\mathbb{R}^r$-valued Gaussian distribution that satisfying

$$E\left[\mathcal{S}(\boldsymbol{\phi}^*)\mathcal{S}'(\boldsymbol{\phi}^*)\right] = E\left[\varsigma_i(\boldsymbol{\phi}^*)\varsigma_i'(\boldsymbol{\phi}^*)\right] = \mathcal{I}(\boldsymbol{\phi}^*).$$

(4.3) $\mathcal{I}(\boldsymbol{\phi}^*)$ 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) If $n^{\frac{1}{2}}\lambda(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$, then $R_n(\boldsymbol{\phi}, \boldsymbol{\theta}) = O_p(n^{-a})$ for some a such that $\frac{1}{2} \geq a > 0$.

[6] We explicitly do not consider cases in which $m < q$ since they may cause underidentification, but not just of finite order.

Theorem 2 *Under the null, Assumptions 1 and 4.1–4, there exist compact sets* Φ *and* Θ *such that*

$$(\boldsymbol{\phi}^*, \mathbf{0}) \in int(\Phi \times \Theta) \subset \Phi \times \Theta \subset \mathcal{N}, \tag{2}$$

and then,

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

where

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

If, in addition, Assumption 4.5 holds, then

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

bla,bla

3 Examples

Given that LM tests 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, as reflected in the surveys by Breusch and Pagan (1980), Engle (1983), and Godfrey (1988). In addition to computational considerations, an important advantage of LM tests is that they are often easy to interpret as moment tests, so that rejections provide a clear indication of the specific directions along which modelling efforts should focus. As we mentioned in the introduction, though, standard LM tests cannot be computed when the information matrix is singular. In what follows, we discuss the application of our proposed tests as specification tests of two models of empirical interest, namely, a bivariate generalization of the Tobit II model with selectivity in Lee and Chesher (1986), and THE SNP. In addition, we consider a third example in which the objective is to detect non-linear predictability. Moreover, in Supplemental Appendix D we consider testing for multiplicative seasonal serial correlation in univariate time series as well as testing a multivariate normal copula against a Hermite expansion copula.

3.1 Testing for selectivity in a bivariate type II Tobit

Consider the following bivariate generalization of the type II Tobit model in Lee and Chesher (1986):

$$\begin{aligned} y_1 &= \mathbf{1}[y_1^* \geq 0] \\ y_2 &= y_2^*\mathbf{1}[y_1^* \geq 0] \\ y_3 &= y_3^*\mathbf{1}[y_1^* \geq 0] \end{aligned}$$

where

$$\begin{aligned} y_1^* &= \mathbf{x}_1'\boldsymbol{\varphi}_1 + u_1 \\ y_2^* &= \mathbf{x}_2'\boldsymbol{\varphi}_2 + u_2 \\ y_3^* &= \mathbf{x}_3'\boldsymbol{\varphi}_3 + u_3 \end{aligned} \quad \text{with} \quad \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \sim N\left[\mathbf{0}, \begin{pmatrix} 1 & \vartheta_1\sqrt{\varphi_4} & \vartheta_2\sqrt{\varphi_5} \\ \vartheta_1\sqrt{\varphi_4} & \varphi_4 & \varphi_6\sqrt{\varphi_4\varphi_5} \\ \vartheta_2\sqrt{\varphi_5} & \varphi_6\sqrt{\varphi_4\varphi_5} & \varphi_5 \end{pmatrix}\right]$$

(see Amemiya (1984) for a taxonomy of Tobit models). Under $H_0 : \vartheta_1 = \vartheta_2 = 0$, there is no selection bias, and one can jointly estimate $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$, φ_4, φ_5 and φ_6 by Seemingly Unrelated Regression Equations (SURE) using the non-zero observed values of y_2^* and y_3^*, while $\boldsymbol{\varphi}_1$ can be obtained from a univariate probit for y_1^*.

Observation i's log likelihood contribution is

$$(1-y_{1i})\log\Phi(-\mathbf{x}'_{1i}\boldsymbol{\varphi}_1)$$
$$+y_{1i}\left\{-\frac{1}{2}\log[(1-\varphi_6^2)\varphi_4\varphi_5]-\frac{1}{2}\mathbf{u}'_i(\boldsymbol{\varphi})\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\varphi})\mathbf{u}_i(\boldsymbol{\varphi})+\log\Phi\left[\frac{\mathbf{x}'_{1i}\boldsymbol{\varphi}_1+\boldsymbol{\upsilon}'(\boldsymbol{\varrho})\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\varphi})\mathbf{u}(\boldsymbol{\varphi})}{\sqrt{1-\boldsymbol{\upsilon}'(\boldsymbol{\varrho})\boldsymbol{\Upsilon}^{-1}(\boldsymbol{\varphi})\boldsymbol{\upsilon}(\boldsymbol{\varrho})}}\right]\right\},$$

where

$$\mathbf{u}_i(\boldsymbol{\varphi})=\begin{pmatrix} y_{2i}-\mathbf{x}'_{2i}\boldsymbol{\varphi}_2 \\ y_{3i}-\mathbf{x}'_{3i}\boldsymbol{\varphi}_3\end{pmatrix},\quad \boldsymbol{\upsilon}(\boldsymbol{\varrho})=\begin{pmatrix}\vartheta_1\sqrt{\varphi_4}\\ \vartheta_2\sqrt{\varphi_5}\end{pmatrix}\quad\text{and}\quad \boldsymbol{\Upsilon}(\boldsymbol{\varphi})=\begin{pmatrix}\varphi_4 & \varphi_6\sqrt{\varphi_4\varphi_5}\\ \varphi_6\sqrt{\varphi_4\varphi_5} & \varphi_5\end{pmatrix}.$$

Consider the case when $x_{1i}=1$ and both $\mathbf{x}_{2i}$ and $\mathbf{x}_{3i}$ contain a constant term. Straightforward algebra shows that if we evaluate all the scores at $\vartheta_1=\vartheta_2=0$, then

$$s_{\vartheta_1}-\sqrt{\varphi_4}M_1(\varphi_1)s_{\varphi_{21}}=0,$$
$$s_{\vartheta_2}-\sqrt{\varphi_5}M_1(\varphi_1)s_{\varphi_{31}}=0,$$

where φ_{21} and φ_{31} are the constants in the conditional means of y_{2i}^* and y_{3i}^*, respectively and $M_1(\varphi_1)=\Phi^{-1}(x_1\varphi_1)\phi(x_1\varphi_1)$. Such a singularity also arises when $\mathbf{x}_1$ is a set of dummy variables and $\mathbf{x}_2$ and $\mathbf{x}_3$ contain the same set of dummy variables. Intuitively, the problem occurs when Heckman's (1976) selectivity correction is perfectly collinear with the regressors that appear in the conditional means of y_{1i}^* and y_{2i}^*.

In this case, the three elements of the Hessian corresponding to ϑ_1 and ϑ_2 are all 0 too, so we need to do a second reparametrization to get the desired results. We can show that a suitable combined reparametrization would be

$$\begin{aligned}
\varphi_1&=\phi_1\\
\varphi_{21}&=\phi_{21}-\sqrt{\phi_4}M_1(\phi_1)\theta_{31}\\
\varphi_{22}&=\phi_{22}\\
\varphi_{31}&=\phi_{31}-\sqrt{\phi_5}M_1(\phi_1)\theta_{32}\\
\varphi_{32}&=\phi_{32}\\
\varphi_4&=\phi_4+\phi_4M_1(\phi_1)[M_1(\phi_1)+\phi_1]\theta_{31}^2\\
\varphi_5&=\phi_5+\phi_5M_1(\phi_1)[M_1(\phi_1)+\phi_1]\theta_{32}^2\\
\varphi_6&=\phi_6-.5[M_1(\phi_1)+\phi_1]M_1(\phi_1)\left(\phi_6\theta_{31}^2+\phi_6\theta_{32}^2-2\theta_{31}\theta_{32}\right)\\
\vartheta_1&=\theta_{31}\\
\vartheta_2&=\theta_{32}.
\end{aligned}$$

Then, we can show that

$$\left.\frac{\partial^{i+j}l}{\partial\theta_{31}^{i}\partial\theta_{32}^{j}}\right|_{\vartheta_1=\vartheta_2=0}=0,\ \ i=0,1,2,\ \ j=0,1,2,\ \ \text{and}\ \ 1\le i+j\le 2.$$

In addition, we can also show that the asymptotic variance of

$$\frac{\partial l}{\partial\phi_1},\frac{\partial l}{\partial\boldsymbol{\phi}_2},\frac{\partial l}{\partial\boldsymbol{\phi}_3},\frac{\partial l}{\partial\phi_4},\frac{\partial l}{\partial\phi_5},\frac{\partial l}{\partial\phi_6},\frac{\partial^3 l}{\partial\theta_{31}^3},\frac{\partial^3 l}{\partial\theta_{31}^2\partial\theta_{32}},\frac{\partial^3 l}{\partial\theta_{31}\partial\theta_{32}^2}\ \text{and}\ \frac{\partial^3 l}{\partial\theta_{32}^3}$$

has full rank.

3.2 Gallant and Nychka's semi-nonparametric MLE

In this note, I use HC for the high level conditions and A for primitive assumptions. The pdf is

$$f(y;\boldsymbol{\varrho})=f_N(y-\varphi_1,\varphi_2)\left(\varepsilon+\frac{(1-\varepsilon)\left\{P\left[\left(\frac{y-\varphi_1}{\sqrt{\varphi_2}}\right);\boldsymbol{\vartheta}\right]\right\}^2}{\int\{P[u;\boldsymbol{\vartheta}]\}^2\phi[u;0_2,\varphi_2]\,du}\right)$$

where

$$P(u;\boldsymbol{\vartheta})=1+\vartheta_1H_1(u)+\vartheta_2H_2(u).$$

and ε is a known small number to bound the density below from 0 (as a proof tool only, see Gallant and Nychka (1987) for details).[7] In what follows, to ease the presentation we ignore ε. Assuming that the paremeter space for $\boldsymbol{\varrho}$ is $[-\bar{\varphi}_1,\bar{\varphi}_1]\times[\underline{\varphi}_2,\bar{\varphi}_2]\times[-\bar{\vartheta},\bar{\vartheta}]^2$, after some reparameterization from $(\boldsymbol{\varphi},\boldsymbol{\vartheta})$ to $(\boldsymbol{\phi},\boldsymbol{\theta})$ as shown in Appendix B, we can show that at $(\boldsymbol{\phi},\boldsymbol{\theta})\in\Phi\times\mathbf{0}$, and letting $u=(y-\phi_1)/\sqrt{\phi_2}$, we will have that the score with respect to $\boldsymbol{\phi}$,

$$\frac{\partial l}{\partial\phi_1}=\frac{1}{\sqrt{\varphi_2}}H_1(u),\ \frac{\partial l}{\partial\phi_2}=\frac{1}{\sqrt{2}\varphi_2}H_2(u),$$

while for those in $\boldsymbol{\theta}$, the relevant quantities are

$$\frac{1}{2}\frac{\partial^2 l}{\partial\theta_2^2}=-\sqrt{6}H_4(u),\ \frac{\partial^2 l}{\partial\theta_1\partial\theta_2}=-2\sqrt{3}H_3(u)\ \ \text{and}\ \ \frac{1}{4!}\frac{\partial^4 l}{\partial\theta_1^4}=\frac{\sqrt{6}}{9}H_4(u).$$ [8]

Letting $\tilde{u}=(y-\tilde{\phi}_1)\sqrt{\tilde{\phi}_2}$ and $\mathrm{H}_j(u)=\sum_{i=1}^n H_j(u)$, we can easily show that the LR test is asymptotically equivalent to the usual Jarque-Bera test, i.e.

$$L_n(\hat{\boldsymbol{\rho}})-L_n(\tilde{\boldsymbol{\rho}})=\frac{1}{n}\frac{\mathrm{H}_3(\tilde{u})}{V_3}+\frac{1}{n}\frac{\mathrm{H}_4(\tilde{u})}{V_4}+o_p(1)$$

To prove this, we first verify the high level assumptions.

[7] In practice, we can ignore ε for the purpose of our testing procedure. To get the GET, we will ignore ε to simplify the notation and calculation, but the same method applies with $\varepsilon>0$.

[8] Since clearly

$$\frac{\partial l}{\partial\theta_1}=\frac{\partial l}{\partial\theta_2}=\frac{\partial^2 l}{\partial\theta_1^2}=\frac{\partial^3 l}{\partial\theta_1^3}=0$$

*HC 1: It's trivial.

*HC 2: Let $l(\boldsymbol{\phi}, \boldsymbol{\theta}) = E_{\boldsymbol{\rho}^*}\left[l(\boldsymbol{\phi}, \boldsymbol{\theta})\right]$. We first prove a uniform convergence result:

$$\sup_{\boldsymbol{\varrho}} \left| \frac{1}{n} L_n(\boldsymbol{\varrho}) - l_0(\boldsymbol{\varrho}) \right| \xrightarrow{p} 0 \tag{3}$$

Recall that the paremeter space for $\boldsymbol{\varrho}$ is $[-\bar{\varphi}_1, \bar{\varphi}_1] \times [\underline{\varphi}_2, \bar{\varphi}_2] \times [-\bar{\vartheta}, \bar{\vartheta}]^2$. Then, we have

$$\begin{aligned}
l(\boldsymbol{\phi}, \boldsymbol{\theta}) &= \log\left\{ f_N\left(y - \varphi_1, \varphi_2\right)\left(\varepsilon + \frac{(1-\varepsilon)\left\{P\left[\left(\frac{y-\varphi_1}{\sqrt{\varphi_2}}\right); \vartheta\right]\right\}^2}{\int \left\{P\left[u; \vartheta\right]\right\}^2 \phi\left[u; 0_2, \varphi_2\right] du}\right)\right\} \\
&\geq \log\left\{ f_N\left(y - \varphi_1, \varphi_2\right) \varepsilon \right\} \\
&= \log\left\{ \frac{1}{\sqrt{\varphi_2}} \exp\left[-\frac{(y-\varphi_1)^2}{2\varphi_2}\right]\right\} + C_1 \\
&\geq -\frac{(|y| + \bar{\varphi}_1)^2}{2\underline{\varphi}_2} + C_2
\end{aligned}$$

where C's are known constants that does not depend on $\boldsymbol{\varrho}$ and y, and where the first inequality follows from the monotonicity of the logarithm, while the last one from the parameter space.

Also,

$$\begin{aligned}
l(\boldsymbol{\phi}, \boldsymbol{\theta}) &= \log\left\{ f_N\left(y - \varphi_1, \varphi_2\right)\left(\varepsilon + \frac{(1-\varepsilon)\left\{P\left[\left(\frac{y-\varphi_1}{\sqrt{\varphi_2}}\right); \vartheta\right]\right\}^2}{\int \left\{P\left[u; \vartheta\right]\right\}^2 \phi\left[u; 0_2, \varphi_2\right] du}\right)\right\} \\
&\leq \log\left\{ f_N\left(y - \varphi_1, \varphi_2\right)\right\} + \varepsilon + \frac{(1-\varepsilon)\left\{1 + \vartheta_1 H_1(\frac{y-\varphi_1}{\sqrt{\varphi_2}}) + \vartheta_2 H_2(\frac{y-\varphi_1}{\sqrt{\varphi_2}})\right\}^2}{1 + \vartheta_1^2 + \vartheta_2^2} - 1 \\
&= -\frac{1}{2}\log(\varphi_2) - \frac{(y-\varphi_1)^2}{2\varphi_2} + \left\{1 + \vartheta_1 H_1(\frac{y-\varphi_1}{\sqrt{\varphi_2}}) + \vartheta_2 H_2(\frac{y-\varphi_1}{\sqrt{\varphi_2}})\right\}^2 + C \\
&\leq -\frac{1}{2}\log(\underline{\varphi}_2) + \left\{1 + \bar{\vartheta}\left|\frac{y + \bar{\varphi}_1}{\sqrt{\underline{\varphi}_2}}\right| + \frac{\bar{\vartheta}}{\sqrt{2}}\left(\frac{y + \bar{\varphi}_1}{\sqrt{\underline{\varphi}_2}}\right)^2\right\}^2 + C
\end{aligned}$$

where the first inequality follows from $\log(x) \leq x - 1$, and letting

$$d(x) = \frac{(|y| + \bar{\varphi}_1)^2}{2\underline{\varphi}_2} + \left\{1 + \bar{\vartheta}\left|\frac{y + \bar{\varphi}_1}{\sqrt{\underline{\varphi}_2}}\right| + \frac{\bar{\vartheta}}{\sqrt{2}}\left(\frac{y + \bar{\varphi}_1}{\sqrt{\underline{\varphi}_2}}\right)^2\right\}^2 + C,$$

where C is another constant, it is then straightforward to notice that $|l(\boldsymbol{\theta})| \leq d(x)$ and $E[|d(x)|] < \infty$. Thus, by Lemma 2.4 in Newey and McFadden (1994),

$$\sup_{\boldsymbol{\varrho}} \left| \frac{1}{n} L_n(\boldsymbol{\varrho}) - l_0(\boldsymbol{\varrho}) \right| \xrightarrow{p} 0 \tag{4}$$

as desired. Moreover, for all $\epsilon > 0$,

$$l_0\left(\boldsymbol{\varphi}^*, \mathbf{0}\right) > \sup_{\|\boldsymbol{\varrho} - \boldsymbol{\varrho}^*\| > \epsilon, \boldsymbol{\varrho} \in \mathrm{P}} l_0(\boldsymbol{\varrho}) \tag{5}$$

(i.e. well seperated maximum), which follows from the fact that $\boldsymbol{\varrho}^*$ is the unique maximizer, $l_0(\boldsymbol{\varrho})$ is continuous and the parameter space is compact. Since $\boldsymbol{\varrho}$ and $\boldsymbol{\rho}$ are one-to-one continuous mapping, we have $\hat{\boldsymbol{\rho}}_n \xrightarrow{p} \boldsymbol{\rho}^*$.

To verify HC 3, let $S_{\boldsymbol{\phi}} = \left(\frac{1}{\sqrt{\phi_2^*}}\mathrm{H}_1(u^*), \frac{1}{\sqrt{2}\phi_2^*}\mathrm{H}_2(u^*)\right)'$, $S_{\boldsymbol{\theta}} = (\mathrm{H}_3(u^*), \mathrm{H}_4(u^*))'$ and $\lambda_{\boldsymbol{\phi}}(\boldsymbol{\theta}) = \mathbf{0}$ and $\lambda_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(-2\sqrt{3}\theta_1\theta_2, (\frac{\sqrt{6}}{9}\theta_1^4 - \sqrt{6}\theta_2^2)\right)$.

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\phi\phi}} & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_{\boldsymbol{\theta\theta}} \end{bmatrix} = \mathrm{diag}\left(\frac{1}{\varphi_2}\mathcal{I}_1, \frac{1}{2\varphi_2^2}\mathcal{I}_2, \mathcal{I}_3, \mathcal{I}_4\right)$$

where $\mathcal{I}_k = Var\,(H_k)$. Note that $(\boldsymbol{\phi}^*, \mathbf{0})$ is the unique minimizer of $\lambda(\boldsymbol{\rho})$, thus HC3(a) holds.

*HC3(a): Since $\lambda\,(\boldsymbol{\phi}, \boldsymbol{\theta})$ is a continuous function and $\{\boldsymbol{\rho} \in \mathrm{P} : ||\,(\boldsymbol{\phi}, \boldsymbol{\theta}) - (\phi^*, \mathbf{0})\,|| \geq \epsilon\}$ is a compact set, it sufficies to check that $(\boldsymbol{\phi}^*, \mathbf{0})$ is the unique solution to $\|\lambda\,(\boldsymbol{\phi}, \boldsymbol{\theta})\| = 0$. First notice that $||\lambda\,(\boldsymbol{\phi}, \boldsymbol{\theta})\,|| = 0$ is equivalent to $\lambda\,(\boldsymbol{\phi}, \boldsymbol{\theta}) = \mathbf{0}$. Thus we need

$$\boldsymbol{\phi} - \boldsymbol{\phi}^* = \mathbf{0};\ -2\sqrt{3}\theta_1\theta_2 = 0;\ -\sqrt{6}\theta_2^2 + \frac{\sqrt{6}}{9}\theta_1^4 = 0.$$

By the last two equation we have $\theta_1 = \theta_2 = 0$ and then by the first equality $\boldsymbol{\phi} = \boldsymbol{\phi}^*$. Thus, we have the unique solution is $(\boldsymbol{\phi}^*, \mathbf{0})$.

HC (b) (c) hold trivially. To verify HC 3(d), first note that by a 8-th order Taylor expansion, we have

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n} \frac{|R_n\,(\boldsymbol{\phi}, \boldsymbol{\theta})|}{\max\{1, n||\boldsymbol{\phi} - \boldsymbol{\phi}^*||^2, n\theta_1^8, n\theta_1^2\theta_2^2, n\theta_2^4\}} = o_p(1)$$

we can also verify that

$$\max_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n} \frac{\max\{1, n||\boldsymbol{\phi} - \boldsymbol{\phi}^*||^2, n\theta_1^8, n\theta_1^2\theta_2^2, n\theta_2^4\}}{(1 + ||n^{\frac{1}{2}}\lambda\,(\boldsymbol{\phi}, \boldsymbol{\theta})\,||)^2} = O(1)$$

thus HC3(d) holds.

*HC 3(d): We want to show that

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

We first show that for all $\gamma_n \to 0$,

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n} \frac{|R_n\,(\boldsymbol{\phi}, \theta)|}{h(\boldsymbol{\phi}, \theta)} = o_p(1) \tag{6}$$

where $h(\boldsymbol{\phi}, \boldsymbol{\theta}) = \max\{1, n||\boldsymbol{\phi} - \boldsymbol{\phi}^*||^2, n\theta_1^8, n\theta_1^2\theta_2^2, n\theta_2^4\}$. Doing a eigth order Taylor expansion, we get

$$L_n(\boldsymbol{\rho}) - L_n(\boldsymbol{\phi}^*, \mathbf{0}) = \sum_{i=1}^{9} A_i + \sum_{i=1}^{15} B_i$$

where

$$A_1 = \frac{\partial L_n}{\partial \boldsymbol{\phi}'}(\boldsymbol{\phi} - \boldsymbol{\phi}^*),\ \ A_2 = \sum_{j_1+j_2=2} E\left[l^{[j_1,j_2,0]}\right] n(\phi_1 - \phi_1^*)^{j_1}(\phi_2 - \phi_2^*)^{j_2} = -\frac{1}{2}(\boldsymbol{\phi} - \boldsymbol{\phi}^*)'\mathcal{I}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi} - \boldsymbol{\phi}^*)$$

$$A_3 = L_n^{[0_2,4,0]}\theta_1^4,\ \ A_4 = E\left[l^{[0_2,8,0]}\right]n\theta_1^8 = -\frac{1}{2}E\left[\left(l^{[0_2,4,0]}\right)^2\right]n\theta_1^8;$$

$$A_5 = L_n^{[0_3,2]}\theta_2^2,\ \ A_6 = E\left[l^{[0_3,4]}\right]n\theta_2^4 = -\frac{1}{2}E\left[\left(l^{[0_3,2]}\right)^2\right]n\theta_2^4;$$

$$A_7 = L_n^{[0_2,1,1]}\theta_1\theta_2,\ \ A_8 = E\left[l^{[0_2,2,2]}\right]n\theta_1^2\theta_2^2 = -\frac{1}{2}E\left[\left(l^{[0_2,1,1]}\right)^2\right]n\theta_1^2\theta_2^2;$$

$$A_9 = E\left[l^{[0_2,4,2]}\right]n\theta_1^4\theta_2^2 = -E\left[l^{[0_2,4,0]}l^{[0_3,2]}\right]n\theta_1^4\theta_2^2$$

$$B_1 = \sum_{j_1+j_2=2}\left(\frac{1}{n}L_n^{[j_1,j_2,0]} - E\left[l^{[j_1,j_2,0]}\right]\right)n(\phi_1-\phi_1^*)^{j_1}(\phi_2-\phi_2^*)^{j_2}$$

$$B_2 = \sum_{j_1+j_2=3}^{8}\left\{\frac{1}{n}L_n^{[j_1,j_2,0]}\right\}n(\phi_1-\phi_1^*)^{j_1}(\phi_2-\phi_2^*)^{j_2}$$

$$B_3 = \left(\frac{1}{n}L_n^{[0_2,8,0]} - E\left[l^{[0_2,8,0]}\right]\right)n\theta_1^8,\ \ B_4 = \sum_{j=5}^{7}\left\{\frac{1}{\sqrt{n}}L_n^{[0_2,j,0]}\right\}\sqrt{n}\theta_1^j,\ \ B_5 = \left\{\frac{1}{\sqrt{n}}L_n^{[0_3,3]}\right\}\sqrt{n}\theta_2^3$$

$$B_6 = \left(\frac{1}{n}L_n^{[0_3,4]} - E\left[l^{[0_3,4]}\right]\right)n\theta_2^4,\ \ B_7 = \sum_{j=5}^{8}\left\{\frac{1}{n}L_n^{[0_3,j]}\right\}n\theta_2^j$$

$$B_8 = \left(\frac{1}{n}L_n^{[0_2,2,2]} - E\left[l^{[0_2,2,2]}\right]\right)n\theta_1^2\theta_2^2,\ \ B_9 = \sum_{j=2}^{3}\left\{\frac{1}{\sqrt{n}}L_n^{[0_2,1,j]}\right\}\sqrt{n}\theta_1\theta_2^j$$

$$B_{10} = \sum_{j=4}^{8}\left\{\frac{1}{n}L_n^{[0_2,1,j]}\right\}n\theta_1\theta_2^j,\ \ B_{11} = \sum_{j=2}^{5}\left\{\frac{1}{\sqrt{n}}L_n^{[0_2,j,1]}\right\}\sqrt{n}\theta_1^j\theta_2$$

$$B_{12} = \sum_{j=6}^{8}\left\{\frac{1}{n}L_n^{[0_2,1,j]}\right\}n\theta_1\theta_2^j; B_{13} = \sum_{\substack{j_1+j_2=5\\ j_1,j_2\geq 2}}^{8}\left\{\frac{1}{n}L_n^{[0_2,j_1,j_2]} - \mathbf{1}_{(4,2)}E\left[l^{[0_2,4,2]}\right]\right\}n\theta_1^{j_1}\theta_2^{j_2}$$

$$B_{14} = \sum_{\substack{|j|=2\\ \min(j_1+j_2,j_3+j_4)\geq 1}}^{8}\left(\frac{1}{n}L_n^{[j]}\right)n(\phi_1-\phi_1^*)^{j_1}(\phi_2-\phi_2^*)^{j_2}\theta_1^{j_3}\theta_2^{j_4}$$

$$B_{15} = \sum_{|j|=8}\left(\frac{1}{n}L_n^{[j_1,j_2,j_3,j_4]}(\bar{\boldsymbol{\rho}}) - \frac{1}{n}L_n^{[j_1,j_2,j_3,j_4]}\right)n\phi_1^{j_1}\phi_2^{j_2}\theta_1^{j_3}\theta_2^{j_4}$$

with $\bar{\boldsymbol{\rho}}$ between $\boldsymbol{\rho}$ and $(\boldsymbol{\phi}^*, \mathbf{0})$. Note that $\sum_{i=1}^{9} A_i = \frac{1}{2}LM_n(\boldsymbol{\theta})$ following from Faa di Bruno's formula, and thus by defnition, $R_n(\theta) = 2\sum_{i=1}^{15} B_n$. Note that in B_i terms, the caligraphic brackets implies $O_p(1)$ and round brackets implies $o_p(1)$, thus it's easy to see that (6) holds.

Then we verify that

$$\max_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n}\frac{h(\boldsymbol{\phi},\boldsymbol{\theta})}{(1+||n^{\frac{1}{2}}\lambda(\boldsymbol{\phi},\boldsymbol{\theta})||)^2} = \frac{h(\boldsymbol{\phi}_n,\boldsymbol{\theta}_n)}{(1+||n^{\frac{1}{2}}\lambda(\boldsymbol{\phi}_n,\boldsymbol{\theta}_n)||)^2} = O(1) \tag{7}$$

where $(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)$ is the maximizer. If $h(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$, (7) holds trivially. If $h(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$, we can find a subsequence $\{u_n\}$ of $\{n\}$ such that one of the follows five equations hold

$$h(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) = u_n\theta_{1,u_n}^2\theta_{2,u_n}^2 \to \infty \tag{8}$$

$$h(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n}) = u_n \left\|\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*_{u_n}\right\|^2 \to \infty \tag{9}$$

$$h(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n}) = u_n \theta^8_{1,u_n} \to \infty \tag{10}$$

$$h(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n}) = u_n \theta^4_{2,u_n} \to \infty \tag{11}$$

If (8) holds, then

$$\max_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\le\gamma_{u_n}} \frac{h(\boldsymbol{\phi},\boldsymbol{\theta})}{1+n||\lambda(\boldsymbol{\phi},\boldsymbol{\theta})||^2} = \frac{h(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})}{1+u_n||\lambda(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})||^2} \le \frac{u_n\theta^2_{1,u_n}\theta^2_{2,u_n}}{1+12n\theta_1^2\theta_2^2} = O(1)$$

Similarly, if (9) holds, then

$$\max_{(\phi,\boldsymbol{\theta})\in\mathbf{P}:\|(\phi,\boldsymbol{\theta})-(\phi^*,\mathbf{0})\|\le\gamma_{u_n}} \frac{h(\boldsymbol{\phi},\boldsymbol{\theta})}{1+n\left\|\lambda\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)\right\|^2} = \frac{h(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})}{1+u_n||\lambda(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})||^2} \le \frac{u_n\left\|\phi_{1,u_n}-\phi^*_{1,u_n}\right\|^2}{1+u_n\left\|\phi_{1,u_n}-\phi^*_{1,u_n}\right\|^2} = O(1)$$

If (10) holds, then

$$u_n\theta^2_{1,u_n}\theta^2_{2,u_n} = \frac{\sqrt{u_n\theta^8_{1,u_n}u_n\theta^4_{2,u_n}}}{\theta^2_{1,u_n}} \le u_n\theta^8_{1,u_n} \Rightarrow \theta^2_{2,u_n} \le \theta^2_{1,u_n}\theta^4_{1,u_n} \le \gamma_n^2\theta^4_{1,u_n}$$

Thus $\exists N$ s.t. $\forall n > N$, we have $\theta^2_{2,u_n} \le \frac{1}{18}\theta^4_{1,u_n}$ and thus $\left|-\sqrt{6}\theta_2^2 + \frac{\sqrt{6}}{9}\theta_1^4\right| \ge \frac{\sqrt{6}}{18}\theta_1^4$. Then

$$\frac{h(\phi_{u_n},\boldsymbol{\theta}_{u_n})}{1+u_n\left\|\lambda\left(\phi_{u_n},\boldsymbol{\theta}_{u_n}\right)\right\|^2} \le \frac{u_n\theta^8_{1,u_n}}{1+u_n\left(-\sqrt{6}\theta_2^2+\frac{\sqrt{6}}{9}\theta_1^4\right)^2} \le \frac{u_n\theta^8_{1,u_n}}{1+u_n\left(\frac{\sqrt{6}}{18}\theta_1^4\right)^2} = O(1).$$

If (11) holds, (if $\theta_{1,u_n} = 0$, then (12) holds trivially; if $\theta_{1,u_n} \neq 0$) then

$$u_n\theta^2_{1,u_n}\theta^2_{2,u_n} = \frac{\sqrt{u_n\theta^8_{1,u_n}u_n\theta^4_{2,u_n}}}{\theta^2_{1,u_n}} \le u_n\theta^4_{2,u_n} \Rightarrow \theta^4_{1,u_n} \le \theta^2_{1,u_n}\theta^2_{2,u_n} \le \gamma_n^2\theta^2_{2,u_n}$$

Thus $\exists N$ s.t. $\forall n > N$, we have $\theta^4_{1,u_n} \le \theta^2_{2,u_n}$ and thus $\left|\sqrt{6}\theta_2^2 - \frac{\sqrt{6}}{9}\theta_1^4\right| \ge \frac{8\sqrt{6}}{9}\theta^2_{2,u_n}$. Then

$$\frac{h(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})}{1+u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})||^2} \le \frac{u_n\theta^4_{2,u_n}}{1+u_n\left(-\sqrt{6}\theta_2^2+\frac{\sqrt{6}}{9}\theta_1^4\right)^2} \le \frac{u_n\theta^4_{2,u_n}}{1+u_n\left(\frac{8\sqrt{6}}{9}\theta^2_{2,u_n}\right)^2} = O(1). \tag{12}$$

Thus by Theorem 2 and $\mathcal{I}_{12} = \mathbf{0}$, we have

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

with $S_\theta = (\mathrm{H}_3(u^*), \mathrm{H}_4(u^*))'$ and $\lambda_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(-2\sqrt{3}\theta_1\theta_2, -\sqrt{6}\theta_2^2 + \frac{\sqrt{6}}{9}\theta_1^4\right)$. First, it's easy to see that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\{2S'_{\boldsymbol{\theta},n}\lambda_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right) - n\lambda_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)'\mathcal{I}_{\boldsymbol{\theta\theta}}\lambda_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)\right\} \le S'_{\boldsymbol{\theta},n}\mathcal{I}^{-1}_{\boldsymbol{\theta\theta}}S_{\boldsymbol{\theta},n}$$

Second, solving for

$$\begin{cases} -2\sqrt{3}\theta_1\theta_2 = n^{-1}\frac{\mathrm{H}_3(u^*)}{V_3} + o_p(1) \\ -\sqrt{6}\theta_2^2 + \frac{\sqrt{6}}{9}\theta_1^4 = n^{-1}\frac{\mathrm{H}_4(u^*)}{V_4} + o_p(1), \end{cases}$$

one feasible solution is

$$\theta_1^{\dagger} = \begin{cases} n^{-\frac{1}{8}}\left(\frac{9}{\sqrt{6}}n^{-\frac{1}{2}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{4}} = n^{-\frac{1}{4}}\left(\frac{9}{\sqrt{6}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{4}} & \text{if } \mathrm{H}_4(u^*) > 0 \\ n^{-\frac{1}{4}}\frac{-n^{-\frac{1}{2}}\frac{\mathrm{H}_3(u^*)}{V_3}}{2\sqrt{3}\left(\frac{-1}{\sqrt{6}}n^{-\frac{1}{2}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{2}}} = n^{-\frac{1}{2}}\frac{-\frac{\mathrm{H}_3(u^*)}{V_3}}{2\sqrt{3}\left(\frac{-1}{\sqrt{6}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{2}}} & \text{if } \mathrm{H}_4(u^*) < 0 \end{cases}$$

and

$$\theta_2^{\dagger} = \begin{cases} n^{-\frac{3}{8}}\frac{-n^{-\frac{1}{2}}\frac{\mathrm{H}_3(u^*)}{V_3}}{2\sqrt{3}\left(\frac{9}{\sqrt{6}}n^{-\frac{1}{2}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{4}}} = n^{-\frac{3}{4}}\frac{-\frac{\mathrm{H}_3(u^*)}{V_3}}{2\sqrt{3}\left(\frac{9}{\sqrt{6}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{4}}} & \text{if } \mathrm{H}_4(u^*) > 0 \\ n^{-\frac{1}{4}}\left(\frac{-1}{\sqrt{6}}n^{-\frac{1}{2}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{2}} = n^{-\frac{1}{2}}\left(\frac{-1}{\sqrt{6}}\frac{\mathrm{H}_4(u^*)}{V_4}\right)^{\frac{1}{2}} & \text{if } \mathrm{H}_4(u^*) < 0 \end{cases}$$

and $2S_{2,n}'\lambda_2\left(\boldsymbol{\theta}^{\dagger}\right) - n\lambda_2\left(\boldsymbol{\theta}^{\dagger}\right)'\mathcal{I}_{22}\lambda_2\left(\boldsymbol{\theta}^{\dagger}\right) = S_{2,n}'\mathcal{I}_{22}^{-1}S_{2,n} + o_p(1)$. Thus, we have

$$LR_n = S_{\boldsymbol{\theta},n}'\mathcal{I}_{\boldsymbol{\theta\theta}}^{-1}S_{\boldsymbol{\theta},n} + o_p(1) = n^{-1}\frac{\mathrm{H}_3(u^*)^2}{\mathcal{I}_3} + n^{-1}\frac{\mathrm{H}_4(u^*)^2}{\mathcal{I}_4} + o_p(1)$$

Finally, noticing that $n^{-1}\mathrm{H}_3(u^*)^2 = n^{-1}\mathrm{H}_3(\tilde{u})^2 + o_p(1)$ and $n^{-1}\mathrm{H}_4(u^*)^2 = n^{-1}\mathrm{H}_4(\tilde{u})^2 + o_p(1)$, we end up with

$$LR_n = n^{-1}\frac{\mathrm{H}_3(\tilde{u})^2}{\mathcal{I}_3} + n^{-1}\frac{\mathrm{H}_4(\tilde{u})^2}{\mathcal{I}_4} + o_p(1).$$

3.3 Purely non-linear predictive regression

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

$$f(y_3|y_1, y_2;\boldsymbol{\theta}) = \phi\left[y_3 - \exp\left(\theta_1 y_1 + \theta_2 y_2\right) + \theta_1 y_1 + \theta_2 y_2 + \frac{1}{2}\theta_2^2 y_2^2\right], \tag{13}$$

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 (viz 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 (13) 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).

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

$$LR_n = \text{GET}_n + O_p(n^{-\frac{1}{6}})$$

$$= \sup_{\theta_1,\theta_2} 2(\theta_1^2, \theta_1\theta_2, \theta_2^3) \begin{pmatrix} L_n^{[2,0]} \\ L_n^{[1,1]} \\ L_n^{[0,3]} \end{pmatrix} - n(\theta_1^2, \theta_1\theta_2, \theta_2^3) \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix} \begin{pmatrix} \theta_1^2 \\ \theta_1\theta_2 \\ \theta_2^3 \end{pmatrix} + O_p(n^{-\frac{1}{6}}), \quad (14)$$

where

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

Unlike in the two previous examples, in this case we would 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 a much simpler but asymptotically equivalent 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 **??** in Appendix A. If p_1 is negligible, then (14) 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 (14) 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 $\text{GET}_n = \max\{supET_{1n}, supET_{2n}\}$.

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

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

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

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

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 $Z_1 \geq 0$).[9]

4 Simulation evidence

In this section we study the finite sample size and power properties of the testing procedures we introduced in section 2 by means of several extensive Monte Carlo exercises. We do so in the context of the three different examples discussed in the previous section. For each distributional assumption, we generate 10,000 samples of size n and compute the parameter estimators and tests.[10] When no nuisance parameters are involved, we compute the exact finite sample distribution using 10,000 simulated samples. Otherwise, we employ a parametric bootstrap procedure based on the same number of simulated samples, so that we can automatically compute size-adjusted rejection rates, as forcefully argued by Horowitz and Savin (2000).

4.1 Bivariate type II Tobit

[9] If we further assume that the regressors y_1 anf 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.

[10] Given the number of Monte Carlo 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.

4.2 Gallant and Nychka's semi-nonparametric MLE

4.3 Non-linear predictive regression

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 (13). And like in the skew normal example, 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 ZZZ we compare the results of the two versions of our tests discussed in section 3.3 with the GMM test mentioned at the end of section 2.3 and two simple alternative procedures. First, a standard LM test based on pseudo-Gaussian ML that checks the joint significance of y_{1t}^2 and $y_{1t}y_{2t}$ in the OLS regression of y_{3t} 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_{1t}^3, $y_{1t}^2y_{2t}$, $y_{1t}y_{2t}^2$ and y_{2t}^3. We refer to these tests as OLS_1 and OLS_2, respectively.

As in previous examples, the first three columns of Table ZZZ 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 our simulated critical values are reliable for both sample sizes. In turn, the last six columns present the rejection rates at the 1%, 5% and 10% levels for the two previously mentioned alternatives. Once again, the behavior of the different test statistics is in accordance with expectations. In particular, our proposed statistics are the most powerful in both cases. Part of the reason has to do with the fact that the linear regressions only provide an approximation to the true non-linear conditional expectation. However, the fraction of the theoretical variance of y_{3t} explained by y_{1t}^2, $y_{1t}y_{2t}$, y_{1t}^3, $y_{1t}^2y_{2t}$, $y_{1t}y_{2t}^2$ and y_{2t}^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 $\chi_1^{2\prime}s$– 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. Figure 1, which displays those plots for the two sample sizes we consider, confirms 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.

5 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 those in Rotnitzky et al. (2000). However, when the information matrix is repeatedly singular, our procedures provide a computationally convenient alternative to the LR test. Our proposed test statistic is a sup type test over a space whose dimension is the nullity of the information matrix minus one when all parameters show the same degree of underidentification, and the nullity otherwise, 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 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. 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 $2r^{th}$-order expansion under the alternative.

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 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 valuable extension.

References

Adcock, C., Eling, M. and Loperfido, N. (2015): "Skewed distributions in finance and actuarial science: a review", *The European Journal of Finance* 21, 1253–1281.

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

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

Amengual, D. Sentana, E. and Tian, Z. (2019): "Gaussian rank correlation and regression", forthcoming in *Advances in Econometrics.*

Amsler, C., Prokhorov, A. and Schmidt, P. (2016): "Endogeneity in stochastic frontier models", *Journal of Econometrics* 190, 280-288.

Arellano-Valle, R.B. and Azzalini, A. (2008): "The centred parametrization for the multivariate skew-normal distribution", *Journal of Multivariate Analysis* 99, 1362–1382.

Azzalini, A. and Dalla Valle, A. (1996): "The multivariate skew-normal distribution", *Biometrika* 83, 715–726.

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.

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.

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): *The théorie générale de l'élimination.* De Leiber & Faraquet.

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

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.

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

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

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.

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

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.

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

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

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

Lemmata

Lemma 1 *If Assumptions 1 and 4.1–4 hold, then (i)* $\boldsymbol{\rho}_n^{LM} \xrightarrow{p} \mathbf{0}$ *and (ii)* $n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}_n^{LM}) = O_p(1)$.

Proof. Let us start by Lemma 1.(ii). By Assumption 4.2, we have that $n^{-\frac{1}{2}}\mathcal{S}_n = O_p(1)$; that is, there exists M_1 such that for all $n \geq N$,

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

Next, let $M = (2M_1+1)/e_{\min}(\mathcal{I})$, which is bounded because of Assumption 4.3. Then, noticing that if $||n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho})|| > M$ and $||n^{-\frac{1}{2}}\mathcal{S}_n|| \leq M_1$, we will have

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

which implies that

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

Then,

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

where from (A3) to (A4) we have used (A1) and (A2). Therefore, (A4) trivially implies that Lemma 1.(ii) holds.

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

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

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* $n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}_n) = O_p(1)$.

Proof. Assumptions 1 and 4.4 imply that

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

Thus, for given $\epsilon > 0$ there exists N such that for all $n > N$,

$$\Pr\left(A_n\right) \geq 1 - \frac{\epsilon}{4} \tag{A5}$$

with

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

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

$$\Pr(||n^{-\frac{1}{2}}\mathcal{S}_n|| \geq M_1) < \frac{\epsilon}{4}. \tag{A6}$$

Letting $M = (6M_1 + 3)/e_{\min}(\mathcal{I})$ and noticing that when

$$||N^{\frac{1}{2}}\boldsymbol{\lambda}\left(\boldsymbol{\rho}\right)|| \geq M, \;\; ||n^{-\frac{1}{2}}\mathbf{S}_n|| \leq M_1 \;\text{ and }\; \left|\frac{R_n(\boldsymbol{\rho})}{1 + n\left\|\lambda\left(\boldsymbol{\rho}\right)\right\|^2}\right| \leq \frac{1}{6}e_{\min}(\mathcal{I}),$$

we have that

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

As a consequence,

$$\Pr\left(\{||n^{\frac{1}{2}}\lambda\left(\hat{\boldsymbol{\rho}}_n\right)|| > M\} \cap \{||n^{-\frac{1}{2}}\mathcal{S}_n|| \leq M_1\} \cap A_n\right) = o(1) \tag{A7}$$

and, therefore,

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

where the inequalities follow from (A5), (A6) and (A7). □

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

Proof. We will show $\forall\epsilon_1 > 0, \forall\epsilon_2 > 0$, $\exists N$ s.t. $\forall n > N$

$$\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}_n) - LM_n(\boldsymbol{\rho}_n^{LM})\right| < \epsilon_1\right) > 1 - \epsilon_2$$

We know that $\max\{n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}_n), n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}_n^{LM})\} = O_p(1)$, so that for $\epsilon_2 > 0$, there exists M such that for all n,

$$\Pr(\max\{n^{\frac{1}{2}}\boldsymbol{\lambda}(\hat{\boldsymbol{\rho}}_n), n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\rho}_n^{LM})\} \le M) > 1 - \frac{\epsilon_2}{2}. \tag{A8}$$

Letting $A_n = \{\boldsymbol{\rho} \in \mathbf{P} : n^{\frac{1}{2}} \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \le M\}$, by Assumption 4.1 we can choose a seqence of $\gamma_n \to 0$ with $\delta_{\gamma_n} \ge 2M/\sqrt{n}$ for all but finite n satisfying

$$\inf_{\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\ge\gamma_n} \{\|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\| \ge \delta_{\gamma_n}\} > \frac{2M}{\sqrt{n}},$$

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

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

where the last equality follows from $\gamma_n \to 0$ and Assumption 4.4. Thus, there exists N such that for all $n > N$,

$$\Pr\left(\sup_{\boldsymbol{\rho}\in A_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon_1\right) > 1 - \frac{\epsilon_2}{2}. \tag{A9}$$

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

$$\begin{aligned}
&\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}_n) - LM_n(\boldsymbol{\rho}_n^{LM})\right| < \epsilon_1\right) \\
\ge& \Pr\left(\left\{\left|LR_n(\hat{\boldsymbol{\rho}}_n) - LM_n(\boldsymbol{\rho}_n^{LM})\right| < \epsilon_1\right\} \cap \{\hat{\boldsymbol{\rho}}_n \in A_n\} \cap \{\boldsymbol{\rho}_n^{LM} \in A_n\}\right) \\
\ge& \Pr\left(\left\{\sup_{\boldsymbol{\rho}\in A_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon_1\right\} \cap \{\hat{\boldsymbol{\rho}}_n \in A_n\} \cap \{\boldsymbol{\rho}_n^{LM} \in A_n\}\right) && \text{(A10)} \\
\ge& \Pr\left(\sup_{\boldsymbol{\rho}\in A_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < \epsilon_1\right) + P\left(\{\hat{\boldsymbol{\rho}}_n \in A_n\} \cap \{\boldsymbol{\rho}_n^{LM} \in A_n\}\right) - 1 && \text{(A11)} \\
\ge& 1 - \frac{\epsilon_2}{2} + 1 - \frac{\epsilon_2}{2} - 1 = 1 - \epsilon_2, && \text{(A12)}
\end{aligned}$$

where from (A10) to (A11), we have used $\Pr(E_1 \cap E_2) \ge \Pr(E_1) + \Pr(E_2) - 1$, while from (A11) to (A12), we used (A8) and (A9). □

Lemma 4 *If Assumptions 1 and 4.1–5 hold, then* $LR_n(\hat{\boldsymbol{\rho}}_n) = LM_n(\boldsymbol{\rho}_n^{LM}) + O_p(n^{-a})$.

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

$$\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}_n) - LM_n(\boldsymbol{\rho}_n^{LM})\right| \le K_\epsilon n^{-\frac{1}{2r}}\right) \ge 1 - \epsilon.$$

The proof is almost the same as the one of Lemma 3 (**XB: shall we just skip this part?**). Let A_n be as the one in Lemma 3, but with ϵ instead of ϵ_2. Then, by Assumption 4.5,

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

or, equivalently, there exists K_ϵ such that for all n,

$$\Pr\left(\sup_{\boldsymbol{\rho}\in A_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-a}\right) > 1 - \frac{\epsilon}{2} \tag{A13}$$

Thus for $n > N$,

$$\begin{aligned}
&\Pr\left(\left|LR_n(\hat{\boldsymbol{\rho}}_n) - LM_n(\boldsymbol{\rho}_n^{LM})\right| < K_\epsilon n^{-a}\right) \\
\geq &\Pr\left(\left\{\left|LR_n(\hat{\boldsymbol{\rho}}_n) - LM_n(\boldsymbol{\rho}_n^{LM})\right| < K_\epsilon n^{-a}\right\} \cap \{\hat{\boldsymbol{\rho}}_n \in A_n\} \cap \left\{\boldsymbol{\rho}_n^{LM} \in A_n\right\}\right) \\
\geq &\Pr\left(\left\{\sup_{\boldsymbol{\rho}\in A_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-a}\right\} \cap \{\hat{\boldsymbol{\rho}}_n \in A_n\} \cap \left\{\boldsymbol{\rho}_n^{LM} \in A_n\right\}\right) && \text{(A14)} \\
\geq &\Pr\left(\sup_{\boldsymbol{\rho}\in A_n} |LR_n(\boldsymbol{\rho}) - LM_n(\boldsymbol{\rho})| < K_\epsilon n^{-a}\right) + \Pr\left(\{\hat{\boldsymbol{\rho}}_n \in A_n\} \cap \left\{\boldsymbol{\rho}_n^{LM} \in A_n\right\}\right) - 1 && \text{(A15)} \\
\geq &1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 = 1 - \epsilon && \text{(A16)}
\end{aligned}$$

as desired. □

Lemma 5 *If Assumptions 1 and 4.1–4 hold, then* $LR_n(\tilde{\boldsymbol{\phi}}_n, \mathbf{0}) = \sup_{(\boldsymbol{\phi},\mathbf{0})\in\boldsymbol{P}} LM_n(\boldsymbol{\phi}, \mathbf{0}) + o_p(1)$. *Moreover, if in addition Assumption 4.5 holds, then* $LR_n(\tilde{\boldsymbol{\phi}}_n, \mathbf{0}) = \sup_{(\boldsymbol{\phi},\mathbf{0})\in\boldsymbol{P}} LM_n(\boldsymbol{\phi}, \mathbf{0}) + O_p(n^{-a})$.

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

Proof of Theorem 2

By Assumption 1, we have that $(\hat{\boldsymbol{\phi}}_n, \hat{\boldsymbol{\theta}}_n) \in \Phi \times \Theta$ and $\tilde{\boldsymbol{\phi}}_n \in \Phi$ with probability approaching 1 (wpa 1), with Θ and Φ defined in (2). It is then easy to verify that

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

where the first equality follows from $\tilde{\boldsymbol{\phi}}_n \in \Phi$ wpa 1, and the second one follows from $n^{-\frac{1}{2}}\mathcal{I}^{-1}_{\boldsymbol{\phi}\boldsymbol{\phi}}\mathcal{S}_{\boldsymbol{\phi},n} \in \{n^{\frac{1}{2}}(\boldsymbol{\phi} - \boldsymbol{\phi}^*) : \boldsymbol{\phi} \in \Phi\}$ wpa 1. Similarly, we have

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

Thus

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

And by the same argument, when Assumption 4.5 also holds, we will have

$$
LR = \sup_{\boldsymbol{\theta} \in \Theta} \left\{ 2(\mathcal{S}_{\boldsymbol{\theta}, n} - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}} \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1} \mathcal{S}_{\boldsymbol{\phi}, n})'_{\boldsymbol{\theta}} \boldsymbol{\lambda}(\boldsymbol{\theta}) - n \boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}) (\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}} \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1} \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}) \boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \right\} + O_p(n^{-a})
$$

as desired. □

Proof of Theorem 1

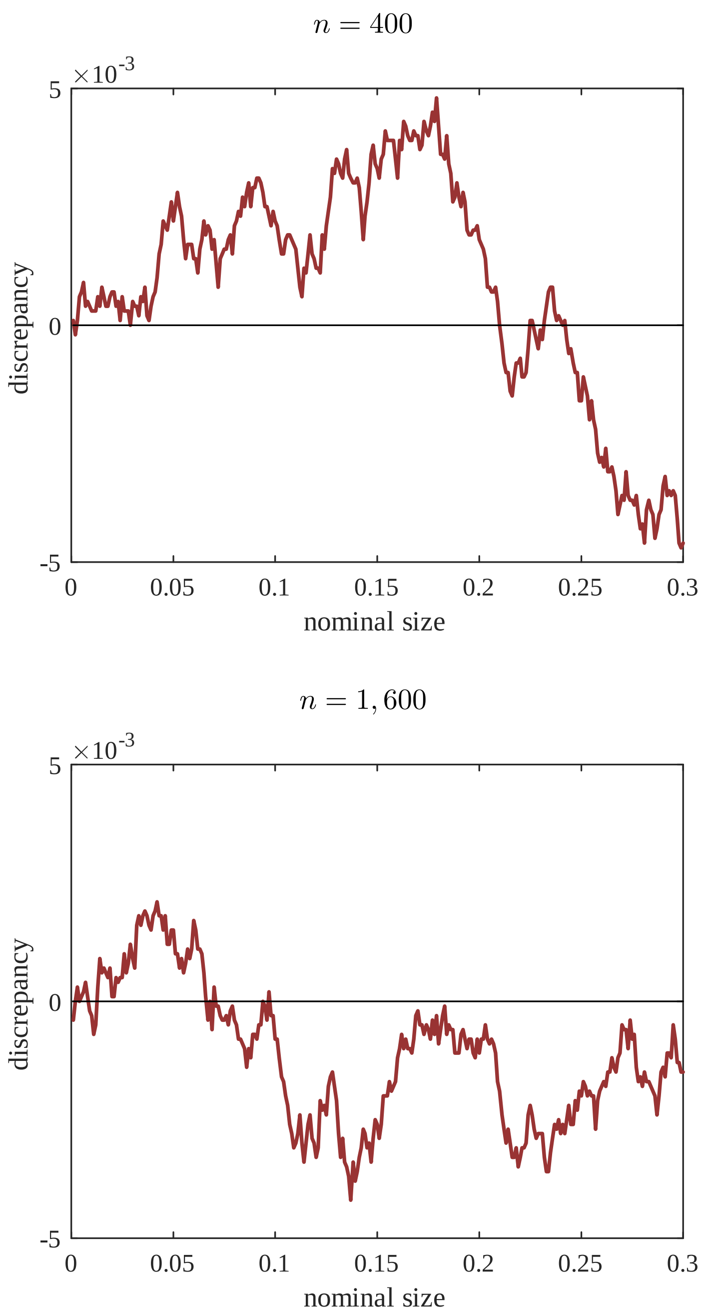
□

Table 1: Monte Carlo rejection rates (in %) under null and alternative hypotheses for white noise versus a purely nonlinear regression

	Null hypothesis			Alternative hypotheses H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
				$n = 400$					
GET	1.0	5.1	10.1	9.3	24.4	36.7	33.0	56.3	68.9
GET_2	1.1	5.3	10.1	8.0	21.8	32.4	31.5	54.7	66.2
OLS_1	1.0	4.9	10.0	5.3	16.7	25.6	1.7	7.8	14.0
OLS_2	1.1	5.1	10.0	3.6	12.1	19.8	20.1	40.4	53.1
GMM	1.0	5.2	10.4	6.9	19.8	30.1	27.9	51.6	64.0
				$n = 1,600$					
GET	1.1	5.3	9.5	70.3	88.0	92.8	82.7	93.9	96.6
GET_2	1.0	5.3	9.7	68.8	86.5	91.7	81.8	93.1	96.2
OLS_1	0.9	4.9	9.9	48.9	72.4	81.9	0.8	5.1	10.2
OLS_2	1.1	4.9	9.9	33.7	57.6	69.4	66.1	84.0	90.3
GMM	1.2	5.0	10.0	66.5	84.3	90.5	79.8	91.9	95.3

Notes: Results based on 10,000 samples. GET and GET_2 are defined in section 3.3. OLS_1 denotes a standard LM test that checks the joint significance of y_{1t}^2 and $y_{1t}y_{2t}$ in the OLS regression of y_{3t} 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_{1t}^3, $y_{1t}^2 y_{2t}$, $y_{1t}y_{2t}^2$ and y_{2t}^3. GMM refers to the J-test based on the influence functions underlying GET. Finite sample critical values are computed by simulation. DGPs: $(y_1 y_2) \sim i.i.d.\ N(\mathbf{0}, \mathbf{I}_2)$ under both the null and alternative hypotheses. In turn, $y_3|y_2, y_1$ is $i.i.d.$ standard normal under the null but under the alternative we consider $\theta_1 = 0.3, \theta_2 = 0$ (H_{a1}) and $\theta_1 = 0, \theta_2 = 0.5$ (H_{a2}).

Figure 1: p-value discrepancy plot for the white noise versus nonlinear predictability test



Notes: Results based on 10,000 simulated samples of size n of $(y_1, y_2, y_3) \sim i.i.d.\ N(0, I_3)$. GET is computed as defined in section 3.3. To compute the exact distribution for each sample size, we simulate $(Z_1, Z_2, Z_3) \sim N(0, I_3)$ 10^7 times, calculate $T = \max\{Z_1^2 \mathbf{1}\{Z_1 \geq 0\}, Z_3^2\} + Z_2^2$ each time, and obtain the α^{th} quantile of T, $Q_{T,\alpha}$.

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>

March 2021

B Reparametrization

B.1 Notation and a simple example

DA: This is taken from the introduction of Singular2001

Consider the estimation of the $p \times 1$ parameter vector $\boldsymbol{\varrho}$ characterizing the probability density function (pdf) of the original model characterized by the *i.i.d.* random vector $\mathbf{y}$, $f(\mathbf{y}; \boldsymbol{\varrho})$. In what follows,

$$s_{\varrho_j i}(\boldsymbol{\varrho}) = \frac{\partial \mathrm{l}_i(\boldsymbol{\varrho})}{\partial \varrho_j} = \frac{\partial \log f(\mathbf{y}_i; \boldsymbol{\varrho})}{\partial \varrho_j}$$

denotes the contribution of observation i to the score with respect to ϱ_j, $1 \leq j \leq$. To keep the notation to a minimum,by considering the simplest possible case, let us partition $\boldsymbol{\varrho}$ into two blocks: 1) $\boldsymbol{\varphi}$, which contains the $(p-q) \times 1$ vector of parameters estimated under H_0; and 2) $\boldsymbol{\vartheta}$, which is the $q \times 1$ vector of parameters such that the null hypothesis can be written in explicit form as $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$. We maintain throughout the assumption that the first $p-q$ scores, $\mathbf{s}_{\boldsymbol{\varphi}i}(\boldsymbol{\varphi}, \mathbf{0})$, are linearly independent under the null. In contrast, we initially assume that the remaining scores are a linear combination of those, so that

$$\mathbf{M}(\boldsymbol{\varphi})\mathbf{s}_{\boldsymbol{\varphi}i}(\boldsymbol{\varphi}, \mathbf{0}) + \mathbf{s}_{\boldsymbol{\vartheta}i}(\boldsymbol{\varphi}, \mathbf{0}) = \mathbf{0} \tag{B1}$$

for some $q \times (p-q)$ matrix $\mathbf{M}$, whose elements may be functions of $\boldsymbol{\varphi}$. In this context, the rank of the information matrix $E[\mathbf{s}_{\boldsymbol{\varrho}i}(\boldsymbol{\varphi}, \mathbf{0})\mathbf{s}'_{\boldsymbol{\varrho}i}(\boldsymbol{\varphi}, \mathbf{0})|(\boldsymbol{\varphi}, \mathbf{0})]$ is $p-q$ and its nullity q.

The first thing we do is to reparametrize the model so that the singularity is confined to the last elements of a new parameter vector. Specifically, we can reparametrize from $\boldsymbol{\varrho}$ to $\boldsymbol{\rho} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$ as

$$\boldsymbol{\varphi} = \boldsymbol{\phi} + \mathbf{M}'(\boldsymbol{\phi})\boldsymbol{\theta}, \quad \text{and} \quad \boldsymbol{\vartheta} = \boldsymbol{\theta}, \tag{B2}$$

so that $\boldsymbol{\varphi} = \boldsymbol{\phi}$ under the null. Defining $\mathrm{l}_i(\boldsymbol{\varrho}) = l_i(\boldsymbol{\rho})$ and assuming that this transformation is a continuous second-order diffeomorphism (which needs to be verified for each example), we can easily use the chain rule for first and second derivatives to show that evaluated under the null

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

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

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

B.2 Sequential reparametrization method

In this subsection, we show how to obtain the reparametrization described above but in more general context, using a sequential approach, provided:

Assumption 5 *1) The asymptotic variance of the sample averages of* $(\boldsymbol{s}_{\boldsymbol{\varphi}}, \boldsymbol{s}_{\boldsymbol{\vartheta}_1})$ *evaluated at* $(\boldsymbol{\varphi}, \mathbf{0})$ *scaled by* $\sqrt{n}$ *has full rank.*

2) $\left.\frac{\partial^{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r}} l}{\partial \boldsymbol{\vartheta}_r^{\mathbf{j}_{\boldsymbol{\theta}_r}}}\right|_{(\boldsymbol{\varphi},\mathbf{0})} = 0$*, for all index vectors such that* $\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r} < r-1$.

3) There exists a set of coefficients $\{m_k^{\mathbf{j}_{\boldsymbol{\theta}_r}}\}_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r}=r-1,k=1,\dots,p-q_r}$ *which may be functions of* $\boldsymbol{\varphi}$ *such that* $\forall \boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r} = r-1$

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

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

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

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

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

Then, if we use the chain rule we can show that

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

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

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

If not, we should check whether either:

1) there is a new set of coefficients $\{m_k^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}}\}_{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r}=r,k=1,\dots,p-q_r}$ which may be functions of $\boldsymbol{\phi}$ such that

$$m_1^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\phi_1} + \dots + m_{p-q}^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\phi_{p-q}} + m_{p-q+1}^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\theta_{11}} + \dots + m_{p-r}^{\dagger\mathbf{j}_{\boldsymbol{\theta}_r}} s_{\theta_{1q_1}} + \frac{\partial^{\boldsymbol{\iota}'_{q_r}\mathbf{j}_{\boldsymbol{\theta}_r}} l}{\partial \theta^{\mathbf{j}_{\boldsymbol{\theta}_r}}} = 0 \qquad \text{(B5)}$$

when evaluated under the null, in which case we can do further reparametrization from $(\boldsymbol{\phi}, \boldsymbol{\theta})$ to $(\boldsymbol{\phi}^\dagger, \boldsymbol{\theta}^\dagger)$ that sets 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.3 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 contribution of to the log-likelihood of observation i.

Define $\boldsymbol{\varrho} = (\boldsymbol{\varphi}, \boldsymbol{\vartheta})$ as the original parameter vector of dimension p, where $\boldsymbol{\varphi}$ is $(p-q) \times 1$ 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 $\boldsymbol{s}_{\boldsymbol{\varphi}}$ has full rank, 2) there is a $q \times (p-q)$ matrix $\mathbf{M}(\boldsymbol{\varphi})$ of possible functions of $\boldsymbol{\varphi}$ such that (B1) holds, and 3) the asymptotic variance of the sample average of

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

has full rank under the null for all $\boldsymbol{\lambda}$ such that $||\boldsymbol{\lambda}|| \neq 0$.

As usual, if we reparametrize from $\boldsymbol{\varrho}$ to $\boldsymbol{\rho}$ as in (B2), then, we can easily check that (B3) and (B4) hold when evaluated under the null, with

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

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 $\mathrm{GET}_n^{\boldsymbol{\rho}} = \sup_{||\boldsymbol{\lambda}|| \neq 0} ET_n^{\boldsymbol{\rho}}(\boldsymbol{\lambda})$, where

$$\begin{aligned} ET_n^{\boldsymbol{\rho}}(\boldsymbol{\lambda}) &= \frac{\left[\boldsymbol{\lambda}' \mathbb{H}(\tilde{\boldsymbol{\varphi}}) \boldsymbol{\lambda}\right]^2 \mathbf{1}\left[\boldsymbol{\lambda}' \mathbb{H}(\tilde{\boldsymbol{\varphi}}) \boldsymbol{\lambda} \geq \mathbf{0}\right]}{\mathcal{V}(\boldsymbol{\lambda}, \tilde{\boldsymbol{\varphi}})}, \\ \mathbb{H}(\boldsymbol{\varphi}) &= \begin{pmatrix} \mathbf{M}(\boldsymbol{\varphi})' \\ \mathbf{I}_q \end{pmatrix}' \left. \frac{\partial^2 \mathrm{l}(\boldsymbol{\varrho})}{\partial \boldsymbol{\varrho} \partial \boldsymbol{\varrho}'} \right|_{(\boldsymbol{\varphi}, \mathbf{0})} \begin{pmatrix} \mathbf{M}(\boldsymbol{\varphi})' \\ \mathbf{I}_q \end{pmatrix}, \end{aligned} \tag{B6}$$

and

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

is the adjusted variance of $\boldsymbol{\lambda}' \mathbb{H}(\boldsymbol{\varphi}) \boldsymbol{\lambda}$.

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 one-to-one mapping, 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) $\boldsymbol{\lambda}' \frac{\partial^2 l}{\partial \boldsymbol{\theta}^\dagger \partial \boldsymbol{\theta}^{\dagger\prime}} \boldsymbol{\lambda}$ is linearly independent of $\mathbf{s}_{\boldsymbol{\phi}^\dagger}$ $\forall ||\boldsymbol{\lambda}|| \neq 0$.

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

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

where we have used the chain rule in the first equality and (B1) in the second one, we need to assume that

$$\det \left(\frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\phi}^\dagger} - \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\phi}^\dagger} \mathbf{M} \right) \neq 0 \tag{B7}$$

for $\partial l/\partial \boldsymbol{\phi}^{\dagger}$ to have full rank. Similarly, given that (B1) 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

$$\frac{\partial \mathbf{g}^{\boldsymbol{\phi}\prime}}{\partial \boldsymbol{\theta}^{\dagger}} = \frac{\partial \mathbf{g}^{\boldsymbol{\theta}\prime}}{\partial \boldsymbol{\theta}^{\dagger}}\mathbf{M} \tag{B8}$$

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, (B8) and (B1) 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, (B6) implies that

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

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

$$a_i = \boldsymbol{\lambda}'\left(\frac{\partial^2 \mathbf{g}_i^{\boldsymbol{\phi}}}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}} - \mathbf{M}'\frac{\partial^2 \mathbf{g}_i^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\right)\boldsymbol{\lambda},$$

and

$$\boldsymbol{\lambda}^{\dagger} = \frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger\prime}}\boldsymbol{\lambda}.$$

In this context, if we further assume that

$$\det\left(\frac{\partial \mathbf{g}^{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}^{\dagger\prime}}\right) \neq 0, \tag{B9}$$

then it is easy to see that $\boldsymbol{\lambda}'\frac{\partial^2 l}{\partial \boldsymbol{\theta}^{\dagger}\partial \boldsymbol{\theta}^{\dagger}}\boldsymbol{\lambda}$ will be linearly dependent of $\mathbf{s}_{\boldsymbol{\phi}^{\dagger}}$ $\forall \left\|\boldsymbol{\lambda}^{\dagger}\right\| \neq 0$ because (a) $\boldsymbol{\lambda}^{\dagger\prime}\mathbb{H}\boldsymbol{\lambda}^{\dagger}$ 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 (B7), (B8) and (B9) 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 $\boldsymbol{\lambda}'\frac{\partial^2 l}{\partial\boldsymbol{\theta}^\dagger\partial\boldsymbol{\theta}^\dagger}\boldsymbol{\lambda}$ as

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

Then, we will have that

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

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

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

as desired.

B.4 Examples

B.4.1 Example 1: Testing for selectivity in a bivariate type II Tobit

B.4.2 Example 2: Gallant and Nychka's semi-nonparametric MLE

C Implementation details

C.1 Hermite expansion of the Gaussian copula

C.1.1 Influence functions

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 the bivariate 4^{th}-order Hermite polynomials $H_{31}(x_1, x_2; \phi)$, $H_{22}(x_1, x_2; \phi)$ and $H_{13}(x_1, x_2; \phi)$ are defined in (D21) and the $\mathbf{H}$'s in Supplemental Appendix C.1.

C.1.2 Positivity of the Hermite expansion of the Gaussian copula

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 described in section D.2, 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 $x_2 = tx_1$, $\theta_{22} = k\theta_{21}$, $k \geq 0$ so that the polynomial that multiplies the Gaussian pdf simplifies to

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

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 $C_4(k, t, \phi) > 0$. To summarize, $k \in [lb(\phi), ub(\phi)]$ 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

$$LR = ET(\boldsymbol{\theta}^{ET}) + O_p(n^{-\frac{1}{2r}}), \tag{C10}$$

where

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

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

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

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

where

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

with $\lambda_1 = \sin(\omega)$ and $\lambda_2 = \cos(\omega)$ so that $\|\boldsymbol{\lambda}\| = 1$, and

$$\omega_l = \arctan[lb(\tilde{\phi})], \quad \omega_u = \arctan[ub(\tilde{\phi})]. \tag{C11}$$

Let $\theta_{21} = \lambda_1 \eta$ and $\theta_{22} = \lambda_2 \eta$, then

$$\mathcal{ET}_n(\boldsymbol{\theta}_1, \eta, \boldsymbol{\lambda}) = 2 \begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix} \begin{pmatrix} S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0}) \\ \mathcal{S}_{\boldsymbol{\theta}_2}(\tilde{\phi}, 0, \boldsymbol{\lambda}) \end{pmatrix} - n \begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix} \begin{bmatrix} V_{11}(\tilde{\phi}) & \mathcal{V}_{12}(\tilde{\phi}, \boldsymbol{\lambda}) \\ \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{\lambda}) & \mathcal{V}_{22}(\tilde{\phi}, \boldsymbol{\lambda}) \end{bmatrix} \begin{pmatrix} \boldsymbol{\theta}_1 \\ \eta^2 \end{pmatrix}, \tag{C12}$$

with

$$\mathcal{S}_{\boldsymbol{\theta}_2}(\phi, 0, \boldsymbol{\lambda}) = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}' \begin{bmatrix} H_{\theta_{21}\theta_{21}}(\phi, \mathbf{0}) & H_{\theta_{21}\theta_{22}}(\phi, \mathbf{0}) \\ H_{\theta_{21}\theta_{22}}(\phi, \mathbf{0}) & H_{\theta_{22}\theta_{22}}(\phi, \mathbf{0}) \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

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

$$\mathcal{ET}_n(\boldsymbol{\theta}_1^{ET}, \tilde{\eta}, \boldsymbol{\lambda}^{ET}) = \mathcal{ET}_n(\boldsymbol{\theta}_1^{ET}, \eta^{ET}, \boldsymbol{\lambda}^{ET}) + o_p(1). \tag{C13}$$

Next, consider $(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{\lambda}^*) = \text{argmax}_{pc\wedge\{\eta\geq n^{-k}\}}\mathcal{ET}_n(\boldsymbol{\theta}_1, \eta, \boldsymbol{\lambda})$, where $pc = \{(\boldsymbol{\theta}_1, \eta\lambda_1, \eta\lambda_2) \in \Theta\}$. It is easy to see that with probability approaching 1,

$$\mathcal{ET}_n(\boldsymbol{\theta}_1^{ET}, \eta^{ET}, \boldsymbol{\lambda}^{ET}) \geq \mathcal{ET}_n(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{\lambda}^*) \geq \mathcal{ET}_n(\boldsymbol{\theta}_1^{ET}, \tilde{\eta}, \boldsymbol{\lambda}^{ET}) \tag{C14}$$

because $(\boldsymbol{\theta}_1^{ET}, \eta^{ET}, \boldsymbol{\lambda}^{ET}) = \text{argmax}_{pc}\mathcal{ET}_n(\boldsymbol{\theta}_1, \eta, \boldsymbol{\lambda})$ has a larger feasible set, and the event $(\boldsymbol{\theta}_1^{ET}, \tilde{\eta}, \boldsymbol{\lambda}^{ET}) \in pc$ and $\{\tilde{\eta} \geq n^{-k}\}$ happens with probability approaching 1. Combining (C13) and (C14), we have

$$\mathcal{ET}_n(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{\lambda}^*) = \mathcal{ET}_n(\boldsymbol{\theta}_1^{ET}, \eta^{ET}, \boldsymbol{\lambda}^{ET}) + o_p(1), \tag{C15}$$

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

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

$$\text{lim}_n P(\|\boldsymbol{\theta}_1^*\| < n^{-k'} < n^{-k} \leq \eta^*) = 1. \tag{C16}$$

Therefore, this confirms that $\boldsymbol{\theta}_1^*$ is asymptotically irrelevant for the positivity constraints because it is effectively unrestricted. Consequently, (C16) implies that the only relevant restriction will affect the direction of $\boldsymbol{\theta}_2$.

In view of (C12), the first order condition for $\boldsymbol{\theta}_1^*$ for given η^* and $\boldsymbol{\lambda}^*$ implies that

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

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

$$\begin{aligned}\mathcal{ET}_n(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{\lambda}^*) = &\frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0}) \\ &- n^{\frac{1}{2}}\eta^{*2}[\mathcal{V}_{22}(\tilde{\phi}, \boldsymbol{\lambda}^*) - \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{\lambda}^*)V_{11}^{-1}(\tilde{\phi})\mathcal{V}_{12}(\tilde{\phi}, \boldsymbol{\lambda}^*)]n^{\frac{1}{2}}\eta^{*2} \\ &+ 2n^{\frac{1}{2}}\eta^{*2}[n^{-\frac{1}{2}}\mathcal{S}_{\boldsymbol{\theta}_2}(\tilde{\phi}, \mathbf{0}, \boldsymbol{\lambda}^*) - \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{\lambda}^*)V_{11}^{-1}(\tilde{\phi})n^{-\frac{1}{2}}S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0})].\end{aligned} \tag{C17}$$

Given that (C17) is quadratic in η^{*2}, if take into account the restriction $\eta^* \geq n^{-k}$, we obtain

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

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

Thus, if we replace the previous expression for $\eta^*(\boldsymbol{\lambda}^*)$ into (C17), we end up with

$$\begin{aligned}\mathcal{ET}_n(\boldsymbol{\theta}_1^*, \eta^*, \boldsymbol{\lambda}^*) \quad = \quad &\frac{1}{n}S'_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0})V_{11}^{-1}(\tilde{\phi})S_{\boldsymbol{\theta}_1}(\tilde{\phi}, \mathbf{0}) \\ &+ \underbrace{\frac{1}{n}\frac{\mathcal{D}^2(\tilde{\phi}, \boldsymbol{\lambda}^*)\mathbf{1}[\mathcal{D}(\tilde{\phi}, \boldsymbol{\lambda}^*) \geq 0]}{\mathcal{V}_{22}(\tilde{\phi}, \boldsymbol{\lambda}^*) - \mathcal{V}_{21}(\tilde{\phi}, \boldsymbol{\lambda}^*)V_{11}^{-1}(\tilde{\phi})\mathcal{V}_{12}(\tilde{\phi}, \boldsymbol{\lambda}^*)}}_{\text{part 2}} + o_p(1).\end{aligned} \tag{C18}$$

But since part 2 in (C18) is a function of $\boldsymbol{\lambda}^*$, which by definition is a maximizer of $\mathcal{ET}$, we will finally end up with

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

which confirms that

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

in view of (C15).

D Additional examples

D.1 Testing white noise versus multiplicative seasonal AR

Box and Jenkins (1970) introduced the popular multiplicative seasonal ARIMA model to capture the autocorrelation of series with strong seasonal patterns, such as their famous airline passenger example. Suppose that after taking regular and seasonal differences of an observed time series, a researcher would like to formally assess the need for a more complicated dependence structure. Assuming the data is observed at the quarterly frequency, one of the alternatives that she might consider is the following AR(2)-SAR(2) process:

$$(1-\vartheta_1 L)(1-\vartheta_2 L)(1-\vartheta_3 L^4)(1-\vartheta_4 L^4)(y_t-\varphi_1)=\varepsilon_t, \tag{D19}$$

with $E(\varepsilon_t)=0$ and $V(\varepsilon_t)=\varphi_2$, where $y_t=\Delta\Delta_4 x_t$ and x_t is the original data. In this context, $H_0:\vartheta_1=\vartheta_2=\vartheta_3=\vartheta_4=0$.

As usual, non-linear least squares estimation coincides with Gaussian ML, so that the criterion function will be

$$-\frac{T}{2}\ln(2\pi)-\frac{T}{2}\ln\varphi_2-\sum_{t=1}^{T}\frac{[y_t-\mu_t(\varphi_1,\boldsymbol{\vartheta})]^2}{2\varphi_2},$$

where the conditional mean under the alternative is

$$\begin{aligned}\mu_t(\varphi_1,\vartheta) =&\varphi_1+(\vartheta_1+\vartheta_2)(y_{t-1}-\varphi_1)-\vartheta_1\vartheta_2(y_{t-2}-\varphi_1)+(\vartheta_3+\vartheta_4)(y_{t-4}-\varphi_1)\\ &-(\vartheta_1+\vartheta_2)(\vartheta_3+\vartheta_4)(y_{t-5}-\varphi_1)+\vartheta_1\vartheta_2(\vartheta_3+\vartheta_4)(y_{t-6}-\varphi_1)\\ &-\vartheta_3\vartheta_4(y_{t-8}-\varphi_1)+(\vartheta_1+\vartheta_2)\vartheta_3\vartheta_4(y_{t-9}-\varphi_1)-\vartheta_1\vartheta_2\vartheta_3\vartheta_4(y_{t-10}-\varphi_1).\end{aligned}$$

Hence, the scores evaluated under the null will be

$$s_{\varphi_1}(\boldsymbol{\varphi},\mathbf{0})=\frac{y_t-\varphi_1}{\varphi_2},\quad s_{\varphi_2}(\boldsymbol{\varphi},\mathbf{0})=\frac{(y_t-\varphi_1)^2-\varphi_2}{2\varphi_2^2},$$
$$s_{\vartheta_1}(\boldsymbol{\varphi},\mathbf{0})=s_{\vartheta_2}(\boldsymbol{\varphi},\mathbf{0})=\frac{(y_t-\varphi_1)(y_{t-1}-\varphi_1)}{\varphi_2},$$
$$s_{\vartheta_3}(\boldsymbol{\varphi},\mathbf{0})=s_{\vartheta_4}(\boldsymbol{\varphi},\mathbf{0})=\frac{(y_t-\varphi_1)(y_{t-4}-\varphi_1)}{\varphi_2}.$$

As a result:

$$s_{\vartheta_1}(\boldsymbol{\varphi},\mathbf{0})-s_{\vartheta_2}(\boldsymbol{\varphi},\mathbf{0})=0,\ s_{\vartheta_3}(\boldsymbol{\varphi},\mathbf{0})-s_{\vartheta_4}(\boldsymbol{\varphi},\mathbf{0})=0,$$

which shows that the nullity of the information matrix is 2.

Consider the reparametrization from $\boldsymbol{\varrho} = (\varphi_1, \varphi_2, \vartheta_1, ..., \vartheta_4)'$ to $\boldsymbol{\rho} = (\phi_1, \phi_2, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})'$ defined by

$$\varphi_1 = \phi_1, \quad \varphi_2 = \phi_2, \quad \vartheta_1 = \theta_{11} - \theta_{21}, \quad \vartheta_2 = \theta_{21}, \quad \vartheta_3 = \theta_{12} - \theta_{22} \quad \text{and} \quad \vartheta_4 = \theta_{22}.$$

The corresponding derivatives under the equivalent hypothesis $H_0 : \theta_{11} = \theta_{21} = \theta_{12} = \theta_{22} = 0$ are

$$\begin{aligned} \frac{\partial l_t}{\partial \theta_{11}} &= \frac{(y_t - \phi_1)(y_{t-1} - \phi_1)}{\phi_2}, \ \frac{\partial l_t}{\partial \theta_{21}} = 0, \\ \frac{\partial l_t}{\partial \theta_{12}} &= \frac{(y_t - \phi_1)(y_{t-4} - \phi_1)}{\phi_2}, \ \frac{\partial l_t}{\partial \theta_{22}} = 0, \end{aligned}$$

$$\frac{\partial^2 l_t}{\partial \theta_{21}^2} = \frac{2(y_t - \phi_1)(y_{t-2} - \phi_1)}{\phi_2}, \quad \frac{\partial^2 l_t}{\partial \theta_{21} \partial \theta_{22}} = 0, \quad \frac{\partial^2 l_t}{\partial \theta_{22}^2} = \frac{2(y_t - \phi_1)(y_{t-8} - \phi_1)}{\phi_2}.$$

Let $\theta_{21} = \lambda_1 \eta$ and $\theta_{22} = \lambda_2 \eta$ with $\lambda_1^2 + \lambda_2^2 = 1$ and consider the simplified null hypothesis $H_0 : \theta_{11} = \theta_{12} = 0$, $\eta = 0$. In this context, the only relevant quantity associated to η is

$$\frac{\partial^2 l_t}{\partial \eta^2} = 2\lambda_1^2 \frac{(y_t - \varphi_1)(y_{t-2} - \varphi_1)}{\sigma^2} + 2\lambda_2^2 \frac{(y_t - \varphi_1)(y_{t-8} - \varphi_1)}{\sigma^2}.$$

Moreover, given that under the null

$$E\left(\frac{\partial l_t}{\partial \boldsymbol{\phi}} \frac{\partial l_t}{\partial \boldsymbol{\theta}_1'}\right) = \mathbf{0} \quad \text{and} \quad E\left[\frac{\partial l_t}{\partial \boldsymbol{\phi}} vech'\left(\frac{\partial^2 l_t}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2'}\right)\right] = \mathbf{0},$$

we can ignore the parameter uncertainty in estimating ϕ_1 and ϕ_2, at least asymptotically.

In view of the discussion in section 2, the GET statistic will be given by

$$\text{GET}_T = \sup_{||\boldsymbol{\lambda}||=1} T^{-1}[S'_{\boldsymbol{\theta}_1}(\tilde{\boldsymbol{\phi}}, \mathbf{0}), \mathcal{H}_\eta(\tilde{\boldsymbol{\phi}}, 0, \boldsymbol{\lambda})]\mathcal{V}^{-1}(\tilde{\boldsymbol{\phi}}, \boldsymbol{\lambda})[S'_{\boldsymbol{\theta}_1}(\tilde{\boldsymbol{\phi}}, \mathbf{0}), \mathcal{H}_\eta(\tilde{\boldsymbol{\phi}}, 0, \boldsymbol{\lambda})]',$$

where

$$\begin{aligned} S_{\boldsymbol{\theta}_1}(\boldsymbol{\rho}) &= [S_{\theta_{11}}(\boldsymbol{\rho}), S_{\theta_{12}}(\boldsymbol{\rho})]', \\ \mathcal{H}_\eta(\boldsymbol{\phi}, \eta, \boldsymbol{\lambda}) &= \sum_{t=1}^{T} \frac{\partial^2 l_t(\boldsymbol{\rho})}{\partial \eta^2}, \\ \mathcal{V}(\boldsymbol{\phi}, \boldsymbol{\lambda}) &= Var\{T^{-1/2}[S'_{\boldsymbol{\theta}}(\boldsymbol{\phi}, \mathbf{0}), \mathcal{H}_\eta(\boldsymbol{\phi}, 0, \boldsymbol{\lambda})]' | \boldsymbol{\phi}, \mathbf{0}\}. \end{aligned}$$

Interestingly, in this example GET_T can be computed analytically. Specifically, straightforward algebra shows that

$$\text{GET}_T = T \sup_{||\boldsymbol{\lambda}|| \neq 0} \left\{ \tilde{r}_1^2 + \tilde{r}_4^2 + \frac{(\lambda_1^2 \tilde{r}_2 + \lambda_2^2 \tilde{r}_8)^2}{\lambda_1^4 + \lambda_2^4} 1[\lambda_1^2 \tilde{r}_2 + \lambda_2^2 \tilde{r}_8 \geq 0] \right\},$$

where

$$\tilde{r}_j = \frac{1}{T} \sum\nolimits_t \frac{(y_t - \tilde{\phi}_1)(y_{t-j} - \tilde{\phi}_1)}{\tilde{\phi}_2}$$

is the j^{th}-order sample autocorrelation of y_t. In addition, when $\tilde{r}_2 > 0$ or $\tilde{r}_8 > 0$, we can show that the value of $\boldsymbol{\lambda}$ that maximizes the above expression will be proportional to the vector

$$\begin{cases} (\sqrt{\tilde{r}_2 \mathbf{1}\left[\tilde{r}_2 \geq 0\right]}, \sqrt{\tilde{r}_8 \mathbf{1}\left[\tilde{r}_8 \geq 0\right]}), & \text{if } \tilde{r}_2 \geq 0 \text{ or } \tilde{r}_8 \geq 0 \\ (1,1), & \text{otherwise.} \end{cases}$$

As a result, GET_T will be

$$T(\tilde{r}_1^2 + \tilde{r}_4^2 + \tilde{r}_2^2 \mathbf{1}[\tilde{r}_2 \geq 0] + \tilde{r}_8^2 \mathbf{1}[\tilde{r}_8 \geq 0]), \tag{D20}$$

Therefore, the GET statistic is simply focusing on the first two regular sample autocorrelations and the first two seasonal ones, which is very intuitive in view of (D19). The partially one-sided nature of the test arises from the multiplicative nature of the alternative, which forces the roots to be always real. Additive alternatives, which allow for complex roots too, give rise to two-sided tests. Given that these estimated autocorrelations are asymptotically independent under the null, the asymptotic distribution of (D20) will be a mixture of χ_2^2, χ_3^2 and χ_4^2 with weights $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$, respectively. Not surprisingly, we would obtain exactly the same test statistic if we consider multiplicative MA alternatives instead.

Furthermore, we can show that a test of white noise against multiplicative AR(k)-SAR(k_s) for $k \geq 3$ or $k_s \geq 3$ will numerically coincide with the statistic in (D20). The intuition is as follows. We can show that when the null is true, the MLE of an additive AR(3) is such that all three roots of the lag polynomial are real with probability tending to 0, unless one of the roots is forced to be 0. Consequently, the LR for multiplicative AR(3) is asymptotically equivalent to the LR for AR(2), and the same applies to the corresponding GETs.

Finally, it is important to mention that our proposed test, which is based on sample autocorrelations, is numerically invariant to affine transformations of the observed series y_t. Effectively, this means that the finite sample distribution of our test is pivotal with respect to (ϕ_1, ϕ_2). Therefore, we can estimate the sample mean and variance of y_t, and apply our test directly to the standardized series as if they were the observed variables.

D.1.1 Monte Carlo simulations

Without loss of generality, we set the unconditional mean and variance of the innovations ε_t to 0 and 1, respectively, both under the null and alternative hypotheses. We also estimate the mean and variance parameters φ_1 and φ_2 with the sample mean and variance, respectively, which effectively impose the null. As alternative hypotheses we consider the covariance stationary models $(1 - .1L - .1L^2 - .1L^3 - .1L^4)y_t = \varepsilon_t$ (H_{a_1}) and $(1 - .4L)(1 + .4L)(1 - .4L^4)(1 + .4L^4)y_t = \varepsilon_t$ (H_{a_2}). Note that two of the roots of the first process are complex conjugates, so our tests is not ideally designed for it. We approximate the exact finite sample distribution using 10,000 simulated samples under the maintained hypothesis that the innovations are normal. Alternatively, one could consider a non-parametric bootstrap procedure that randomly draws with replacement from the observations, which would eliminate any time series dependence while allowing for any marginal distribution. As in section **??**, either way we do not need to take into account the sensitivity of the critical values to $\tilde{\boldsymbol{\varphi}}$ because the test statistics are numerically invariant to the values of this estimator.

In Table D.1 we compare the results of our tests with three alternative procedures: LM-AR(1) and LM-SAR(4), which denote standard LM tests based on the score of an AR(1) and a Wallis (1972)-style seasonal AR(4), respectively, and the GMM test described at the end of section 2.3.

Following the same structure by columns as in the previous tables, we report the results we have obtained for $n = 100$ (top) and $n = 400$ (bottom). The first three columns make clear that the our simulated finite sample distribution works remarkably well for both sample sizes. In turn, the last six columns present the rejection rates at the 1%, 5% and 10% levels for the two AR alternatives. Once again, the behavior of the different test statistics is in accordance with expectations. In particular, our proposal is the most powerful for H_{a_2}, which is not very surprising given that it is designed to direct power against such multiplicative alternatives with real roots. But it is also the top performer for H_{a_1} even though the process has two complex roots.

Given that in this case our test has a relatively standard asymptotic distribution, we can also compute p-value discrepancy plots to assess the finite sample reliability of this large sample approximation for every possible significance level. The results displayed in Figure D.1 confirm that the asymptotic distribution is also reliable in this context.

D.2 Testing Gaussian vs Hermite copulas

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 (2018) consider 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, which provide a rather flexible alternative.

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^{th}, which are defined as differentials of the multivariate normal density:

$$H_{\mathbf{v}}(\mathbf{x}, \boldsymbol{\varphi}) = f_{NK}(\mathbf{x}; \mathbf{R})^{-1} \left(\frac{-\partial}{\partial \mathbf{x}}\right)^{\mathbf{v}} f_{NK}(\mathbf{x}; \mathbf{R}), \boldsymbol{\iota}_K' \mathbf{v} = p \text{ with } \mathbf{v} \in \mathbb{N}^K, \tag{D21}$$

where $\boldsymbol{\varphi} = vecl(\mathbf{R})$ 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

$$f_H(x_1, x_2; \varphi, \boldsymbol{\vartheta}) = f_{N2}\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \begin{pmatrix} 1 & \varphi \\ \varphi & 1 \end{pmatrix}\right] P(x_1, x_2; \varphi, \boldsymbol{\vartheta}), \tag{D22}$$

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 (D22) 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

$$\left.\begin{array}{l} f_H(x_1; \vartheta_1) = \phi(x_1)[1 + \vartheta_1 H_4(x_1)] \\ f_H(x_2; \vartheta_5) = \phi(x_2)[1 + \vartheta_5 H_4(x_2)] \end{array}\right\}, \tag{D23}$$

where $H_4(x) = x^4 - 6x^2 + 3$ is the fourth-order univariate Hermite polynomial and $\phi(.)$ the standard normal pdf.

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 (D23). When the copula is Gaussian, x_i coincides with the Gaussian rank $\Phi^{-1}(u)$.

The pdf of the pure fourth-order Hermite expansion copula is

$$\frac{f_H(x_1, x_2; \boldsymbol{\varrho})}{f_H(x_1; \vartheta_1) f_H(x_2; \vartheta_5)} = \frac{\phi_2(x_1, x_2; \varphi)[1 + \sum_{j=0}^{4} \vartheta_{j+1} H_{4-j,j}(x_1, x_2; \varphi)]}{\phi_1(x_1)[1 + \vartheta_1 H_4(x_1)]\phi_1(x_2)[1 + \vartheta_5 H_4(x_2)]}.$$

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 those three second 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 (see Supplemental Appendix C.2.1 for details).

Let $\theta_{21} = \lambda_1\eta$ and $\theta_{22} = \lambda_2\eta$ with $\lambda_1^2 + \lambda_2^2 = 1$, 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

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

where

$$\begin{aligned}
\mathcal{D}_n(\boldsymbol{\phi}, \eta, \boldsymbol{\lambda}) &= \mathcal{H}_{\eta n}(\boldsymbol{\phi}, \eta, \boldsymbol{\lambda}) - \mathcal{V}_{\eta 1}(\boldsymbol{\phi}, \eta, \boldsymbol{\lambda})V_{11}^{-1}(\boldsymbol{\phi})S_{1n}(\boldsymbol{\phi}, \mathbf{0}),\\
\mathcal{H}_{\eta n}(\boldsymbol{\phi}, \eta, \boldsymbol{\lambda}) &= \sum_{i=1}^{n}(\lambda_1 \ \lambda_2)\begin{bmatrix} h_{\theta_{21}\theta_{21},i}(\boldsymbol{\rho}) & h_{\theta_{21}\theta_{22},i}(\boldsymbol{\rho}) \\ h_{\theta_{21}\theta_{22},i}(\boldsymbol{\rho}) & h_{\theta_{22}\theta_{22},i}(\boldsymbol{\rho}) \end{bmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix},\\
S_{1n}(\boldsymbol{\phi}, \mathbf{0}) &= [S_{\theta_{11}}(\boldsymbol{\phi}, \mathbf{0}), S_{\theta_{12}}(\boldsymbol{\phi}, \mathbf{0}), S_{\theta_{13}}(\boldsymbol{\phi}, \mathbf{0})]',
\end{aligned}$$

and the omitted arguments are $(\tilde{\boldsymbol{\phi}}, 0, \boldsymbol{\lambda})$ for $\mathcal{D}_n$, $(\tilde{\boldsymbol{\phi}}, \boldsymbol{\lambda})$ 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}.[11]

D.2.1 Positivity

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 $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22})'$.[12] Fortunately, 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 $\theta_{21} = \eta\lambda_1 = \eta\sin(\omega)$ and $\theta_{22} = \eta\lambda_2 = \eta\cos(\omega)$ with $\omega \in [0, 2\pi)$ to ensure a unit norm for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$. As we show in Supplemental Appendix C.1.2, these parameters lead to a positive density when η is small enough if and only if $\omega \in (\omega_l, \omega_u)$, with ω_l and ω_u defined in (C11).

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

$$\frac{1}{n}S'_{1n}V_{11}^{-1}S_{1n} + \frac{1}{n}\sup_{\omega\in(\omega_l,\omega_u)} \mathcal{D}'_n\left(\mathcal{V}_{\eta\eta} - \mathcal{V}_{\eta 1}\mathcal{V}_{11}^{-1}\mathcal{V}_{1\eta}\right)^{-1}\mathcal{D}_n\mathbf{1}\left[\mathcal{D}_n > 0\right]. \tag{D24}$$

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

[11] In view of equation (??), 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.

[12] This is an example in which Assumption 2.1 fails because $\boldsymbol{\rho}_0$ lies at the boundary of the admissible parameter space, and yet we can still derive a LR-equivalent test.

test is far more computationally convenient than the LR test because the positivity constraints effectively become linear under local alternatives.

D.2.2 Testing Gaussian vs Hermite copulas

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 (2018) 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 (2019), which effectively imposes the null. As alternative hypotheses, we consider two Hermite expansion copulas: one with $\boldsymbol{\vartheta}' = (0.04, 0, 0, 0, 0)$ (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 2 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 (2018)), 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 2 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 Hermite expansion copula 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 Hermite expansion copula alternatives. 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 (2018) 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 $\mathbf{L}_n$ implied by the higher-order expansion of the log-likelihood on which it is based, which is partly due to the fact that it takes into account the partially one-sided nature of the alternatives.

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

E Relationship to Dovonon and Renault (2013)

As we mentioned in the concluding section, the results of our paper can be extended to other extremum estimators, such as GMM. In that regard, the purpose of this appendix is to compare the results in Dovonon and Renault (2013) with the implications of our Theorem 1 for the particular case of $r = 2$. To simplify the notation, in what follows we will omit the nuisance parameters $\boldsymbol{\phi}$ from $\boldsymbol{\rho} = (\boldsymbol{\phi}', \boldsymbol{\theta}')'$.

Let Q be the normalized objective function of some extremum estimator $\hat{\boldsymbol{\theta}} \in \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q(\boldsymbol{\theta})$. Specifically, $Q^{GMM}(\boldsymbol{\theta}) = -n\bar{\boldsymbol{\psi}}'(\boldsymbol{\theta})\mathbf{W}_n\bar{\boldsymbol{\psi}}(\boldsymbol{\theta})$ in GMM, where $\boldsymbol{\psi}(\boldsymbol{\theta})$ denotes a vector of H influence functions and $\mathbf{W}_n \xrightarrow{p} \mathbf{W}$, while $Q^{ML}(\boldsymbol{\theta}) = 2L(\boldsymbol{\theta})$ in a likelihood context. For brevity of exposition, we assume that either our Assumptions 1 and 2 hold (likelihood), or Assumptions 1–5 in Dovonon and Renault (2013) hold (GMM).

Let us start by comparison of the rank deficiency conditions. Regarding first-order underidentification (Condition E1 henceforth), we have that $\frac{\partial}{\partial\boldsymbol{\theta}'}E[\boldsymbol{\psi}(\boldsymbol{\theta}_0)] = \mathbf{0}$ [see Proposition 2.1 in Dovonon and Renault (2013)]. In turn, our Assumption 2.1 implies that $\frac{\partial l}{\partial\boldsymbol{\theta}}\big|_{\boldsymbol{\theta}_0} = \mathbf{0}$. As for second-order identification (Condition E2 hereinafter), Lemma 2.3 in Dovonon and Renault (2013) implies that $\left(\boldsymbol{\lambda}' \frac{\partial^2\psi_h}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}_0} \boldsymbol{\lambda}\right)_{h=1,\ldots,H} \neq 0$ for all $||\boldsymbol{\lambda}|| \neq 0$. In the likelihood context instead, we have $\boldsymbol{\lambda}' \frac{\partial^2 l}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}_0} \boldsymbol{\lambda} \neq 0$ for all $||\boldsymbol{\lambda}|| \neq 0$ whenever Assumption 2.2 holds.

Using a fourth-order Taylor expansion of the normalized objective function Q around the true value of the parameter vector, we can show that

$$\begin{aligned} Q(\hat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0) =& \frac{\partial Q}{\partial\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\frac{\partial^2 Q}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &+ \frac{1}{3!}\sum_{\boldsymbol{\imath}_q'\boldsymbol{j}=3}\frac{\partial^3 Q}{\partial\boldsymbol{\theta}^{\boldsymbol{j}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\boldsymbol{j}} + \frac{1}{4!}\sum_{\boldsymbol{\imath}_q'\boldsymbol{j}=4}\frac{\partial^4 Q}{\partial\boldsymbol{\theta}^{\boldsymbol{j}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\boldsymbol{j}} + \delta_n, \end{aligned} \tag{E25}$$

where δ_n is a remainder term, which is zero in the Dovonon and Renault (2013) setup because $\boldsymbol{\psi}$ is a second order polynomial in $\boldsymbol{\theta}$, while we have shown it to be $o_p(1)$ in the likelihood context of our paper.

Next, we look a each of the other terms of the RHS of (E25) in detail.

Regarding the linear term in (E25), we have $\frac{\partial Q^{GMM}}{\partial \boldsymbol{\theta}'} = -2(\sqrt{n}\bar{\boldsymbol{\psi}}_n')\mathbf{W}_n\left(\sqrt{n}\frac{\partial \bar{\boldsymbol{\psi}}}{\partial \boldsymbol{\theta}'}\right)$ in the GMM context, which is $O_p(1)$ by virtue of Condition E1, while the analogous condition in the likelihood context implies that $\frac{\partial Q^{ML}}{\partial \boldsymbol{\theta}} = \mathbf{0}$. Moreover, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$ due to the usual regularity conditions, which implies that the first-order conditions are negligible in both cases.

As for the quadratic term in (E25), we can show that $\frac{1}{\sqrt{n}}\boldsymbol{\lambda}'\frac{\partial^2 Q}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\boldsymbol{\lambda}$ converges in distribution to a non-degenerate normal distribution with zero mean. In Dovonon and Renault (2013), specifically, this fact follows from the form of the GMM criterion function, which implies that

$$\frac{1}{\sqrt{n}}\boldsymbol{\lambda}'\frac{\partial^2 Q}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\boldsymbol{\lambda} = -2\boldsymbol{\lambda}'\frac{\partial \bar{\boldsymbol{\psi}}_n'}{\partial \boldsymbol{\theta}}\mathbf{W}_n\sqrt{n}\frac{\partial \bar{\boldsymbol{\psi}}_n}{\partial \boldsymbol{\theta}'}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}'\frac{\partial \text{vec}'\left(\partial \bar{\boldsymbol{\psi}}_n/\partial \boldsymbol{\theta}'\right)}{\partial \boldsymbol{\theta}}[\mathbf{I}_q \otimes (\sqrt{n}\mathbf{W}_n\bar{\boldsymbol{\psi}}_n)]\boldsymbol{\lambda},$$

while it is a consequence of the information matrix equality in our setup.

In turn, the third-order term in (E25) is dominated by the quadratic one in both cases. Specifically, $\frac{1}{\sqrt{n}}\frac{\partial^3 Q}{\partial \boldsymbol{\theta}^3} = O_p(1)$ holds in MLE by virtue of Lemma **??**, while it holds in GMM thanks to Condition E1. This, together with the fact that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$, allows us to neglect the third-order term.

Finally, regarding the fourth-order term of the expansion (E25), which is the one characterizing the asymptotic variance of the tests, we have that in the GMM context

$$\frac{1}{4!}\sum_{\boldsymbol{\imath}_q'\boldsymbol{j}=4}\frac{\partial^{\boldsymbol{4}} Q}{\partial \boldsymbol{\theta}^{\boldsymbol{j}}}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0)^{\boldsymbol{j}} = -\frac{1}{4}vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')'\left[\mathbf{G}'\mathbf{W}\mathbf{G} + o_p(1)\right]vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')$$

where $\hat{\mathbf{v}} = n^{\frac{1}{4}}(\hat{\boldsymbol{\theta}}_{GMM} - \boldsymbol{\theta}_0)$ and $\mathbf{G} = \left[vec\left(\frac{\partial^2 \psi_1}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right), vec\left(\frac{\partial^2 \psi_2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right), \ldots, vec\left(\frac{\partial^2 \psi_H}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right)\right]'$ (see Dovonon and Renault (2013, p. 2,576)).

Similarly, if we denote $(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)'\frac{\partial^2 Q}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0) = \mathbf{Z}vec[(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)']$, we will have that in the likelihood context

$$\frac{1}{4!}\sum_{\boldsymbol{\imath}_q'\boldsymbol{j}=4}\frac{\partial^{\boldsymbol{\imath}_q'\boldsymbol{j}} Q}{\partial \boldsymbol{\theta}^{\boldsymbol{j}}}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0)^{\boldsymbol{j}} \xrightarrow{p} -\frac{1}{4}vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')'Var(\mathbf{Z})\,vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')$$

by virtue of Lemma **??**.

As a consequence,

$$Q^{GMM}(\hat{\boldsymbol{\theta}}_{GMM}) - Q^{GMM}(\boldsymbol{\theta}_0) = vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')'\left[\underbrace{\mathbf{G}'\mathbf{W}\mathbf{X}}_{A_1} - \frac{1}{4}\underbrace{\mathbf{G}'\mathbf{W}\mathbf{G}}_{A_2}vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')\right] + o_p(1), \tag{E26}$$

where $\mathbf{X} \sim N[\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)]$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ is the asymptotic variance of $\sqrt{n}\bar{\boldsymbol{\psi}}_n(\boldsymbol{\theta}_0)$.

In turn,

$$Q^{ML}(\hat{\boldsymbol{\theta}}_{ML}) - Q^{ML}(\boldsymbol{\theta}_0) = vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')'\left[\underbrace{\mathbf{Z}}_{B_1} - \frac{1}{4}\underbrace{V(\mathbf{Z})}_{B_2}vec(\hat{\mathbf{v}}\hat{\mathbf{v}}')\right] + o_p(1) \tag{E27}$$

where $\mathbf{Z} \sim N[\mathbf{0}, V(\mathbf{Z})]$. Importantly, the term A_2 in (E26) is the variance of A_1 only if one chooses the optimal GMM weighting matrix $\mathbf{W} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$. In contrast, B_2 in (E27) is always the variance of B_1 because of Lemma **??**. Therefore, the asymptotic distribution of $Q^{GMM}(\hat{\boldsymbol{\theta}}_{GMM}) - Q^{GMM}(\boldsymbol{\theta}_0)$ and $Q^{ML}(\hat{\boldsymbol{\theta}}_{ML}) - Q^{ML}(\boldsymbol{\theta}_0)$ will be the same when $\mathbf{W} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$.

While the rank deficiency condition and the asymptotic distribution of $Q(\hat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}_0)$ look quite similar for a likelihood function and an optimal GMM criterion, there are some differences. First, the *expected* Jacobian is zero with rank deficiency q in GMM, while q linear combinations of the score vector are *numerically* zero in the likelihood context. An additional difference between GMM and MLE is that in the latter $\boldsymbol{\theta}$ is the parameter we want to test, while in the former the objective is to test some $H > q$ overidentified moment conditions, with $\boldsymbol{\theta}$ being the parameter vector estimated from those conditions.

Additional references

Amemiya, T. (1984): "Tobit models: a survey" *Journal of Econometrics* 24, 3-61.

Box, G. and G. Jenkins (1970): *Time series analysis: forecasting and control.* Holden-Day.

Heckman, J.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.

Wallis, K.F. (1972): "Testing for fourth order autocorrelation in quarterly regression equations", *Econometrica* 40, 617–636.

F Additional tables and figures

Table D.1: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the white noise versus multiplicative seasonal AR test.

	Null hypothesis			Alternative hypotheses					
				H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
				$n = 100$					
GET	1.0	4.7	9.4	26.7	43.7	54.1	24.6	47.2	60.4
LM-AR(1)	1.2	5.7	10.7	14.6	28.8	38.3	3.2	9.9	16.4
LM-SAR(4)	0.9	4.8	9.9	12.8	27.3	38.2	2.8	9.5	16.0
GMM	1.0	5.2	10.1	24.4	40.4	49.4	20.8	40.0	51.5
				$n = 400$					
GET	1.0	4.8	9.9	88.1	95.1	97.0	92.0	98.0	99.1
LM-AR(1)	1.2	4.4	9.7	60.2	76.4	84.1	3.3	9.9	16.8
LM-SAR(4)	1.1	5.4	9.8	59.2	78.6	86.4	5.6	15.0	22.6
GMM	0.9	5.0	9.9	86.1	93.7	96.1	89.0	96.5	98.5

Notes: Results based on 10,000 samples. The mean and variance parameters φ_1 and φ_2 are estimated under the null using the sample mean and sample variance. LM-AR(1) and LM-SAR(4) denote the Lagrange multiplier tests based on the score of an AR(1) and a seasonal AR(4), respectively. GMM refers to the J-test based on the influence functions underlying GET. Finite sample critical values are computed by simulation. DGPs: the true unconditional mean and the variance of the innovations are set to 0 and 1, respectively, under both the null and alternative hypotheses. As for the alternative hypotheses, $H_{a_1} : (1 - .1L - .1L^2 - .1L^3 - .1L^4)y_t = \varepsilon_t$ and $H_{a_2} : (1 - .4L)(1 + .4L)(1 - .4L^4)(1 + .4L^4)y_t = \varepsilon_t$.

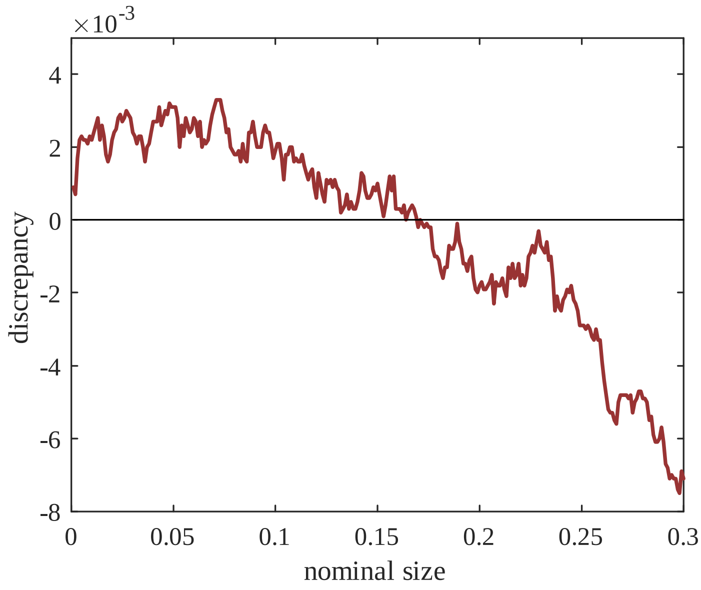
Table D.2: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the Gaussian versus Hermite expansion copula test.

	Null hypothesis			Alternative hypotheses					
				H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
				$n = 400$					
GET	1.1	5.0	10.2	22.6	55.8	69.7	23.9	55.8	69.8
KS	0.8	4.6	9.4	1.1	5.4	10.8	1.1	5.6	10.7
KT-AS	1.0	5.0	9.7	27.7	50.8	63.5	30.0	53.6	66.0
GMM	1.0	5.2	10.1	5.6	43.0	62.0	5.2	45.0	62.7
				$n = 1,600$					
GET	1.0	4.8	9.6	95.3	99.5	99.8	94.6	99.2	99.8
KS	1.1	5.1	10.4	2.0	7.7	14.5	2.4	9.4	17.0
KT-AS	1.1	4.9	10.0	79.8	93.4	96.5	83.8	95.1	97.6
GMM	1.1	5.0	9.8	55.9	97.9	99.6	57.1	97.8	99.3

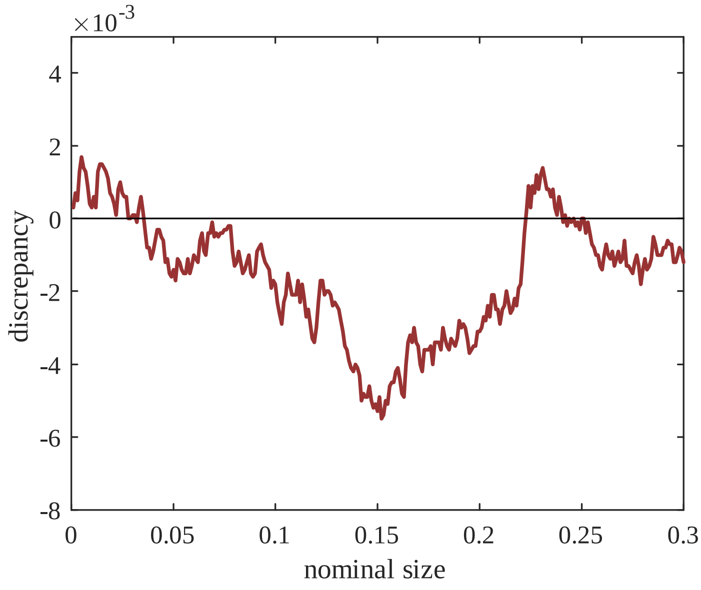
Notes: Results based on 10,000 samples. Margins are assumed to be known. The correlation parameter φ is estimated under the null using the Gaussian rank correlation estimator described in Amengual, Sentana and Tian (2019). KS denotes the Kolmogorov–Smirnov test for copula models (see Rémillard (2017) for details) while KT–AS is the Kuhn-Tucker test based on the score of the symmetric Student t copula (see Amengual and Sentana (2018) 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 Hermite expansion copulas with $\boldsymbol{\vartheta}' = (0.04, 0, 0, 0, 0)$ and $\boldsymbol{\vartheta}' = (0.02, 0, 0, 0, 0.02)$, respectively.

Figure D.1: p-value discrepancy plot for the white noise versus multiplicative seasonal AR test

$n = 100$



$n = 400$



Notes: Results based on 10,000 simulated samples of size n of $y \sim i.i.d.\ N(0,1)$. GET is computed as defined in section D.1. Given that the asymptotic distribution of the GET statistic is a mixture of χ^2_2, χ^2_3 and χ^2_4 with weights $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, we compute the p-values as a linear combination of the p-values of those three random variables with the same weights.