Hypothesis tests with a repeatedly singular information matrix*

Dante Amengual
CEMFI
<amengual@cemfi.es>

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

Enrique Sentana
CEMFI
<sentana@cemfi.es>

Revised: March 2025

Abstract

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

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

JEL: C12, C34, C46, C58.

*We would like to thank Tincho Almuzara, Manuel Arellano, Éric Renault and Andrea Rotnitzky for useful comments and suggestions, as well as audiences at Banco Central de Chile, Boston College, Boston University, CEMFI, Duke University, Harbin Institute of Technology, Harvard, NYU, Universidad de la República, Università di Firenze, Université de Montréal, Universidad Torcuato Di Tella, TAMU, the Econometric Society 2019 China Meeting (Guangzhou) and its 2020 World Congress, the 2019 Galatina Summer Meetings, the VIII Encuentro Anual de la SEU (Montevideo) and the 2023 TSE Financial Econometrics Conference. Of course, the usual caveat applies. The first and third authors gratefully acknowledge financial support from grant PID2021-128963NB-I00 funded by MICIU/AEI/1013039/501100011033 and ERDF/EU.

1 Introduction

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

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

Lee and Chesher (1986) proposed to replace the LM test by what they called an "extremum" test. Their suggestion was to study the restrictions that the null imposes on higher-order optimality conditions. Often, the second derivative will suffice ($r = 2$), but sometimes it might be necessary to study the third or even higher-order ones ($r \geq 3$). They proved the asymptotic equivalence between their extremum tests and the corresponding LR tests under the null and sequences of local alternatives in unrestricted contexts. Using earlier results by Cox and Hinkley (1974), this equivalence intuitively follows from the fact that their tests can often be re-interpreted as standard LM tests of a suitable transformation of the parameter whose first derivative is 0 on average such that the new score is no longer so. In contrast, Wald tests are extremely sensitive to reparametrization under these circumstances. Bera, Ra and Sarkar (1998)

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

provided some additional insights. In turn, Rotnitzky el al (2000) studied the asymptotic distribution of the maximum likelihood (ML) estimators in those contexts. Finally, Bottai (2003) looked at the validity of confidence intervals obtained by inverting the classical test statistics in this setup.

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

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

Repeatedly singular information matrices are not a mere theoretical curiosity. In fact, we illustrate our proposed procedure with several examples of interest that arise in economic and finance when testing: 1) exogenous sample selectivity in multivariate regressions; 2) normality against the flexible semi-nonparametric (SNP) family proposed by Gallant and Nychka (1987); 3) a Gaussian copula against another flexible Hermite expansion; and 4) unpredictability in a multiple regressor version of the purely non-linear model considered by Bottai (2003). Further, Amengual, Bei and Sentana (2022, 2024) discuss the application of the test proposed in this paper to two additional examples of substantial empirical interest: testing for multivariate normality against a skew normal distribution, and testing for neglected serial correlation in univariate time

[2] An exception is the multiplicative seasonal ARMA model considered in Amengual, Bei and Sentana (2024).

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

series models with and without unobserved components, respectively.

Our analysis is reminiscent of but different from the theoretical literature that studies GMM inference in situations in which the expected Jacobian of the influence functions is singular. There are in fact two strands in this literature. One suggests augmenting the influence functions with their first-order derivatives to restore standard asymptotics when the model parameters are first-order underidentified but second-order identified in the terminology of Sargan (1983) (see Lee and Liao (2018) and Sentana (2024) when the singularity direction is known or estimated, respectively). The other one studies the properties of the usual GMM estimators and overidentification restrictions tests in those circumstances (see Dovonon and Renault (2013, 2020), Dovonon and Hall (2018) and Han and McCloskey (2019) for asymptotic results, and Dovonon and Gonçalves (2017) for the correct implementation of the non-parametric bootstrap in this context).[4] Under second-order identification, the asymptotic distribution of our test statistic resembles that of the J test statistic in a GMM model.[5] However, the problems we analyze differ fundamentally from both these strands in two key aspects. First, in our context some of the (average) influence functions underlying the usual LM test not only have a singular Jacobian but are actually identically 0 when evaluated at the true parameter values under the null. Second, we allow the rank deficiency to be higher than one and the identification of the parameters to come from higher-order derivatives, not necessarily of the same order.

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

2 Theoretical results

Consider the estimation of the $d \times 1$ parameter vector $\boldsymbol{\rho}$ characterizing the distribution of an *iid* random vector $\mathbf{y}$. Let $l_i(\boldsymbol{\rho}) = \ln f(\mathbf{y}_i; \boldsymbol{\rho})$ denote the log-likelihood function contribution from observation i, so that the log-likelihood function of a sample of size n is $\mathcal{L}_n = \sum_{i=1}^{n} l_i(\boldsymbol{\rho})$.[6] In what follows,

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

[4] See also Dovonon, Hall and Kleibergen (2020) for a study of the local power properties of the alternative inference procedures proposed by Kleibergen (2005), which restore the χ^2 distribution by orthogonalizing the moment conditions with respect to the Jacobian.

[5] See Supplemental Appendix E of Amengual, Bei and Sentana (2020) for a formal link to the results in Dovonon and Renault (2013).

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

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

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

Using this notation, we henceforth assume:

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

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

2.1 Repeated singularity of the same order

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

Let $\mathbf{j}_d \in \mathbb{N}^d$ denote a vector of indices, $\mathbf{j}_d! = \prod_{i=1}^{d} j_i!$, $\boldsymbol{\iota}_d$ a vector of d ones,

$$l_i^{[\mathbf{j}_d]}(\boldsymbol{\rho}) = \frac{1}{\mathbf{j}_d!} \frac{\partial^{\boldsymbol{\iota}_d' \mathbf{j}_d} l_i(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}^{\mathbf{j}_d}}, \tag{1}$$

$\partial\boldsymbol{\rho}^{\mathbf{j}_d} = \partial\rho_1^{j_1}\ldots\partial\rho_d^{j_d}$ and $L_n^{[\mathbf{j}_d]}(\boldsymbol{\rho}) = \sum_{i=1}^n l_i^{[\mathbf{j}_d]}(\boldsymbol{\rho})$. 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 $l_i^{[\mathbf{j}_d]}(\boldsymbol{\rho})$ *exist for all* $\boldsymbol{\rho}$ *in* $\mathcal{N}$ *and* $\boldsymbol{\iota}_d'\mathbf{j}_d \leq 2r$*, and they satisfy* $E[\sup_{\boldsymbol{\rho}\in\mathcal{N}}|l_i^{[\mathbf{j}_d]}(\boldsymbol{\rho})|] < \infty$*.*
(2.2) For $r \leq \boldsymbol{\iota}_d'\mathbf{j}_d \leq 2r$*,* $E\{[l_i^{[\mathbf{j}_d]}(\boldsymbol{\rho})]^2\} < \infty$ *for all* $\boldsymbol{\rho}$ *in* $\mathcal{N}$*.*
(2.3) When $\boldsymbol{\iota}_d'\mathbf{j}_d = 2r$ *there is some function* $g(\mathbf{y}_i)$ *satisfying* $E[g^2(\mathbf{y}_i)] < \infty$ *such that with probability 1,* $|L_n^{[\mathbf{j}_d]}(\boldsymbol{\rho}) - L_n^{[\mathbf{j}_d]}(\boldsymbol{\rho}^\dagger)| \leq ||\boldsymbol{\rho}-\boldsymbol{\rho}^\dagger||\sum_i g(\mathbf{y}_i)$ *for all* $\boldsymbol{\rho}$ *and* $\boldsymbol{\rho}^\dagger$ *in* $\mathcal{N}$*.*

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

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

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

Moreover, let

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

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

$$V_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}) = \begin{bmatrix} V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \\ \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) \end{bmatrix} - \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) \\ \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) \end{bmatrix} \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}) [\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}_1}(\boldsymbol{\phi})\ \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}_r}(\boldsymbol{\phi})]$$

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

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

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

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

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

has full rank for all possible non-zero values of $\boldsymbol{\theta}_r \in \mathbb{R}^{q_r}$*.*

Intuitively, the rationale for looking at

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

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

Finally, let

$$Q_n(\boldsymbol{\theta}_r,\boldsymbol{\phi}) = \frac{\boldsymbol{\theta}_r^{\otimes r\prime}D_{rn}(\boldsymbol{\phi})D'_{rn}(\boldsymbol{\phi})\boldsymbol{\theta}_r^{\otimes r}}{\boldsymbol{\theta}_r^{\otimes r\prime}[V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_r}(\boldsymbol{\phi}) - V_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}^{-1}(\boldsymbol{\phi})V_{\boldsymbol{\theta}_1\boldsymbol{\theta}_r}(\boldsymbol{\phi})]\boldsymbol{\theta}_r^{\otimes r}}, \tag{2}$$

where

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

is the residual in the least squares projection of $\partial^r L_n(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}_r^{\otimes r}$ onto the linear span of $S_{\boldsymbol{\theta}_{1n}}(\boldsymbol{\phi},\mathbf{0})$.[7] In this context, we can proof the following result:

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

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

where

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

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

Importantly, expression (2), 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. Consequently, Theorem 1 generalizes the results in Lee and Chesher (1986) and Rotnitzky et al. (2000) by allowing for the presence of multiple singularities under the null (see Supplemental Appendix F for further comparisons to the existing literature).

Nevertheless, Theorem 1 does not cover situations in which the degree of underidentification of the different elements of $\boldsymbol{\theta}$ is heterogeneous, which we discuss next.

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

2.2 Repeated singularity of different orders

Let $C \subset \mathbb{N}^q$ denote a finite set of index pairs. In what follows, we use the vector inequality $\mathbf{j_q} < \mathbf{j_q^+}$ if and only if $j_k \leq j_k^+$, for $k = 1, ..., q$ and $\mathbf{j_q} \neq \mathbf{j_q^+}$. Combining these notational conventions with (1), we state the following assumption:

Assumption 4 *1)There exists a set* $C = \{\mathbf{j}_q^1, ..., \mathbf{j}_q^K\}$ *such that* $\forall k \leq K$ *(i)* $l_i^{[\mathbf{0}_{p-q}, \mathbf{j}_q^k]}(\boldsymbol{\phi}, \mathbf{0}) \neq 0$ *with positive probability but (ii)* $l_i^{[\mathbf{0}_{p-q}, \mathbf{j}_q]}(\boldsymbol{\phi}, \mathbf{0}) = 0$ *with probability 1 for all* $\mathbf{j}_q < \mathbf{j}_q^k$.
2) For all $i \leq q$*, there exists* $r_i \in \mathbb{N}$ *such that* $r_i \mathbf{e}_i \in C$*, where* $\mathbf{e}_i$ *is the* i^{th} *element of the canonical basis of order* q.
3) The asymptotic covariance matrix of the sample averages of $\mathbf{s}_{\boldsymbol{\phi}i}(\boldsymbol{\phi}, \mathbf{0})$, $l_i^{[\mathbf{0}_{p-q}, \mathbf{j}_q^1]}(\boldsymbol{\phi}, \mathbf{0})$*,... and* $l_i^{[\mathbf{0}_{p-q}, \mathbf{j}_q^K]}(\boldsymbol{\phi}, \mathbf{0})$ *scaled by* $\sqrt{n}$ *has full rank.*
4) For all $\mathbf{j}_q \in \mathbb{N}^q$*, one of the following holds: i)* $\mathbf{j}_q \in C$*; (ii) there exists* $\mathbf{j}_q' \in C$ *such that* $\mathbf{j}_q < \mathbf{j}_q'$*; (iii) there exists* $\mathbf{j}_q' \in C$ *such that* $\mathbf{j}_q > \mathbf{j}_q'$.

Let $\mathbf{L}_n = (L_n^{[\mathbf{0}p-q, \mathbf{j}_q^1]}, ..., L_n^{[\mathbf{0}_{p-q}, \mathbf{j}_q^K]})'$, $\boldsymbol{\theta}^{\mathbf{j}} = (\boldsymbol{\theta}^{\mathbf{j}_q^1}, ..., \boldsymbol{\theta}^{\mathbf{j}_q^K})$ with $\boldsymbol{\theta}^{\mathbf{j}_q^k} = \prod_{i=1}^q \theta_i^{\mathbf{j}_{q,i}^k}$. The following theorem shows that the LR test admits a linear-quadratic approximation in which the linear term coincides with the influence functions underlying our proposed test and the quadratic form has the variance of the influence functions, $V_{\boldsymbol{\theta\theta}}$, playing the role of the information matrix:

Theorem 2 *If Assumptions 1, 2, and 4 hold with* $r = \max\{r_1, ..., r_q\}$ *and* $C = \{\mathbf{j}_d^1, ..., \mathbf{j}_d^K\}$*, respectively, 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

$$\begin{aligned} GET_n(\boldsymbol{\theta}) &= \sup_{\boldsymbol{\theta}} 2n^{\frac{1}{2}}\boldsymbol{\theta}^{\mathbf{j}\prime} n^{-\frac{1}{2}}\mathbf{L}_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) - n^{\frac{1}{2}}\boldsymbol{\theta}^{\mathbf{j}\prime} V_{\boldsymbol{\theta\theta}}(\tilde{\boldsymbol{\phi}}) n^{\frac{1}{2}}\boldsymbol{\theta}^{\mathbf{j}}, \qquad (3) \\ V_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) &= \mathcal{I}_{\boldsymbol{\theta\theta}} - \mathcal{I}_{\boldsymbol{\theta\phi}}\mathcal{I}_{\boldsymbol{\phi\phi}}^{-1}\mathcal{I}_{\boldsymbol{\phi\theta}} \\ \mathcal{I}(\boldsymbol{\phi}) &= \begin{bmatrix} \mathcal{I}_{\phi\phi}(\boldsymbol{\phi}) & \mathcal{I}_{\phi\theta}(\boldsymbol{\phi}) \\ \mathcal{I}_{\theta\phi}(\boldsymbol{\phi}) & \mathcal{I}_{\theta\theta}(\boldsymbol{\phi}) \end{bmatrix}, \text{ with} \\ \mathcal{I}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}) &= Var\left[l_i^{[\boldsymbol{\iota}_p, \mathbf{0}]}(\boldsymbol{\phi}, \mathbf{0})\right], \\ \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}) &= Var\left\{\begin{bmatrix} l_i^{[\mathbf{0}, \mathbf{j}_q^1]}(\boldsymbol{\phi}, \mathbf{0}) & \dots & l_i^{[\mathbf{0}, \mathbf{j}_q^K]}(\boldsymbol{\phi}, \mathbf{0}) \end{bmatrix}'\right\} \text{ and} \\ \mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}) &= E\left\{l_i^{[\boldsymbol{\iota}_p, \mathbf{0}]}(\boldsymbol{\phi}, \mathbf{0})\begin{bmatrix} l_i^{[\mathbf{0}, \mathbf{j}_q^1]}(\boldsymbol{\phi}, \mathbf{0}) & \dots & l_i^{[\mathbf{0}, \mathbf{j}_q^K]}(\boldsymbol{\phi}, \mathbf{0}) \end{bmatrix}\right\}. \end{aligned}$$

Importantly, we show in the proof of Theorem 2 that we can interpret $L_n(\tilde{\boldsymbol{\phi}}, \mathbf{0}) + GET_n$ as a Taylor approximation of order $2r$ to the log-likelihood function around $\tilde{\boldsymbol{\rho}}$, which means that GET_n is effectively an LR-type test that compares the log-likelihood function under the null to the maximum of its lowest-order polynomial approximation under the alternative capable of identifying the restricted parameters. Unsurprisingly, the rate of convergence of the difference between the LR and GET tests is inversely proportional to the highest order of identification.

Although Theorem 2 is substantially more general than Theorem 1, unfortunately Assumption 4.3 excludes some examples of interest in which the covariance matrix of the influence functions is singular, such as the SNP distribution in section 3.2 below. To be able to consider such cases, next we generalize the conditions in Assumptions 3 and 4. Specifically, let $\varsigma_{\boldsymbol{\phi}i}(\boldsymbol{\phi})$

and $\boldsymbol{\varsigma}_{\boldsymbol{\theta}i}(\boldsymbol{\phi})$ denote two measurable functions of dimensions $p \times 1$ and $m \times 1$, respectively, so that we can define the empirical process

$$\mathcal{S}_n'(\boldsymbol{\phi}) = [\ \mathcal{S}_{\boldsymbol{\phi},n}'(\boldsymbol{\phi}) \quad \mathcal{S}_{\boldsymbol{\theta},n}'(\boldsymbol{\phi})\ ] = \textstyle\sum_{i=1}^n \boldsymbol{\varsigma}_i'(\boldsymbol{\phi}), \quad \text{where} \quad \boldsymbol{\varsigma}_i'(\boldsymbol{\phi}) = [\ \boldsymbol{\varsigma}_{\boldsymbol{\phi}i}'(\boldsymbol{\phi}) \quad \boldsymbol{\varsigma}_{\boldsymbol{\theta}i}(\boldsymbol{\phi})\ ].$$

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

$$\boldsymbol{\lambda}'(\boldsymbol{\phi}, \boldsymbol{\theta}) = [\ (\boldsymbol{\phi} - \boldsymbol{\phi}^*)' + \boldsymbol{\lambda}_{\boldsymbol{\phi}}'(\boldsymbol{\theta}) \quad \boldsymbol{\lambda}_{\boldsymbol{\theta}}'(\boldsymbol{\theta})\ ],$$

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

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

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

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

Assumption 5 *(LQ approximation) L_n has a "linear-quadratic" expansion given by*

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

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

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

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

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

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

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

(5.4) 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 $n \to \infty$.
(5.5) It holds that

$$E\left[\partial \boldsymbol{\varsigma}_i(\boldsymbol{\phi}) / \partial \boldsymbol{\phi}'\right] = -(\ \mathcal{I}_{\boldsymbol{\phi\phi}} \quad \mathcal{I}_{\boldsymbol{\phi\theta}}\ ).$$

(5.6) There exists some function $\mathbf{g}(\mathbf{y})$ satisfying $E\{[\mathbf{g}(\mathbf{y}_i)]^2\} < \infty$ such that

$$\left\| \partial \mathcal{S}_n(\boldsymbol{\phi}^\dagger) / \partial \boldsymbol{\phi}' - \partial \mathcal{S}_n(\boldsymbol{\phi}^*) / \partial \boldsymbol{\phi}' \right\| \leq ||\boldsymbol{\phi}^\dagger - \boldsymbol{\phi}^*|| \textstyle\sum_{i=1}^n \mathbf{g}(\mathbf{y}_i) \tag{4}$$

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

Assumption 5 states that the likelihood ratio can be expressed as the sum of a linear-quadratic approximation and a residual term, R_n. The linear-quadratic part represents a higher-order expansion of the likelihood ratio around $\boldsymbol{\theta} = \mathbf{0}$. Assumption 5.1 captures the local identification condition at the true parameter value. In addition, $\mathcal{K}(\boldsymbol{\phi}^*, \boldsymbol{\phi}^*)$ in Assumptions 5.2 and 5.3 plays the role of a full-rank information matrix, while 5.5 is analogous to the generalized information matrix equality. In turn, Assumption 5.4 ensures that the residual is dominated by the leading terms, and thus, negligible asymptotically, while Assumption 5.6 enables us to substitute the true parameter $\boldsymbol{\phi}^*$ with the restricted estimator $\tilde{\boldsymbol{\phi}}$ after an appropriate adjustment for sampling variability. Finally, Assumption 5.7 allows us to obtain the convergence rate of the linear-quadratic approximation, with a typically associated with the slowest rate of convergence of the parameter estimators under the null.

We can then prove the following result, which nests our first and second theorems:

Theorem 3 *If Assumptions 1 and 5.1– 5.6 hold, then under* $H_0 : \boldsymbol{\theta} = \mathbf{0}$

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

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

In addition, if 5.7 holds, then

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

Finally, it is worth mentioning that even though GET cannot be directly understood as a moment test, a by-product of this theorem is a set of influence functions $\mathcal{S}_n(\boldsymbol{\phi}, \mathbf{0})$ that can be used for that purpose after taking into account the sampling uncertainty in estimating $\boldsymbol{\phi}$ under the null. In fact, we can prove that this moment test, which converges in distribution to a $\chi^2_{\dim[\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})]}$ under H_0, provides an upper bound to GET_n, albeit a rather loose one in most cases.

2.3 Asymptotic validity of the parametric bootstrap

As we mentioned before, the asymptotic distribution of the GET statistic may not be very reliable in finite samples. In addition, given that it often has a non-standard form, obtaining critical values requires simulations in any event. For this reason, we recommend using parametric bootstrap procedures to generate samples from the distribution under the null hypothesis. This is achieved by sampling from the fully specified parametric model of interest, with the unknown true parameter values replaced by their restricted maximum likelihood estimates. To simplify the exposition, in this section we proceed as if there were no regressors in the model, although our proofs explicitly allow for strictly exogenous regressors. In this context, our proposed procedure to obtain critical values is:

1. Simulate a sample $\{\mathbf{y}_i^{(s)}\}_{i=1}^n$ from $f[\mathbf{y}_i; (\tilde{\boldsymbol{\phi}}, \mathbf{0})]$

2. Compute $GET^{(s)}$ using the simulated sample $\{\mathbf{y}_i^{(s)}\}_{i=1}^n$

3. Repeat Step 1 and 2 S times. The critical value cv_n is then defined as the $1-\alpha$ quantile of $GET^{(s)}$.

To prove the asymptotic validity of the parametric bootstrap, we assume that:

Assumption 6 *(Bootstrap Validity)*

1. *There is an open set $\mathcal{N}_{\boldsymbol{\phi}}$ around $\boldsymbol{\phi}^*$ such that for all sequences $\boldsymbol{\phi}_n \in \mathcal{N}_{\boldsymbol{\phi}}$ that converge to $\boldsymbol{\phi}^*$ we have that under DGP $(\boldsymbol{\phi}_n, \mathbf{0})$:*

$$n^{-1}\partial\mathcal{S}_n(\boldsymbol{\phi}_n)/\partial\boldsymbol{\phi}' \xrightarrow{p} -\left[\begin{array}{cc} \mathcal{I}_{\phi\phi}(\boldsymbol{\phi}^*) & \mathcal{I}_{\phi\theta}(\boldsymbol{\phi}^*) \end{array}\right]' \text{ and}$$

$$n^{-1/2}\mathcal{S}_n(\boldsymbol{\phi}_n) \xrightarrow{d} \mathcal{S}(\boldsymbol{\phi}^*).$$

2. *There is some function $g(\mathbf{y})$ satisfying $\sup_{\boldsymbol{\phi}\in\mathcal{N}_{\boldsymbol{\phi}}} E\left[g^2(\mathbf{y}); (\boldsymbol{\phi}, 0)\right] < \infty$ such that with probability 1,*

$$\left\|\partial\mathcal{S}_n(\boldsymbol{\phi})/\partial\boldsymbol{\phi}' - \partial\mathcal{S}_n(\boldsymbol{\phi}^\dagger)/\partial\boldsymbol{\phi}'\right\| \leq \left\|\boldsymbol{\phi} - \boldsymbol{\phi}^\dagger\right\| \sum_i g(\mathbf{y}_i)$$

for all $\boldsymbol{\phi}$ and $\boldsymbol{\phi}^\dagger$ in $\mathcal{N}_{\boldsymbol{\phi}}$.

In addition, we assume that the $\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ that appears in Assumption 5 is well-approximated by a cone $\boldsymbol{\Lambda}$ at $\mathbf{0}$, which allows us to cover the statistic in our general Theorem 3. Specifically

Assumption 7 *(Cone cover) $\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is Chernoff regular at $\mathbf{0}$. Specifically, $\inf_{\mathbf{w}\in\boldsymbol{\Lambda}} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - \mathbf{w}\| = o\left(\|\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\|\right)$ for all $\boldsymbol{\theta} \in \Theta$, and $\inf_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - \mathbf{w}\| = o(\|\mathbf{w}\|)$ for all $w \in \boldsymbol{\Lambda}$.*

The following result confirms the asymptotic validity of the parametric bootstrap procedure above under the assumption that the number of simulations S is so large that the difference between the true $1-\alpha$ quantile of $GET^{(s)}$ and its simulated estimate from S samples is $o_p(1)$:

Theorem 4 *(Validity of parametric bootstrap.)*

1. *If Assumption 1, 5, 6, and 7 hold, then under H_0*

$$GET_n \xrightarrow{d} GET, \tag{5}$$

where $$GET = \sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \left\{2\mathcal{S}'\boldsymbol{\lambda} - \boldsymbol{\lambda}\left[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)\right]\boldsymbol{\lambda}\right\},$$

and for almost all sequences $\{\mathbf{y}_i\}_{i\geq 1}$

$$GET_n^{(s)}|\{\mathbf{y}_i\}_{i=1}^n \xrightarrow{d} GET. \tag{6}$$

2. *If GET has a cumulative distribution function continuous at its $1-\alpha$ quantile, then*

$$\lim_n \Pr\left(GET_n > cv_n\right) = \alpha$$

under the following set of assumptions: (i) 1, 2, 3, and 6; (ii) 1, 2, 4, and 6; or (iii) 1, 5, 6, and 7, with the definition of GET_n adjusted accordingly to match the ones in Theorems 1, 2 and 3, respectively.

Thus, we can automatically compute size-adjusted rejection rates without knowing the true DGP, as recommended by Horowitz and Savin (2000).

2.4 Distribution under local alternatives

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

$$H_{1n} : n^{1/2}[\ \boldsymbol{\lambda}'_{\boldsymbol{\phi}}(\boldsymbol{\theta}_n) \quad \boldsymbol{\lambda}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}_n)\ ] \rightarrow (\ \boldsymbol{\lambda}'_{\phi,\infty} \quad \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}\ ) = \boldsymbol{\lambda}'_{\infty} \in \mathbb{R}^{\dim(\boldsymbol{\lambda}_{\infty})}.$$

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

Theorem 5 *(Distribution under local alternatives)*
(5.1) $P_{\boldsymbol{\theta}_n}$ is contiguous with respect to P_0.
(5.2) Under H_{1n} and Assumptions 1 and 5,

$$n^{-1/2}\mathcal{S}_n(\boldsymbol{\phi}^*) \xrightarrow{d} N[\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_{\infty}, \mathcal{I}(\boldsymbol{\phi}^*)].$$

(5.3) Under H_{1n} and Assumptions 1, 5, 6 and 7,

$$\begin{aligned} GET_n \xrightarrow{d} \sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \big\{ & 2S + \boldsymbol{\lambda}'_{\theta,\infty}[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*)]\boldsymbol{\lambda} \\ & -\boldsymbol{\lambda}\left[\mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*)\right]\boldsymbol{\lambda}\Big\}, \quad \textit{where} \\ S \sim \mathcal{N}[0, & \mathcal{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta\phi}}(\boldsymbol{\phi}^*)\mathcal{I}^{-1}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi\theta}}(\boldsymbol{\phi}^*)]. \end{aligned}$$

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

3 Examples

In this section, we discuss the application of our proposed test to the first two examples of empirical interest that we mentioned in the introduction. Specifically, we derive a test for irrelevant sample selectivity in multivariate regression models, for which Theorem 1 suffices, and a test for normality against SNP alternatives, which requires our more general Theorem 3. In turn, in Supplemental Appendix D we obtain a test of a multivariate normal copula against its Hermite expansion, which is another example of Theorem 1 but with the added difficulty of inequality constraints on the parameters. Finally, in Supplemental Appendix E, we derive a test aimed at detecting non-linear predictability in a multiple regressor version of Bottai (2003), for which Theorem 2 suffices (see also Amengual, Bei and Sentana (2022, 2024) for other empirically-relevant applications of Theorems 1 and 2).

3.1 Example 1: Testing for selectivity in multivariate regressions

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

3.1.1 Model, likelihood and null hypothesis of no selectivity

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

$$\mathbf{y} = \mathbf{y}^* d, \tag{7}$$

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

$$d = \mathbf{1}(\mathbf{w}'\boldsymbol{\varphi}^S + u_S \geq 0), \tag{8}$$

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

$$y_k^* = \boldsymbol{\varphi}_k^{M\prime}\mathbf{x} + \varphi_k^D u_k, \quad k = 1, \ldots, K, \tag{9}$$

$$\begin{pmatrix} \mathbf{u} \\ u_S \end{pmatrix} |\mathbf{x}, \mathbf{w} \sim N\left\{\mathbf{0}, \begin{bmatrix} \boldsymbol{R}(\boldsymbol{\varphi}^L) & \boldsymbol{\vartheta} \\ \boldsymbol{\vartheta}' & 1 \end{bmatrix}\right\}, \tag{10}$$

with $\mathbf{x}$ being a vector of exogenous regressors that may partially overlap with $\mathbf{w}$, $\mathbf{u} = (u_1, ..., u_K)$, so that $\boldsymbol{\varphi}^D = (\varphi_1^D, \ldots, \varphi_K^D)'$ contains the standard deviations of the regression shocks, $\boldsymbol{\varphi}^L$ the correlations between them, and $\boldsymbol{\vartheta}$ the correlations between those shocks and the unobserved component of the selection equation, whose variance we normalize to 1 without loss of generality.

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

$$(1-d)\ln\Phi\left(-\mathbf{w}'\boldsymbol{\varphi}^S\right) + d\ln\Phi\left[\frac{\mathbf{w}'\boldsymbol{\varphi}^S + \boldsymbol{\vartheta}'\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)}{\sqrt{1-\boldsymbol{\vartheta}'\boldsymbol{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\boldsymbol{\vartheta}}}\right]$$
$$-\frac{d}{2}\left[2\sum\nolimits_{k=1}^K \ln\varphi_k^D + \ln\{\det[\boldsymbol{R}\left(\boldsymbol{\varphi}^L\right)]\} + \mathbf{u}'(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)\boldsymbol{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\mathbf{u}(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)\right], \tag{11}$$

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

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

Under the assumption that the unobserved selectivity determinants are uncorrelated with the regression residuals, one can efficiently estimate the multivariate regression coefficients $\boldsymbol{\varphi}^M$ together with the covariance matrix parameters $\boldsymbol{\varphi}^D$ and $\boldsymbol{\varphi}^L$ without selection bias from the non-zero values of $\mathbf{y}$ only using equation by equation OLS without the need to consider the model for d. However, when this assumption does not hold, those OLS estimators will be biased because of the sample selectivity, which justifies testing the null hypothesis $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$.

3.1.2 Singularities, reparametrizations and GET test statistic

To simplify the presentation, we consider the case in which $w = 1$ and the regression equations contain a constant term, so that at $\boldsymbol{\vartheta} = \mathbf{0}$,

$$s_{\vartheta_k} - M_1(\varphi^S)\varphi_k^D s_{\varphi_{k1}^M} = 0 \quad k = 1, \ldots, K \tag{12}$$

where $M_1(\boldsymbol{\varphi}^S)$ represents the standard inverse Mills ratio. Consequently, the nullity of the information matrix is K. As Lee and Chesher (1986) explain in the univariate model that they considered, analogous singularities will arise for example when the observed selectivity determinants $\mathbf{w}$ are given by a set of dummy variables and $\mathbf{x}$ contains those dummy variables too. In general, singularities will be present whenever Heckman's (1976) selectivity correction is perfectly collinear with the regressors that appear in the conditional means of the y^*'s even though the log-likelihood function in (11) is able to locally identify all the model parameters.

In this set-up, the general reparametrization method in Supplemental Appendix C, which is characterized by equations (C23)-(C25), yields:

$$\varphi_{k1}^M = \varphi_{k1}^{M\dagger} - M_1(\boldsymbol{\varphi}^S)\varphi_k^D\vartheta_k^\dagger,\ \ \boldsymbol{\varphi}^D = \boldsymbol{\varphi}^{D\dagger},\ \ \boldsymbol{\varphi}^L = \boldsymbol{\varphi}^{L\dagger},\ \ \boldsymbol{\varphi}^S = \boldsymbol{\varphi}^{S\dagger},\ \ \boldsymbol{\varphi}_{k(-1)}^M = \boldsymbol{\varphi}_{k(-1)}^{M\dagger}, \text{ and } \ \boldsymbol{\vartheta} = \boldsymbol{\vartheta}^\dagger.$$

By construction, we now have $s_{\vartheta_k^\dagger} = 0$ for all k. However, there are $K(K+1)/2$ linear combinations of the scores and the elements of the Hessian corresponding to $\boldsymbol{\vartheta}$ that are 0 too. Specifically, denoting by $\boldsymbol{\varphi}_{ij}^L$ the correlations between u_i and u_j for $i, j = 1, \ldots, K$, we have

$$\frac{\partial^2 \ell}{\left(\partial\vartheta_k^\dagger\right)^2} - M_1\left(\boldsymbol{\varphi}^{S\dagger}\right)\left[\boldsymbol{\varphi}^{S\dagger} + M_1\left(\boldsymbol{\varphi}^{S\dagger}\right)\right]\left(\sum_{j\neq k}\boldsymbol{\varphi}_{jk}^{L\dagger}s_{\boldsymbol{\varphi}_{jk}^{L\dagger}} - \varphi_k^{D\dagger}s_{\varphi_k^{D\dagger}}\right) = 0, \text{ for } k = 1, \ldots, K$$

$$\frac{\partial^2 \ell}{\partial\vartheta_i^\dagger\partial\vartheta_j^\dagger} + M_1\left(\boldsymbol{\varphi}^{S\dagger}\right)\left[\boldsymbol{\varphi}^{S\dagger} + M_1\left(\boldsymbol{\varphi}^{S\dagger}\right)\right]s_{\boldsymbol{\varphi}_{ij}^{L\dagger}} = 0 \text{ for } i > j, \quad i, j = 1, \ldots, K.$$

To circumvent this problem, we can apply (C23)-(C25) again as follows:

$$\varphi_{ij}^{L\dagger} = \phi_{ij}^L - M_1\left(\phi^S\right)\left[\phi^S + M_1\left(\phi^S\right)\right]\left(\frac{1}{2}\phi_{ij}^L\theta_i^2 + \frac{1}{2}\phi_{ij}^L\theta_j^2 - \theta_i\theta_j\right),$$
$$\boldsymbol{\varphi}^{M\dagger} = \boldsymbol{\phi}^M, \ \ \boldsymbol{\varphi}^{S\dagger} = \boldsymbol{\phi}^S, \text{ and } \ \boldsymbol{\vartheta}^\dagger = \boldsymbol{\theta}.$$

After this second reparametrization, Assumption 3 is satisfied with $r = 3$, $q_1 = 0$ and $q_3 = K$, so that we can work with third derivatives. Besides, since these have a full-rank asymptotic covariance matrix under the null, we can apply Theorem 1, which somewhat remarkably, leads to the following result:

Proposition 1 *The difference between LR test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ in model (8)-(10) based on a random sample of n observations on $(\mathbf{y},d)$ and the following test statistic*

$$GET_n = \sup_{\boldsymbol{\upsilon}\neq\mathbf{0}} \left(\sum_{i=1}^n d_i\right)^{-1} \left\{\sum_{i=1}^n d_i H_3[(\boldsymbol{\upsilon}'\boldsymbol{\upsilon})^{-1/2}\boldsymbol{\upsilon}'\boldsymbol{\varepsilon}_i(\tilde{\boldsymbol{\varphi}}^M,\tilde{\boldsymbol{\varphi}}^D)]\right\}^2 \tag{13}$$

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

In simpler terms, our test statistics numerically coincides with the supremum of the moment tests for univariate skewness based on the third Hermite polynomial over all possible linear combinations of the OLS residuals that have 0 mean and unit variance in the sample of observations with $d = 1$. In fact, the standardization is unnecessary because the moment test for univariate skewness is numerically invariant to affine transformations of the observations, which in turn confirms that the test statistic (13) is homogeneous of degree 0 in $\boldsymbol{\upsilon}$. Thus, when $K = 1$ our proposed test reduces to the test for selectivity derived by Lee and Chesher (1986) in the univariate case, which simply assesses the symmetry of the regression residuals by looking at the sample mean of their third power.

The rationale is also analogous in the multivariate case. Equations (7)-(10) imply that the OLS residuals should be approximately multivariate normally distributed when the unobserved component of the sample selection is independent of the shocks to the observed variables. Under the alternative, in contrast, asymmetry becomes a common feature, as in the multivariate skew normal distribution discussed in Amengual, Bei and Sentana (2022). Intuitively, if we orthogonalize the regression residuals with respect to the unobserved component of the selectivity equation, then we end up with $\boldsymbol{\vartheta}u_S$ as a common component, whose distribution conditional on $d = 1$ is asymmetric even though the unconditional distribution of u_S is symmetric.

3.1.3 Local power analysis

Although the null distribution of the test statistic (13) is non-standard, we can still say something about the determinants of its local power. Consider the following sequence of local alternatives:

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

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

Proposition 2 *The local power of the test in Proposition 1 only depends on the magnitude of the quadratic form* $\boldsymbol{\vartheta}_\infty' \mathbf{R}^{-1}\left(\boldsymbol{\varphi}^L\right)\boldsymbol{\vartheta}_\infty$.

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

3.1.4 Simulation evidence

For simplicity, we let $w = x_1 = 1$ and $x_2 \sim N(0,1)$. Given that the restricted MLE of the multivariate regression coefficients is equation by equation OLS, and that all regressions contain an intercept, the sample mean of the multivariate regression residuals $\hat{\mathbf{u}}$ will be a vector of K zeros. Similarly, any orthogonalization of the $\hat{\mathbf{u}}'s$ based on the estimated covariance matrix will have the identity matrix as sample covariance matrix because the MLEs of the residual standard deviations $\boldsymbol{\varphi}^D$ and correlations $\boldsymbol{\varphi}^L$ match perfectly the sample variances and covariances of $\hat{\mathbf{u}}$ with denominator $\sum_{i=1}^n d_i$. Therefore, it is not surprising that the particular square root that orthonormalizes the OLS residuals in the sample is numerically irrelevant. For example, in the bivariate case, we could either define ε_1 as the standardized value of u_1 and ε_2 as the standardized value of the residual in the OLS regression of u_2 on a constant and u_1, or vice versa.

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

If we exploited our knowledge of the values of these two parameters, we could compute exact critical values under the null for any sample size to any degree of accuracy by repeatedly simulating samples from the true distribution. In practice, though, we fix the selection parameter and the correlation coefficient to their estimated values in each sample, as explained in section 2.3. Given that we can verify that the LR test statistic is also numerically invariant to the true values $(\boldsymbol{\varphi}^M, \boldsymbol{\varphi}^D)$, we can approximate its critical value using the same parametric bootstrap procedure.

However, the maximization of the unrestricted likelihood function is highly sensitive to the choice of initial values. In fact, the presence of first and second derivatives that are identically 0 under the null implies that most numerical algorithms will not move if we start from the restricted parameter estimators. For that reason, we employ the MATLAB function `fmincon` using four different sets of initial values in which we set $\boldsymbol{\phi}$ to the restricted MLE $\tilde{\boldsymbol{\phi}}$ and the two elements of $\boldsymbol{\vartheta}$ to the four possible combinations of ± 0.1. We then define the LR statistic as the maximum of the LR statistics associated to these four sets of initial values.[8]

We compare the results of our tests with a bootstrap-based LR test, as well with a *naive*

[8]For data simulated under the null hypothesis, we observe that approximately 35% of these four LR statistics have a value with is less than 1/2 of the largest one irrespectively of whether $n = 400$ or $n = 1600$.

version, denoted as LR_{naive}, that computes p-values as if the limiting distribution of this statistic was a χ^2_2. Moreover, we also consider the moment test that assesses the four influence functions underlying GET mentioned at the end of section 2.2, which we label as GMM_{asy} when we use asymptotic critical values and GMM when we rely on the parametric bootstrap. For each design, we generate 10,000 samples of size n and compute the restricted and unrestricted parameter estimators together with the aforementioned tests.

Panels A and B of Table 1 report the results for samples of length 400 and 1600, respectively. In turn, the first three columns report rejection rates under the null at the 1%, 5% and 10% levels, confirming that our simulated critical values based on 10,000 bootstrap samples work remarkably well for both sample sizes.[9] The size distortions of GMM and GMM_{asy} are also negligible. In contrast, LR_{naive} is far too liberal, highlighting the practical consequences of the singularity of the information matrix.

In turn, the last six columns present the rejection rates at the 1%, 5% and 10% levels for the following two alternatives: $\boldsymbol{\vartheta}' = (0.57, 0.57)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.80, 0)$ (H_{a2}). As can be seen, our proposed test has similar power to the LR test for the two alternatives, and both these tests outperform the GMM one.

We find a Gaussian rank correlation[10] of 0.88 (0.95) between our proposed test statistic and the LR across Monte Carlo simulations of 400 (1600) observations that satisfy the null displayed in Figure 1, which is in line with the slow rate of convergence in Proposition 1. In addition, our results indicate that the LR takes about 10 and 20 times as much CPU time to compute as GET does for $n = 400$ and $n = 1600$, respectively, which makes a huge difference in the calculation of the critical values with the parametric bootstrap.

3.2 Example 2: Testing for normality against SNP alternatives

Gram-Charlier expansions provide flexible and analytically tractable generalizations of the normal distribution. Unfortunately, their truncated versions lead to negative density values, and the parametric restrictions that Jondeau and Rockinger (2001) propose to guarantee positivity are not easy to implement even when the truncation order is low. In contrast, the SNP distributions introduced by Gallant and Nychka (1987) provide a Hermite expansion of the Gaussian density that is positive by construction. Although these authors introduced those distributions for nonparametric estimation purposes, León, Mencía and Sentana (2009) treated them as parametric ones, studied their statistical properties, and used them in option valuation. Still, MLE under normality is much simpler than when the distribution of the shocks follows an SNP. For that reason, we shall derive a test of normality that will also highlight the hidden complications researchers face in this context.

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

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

3.2.1 Model, likelihood and null of normality

The model we consider is

$$y = \mu(\mathbf{x}, \boldsymbol{\alpha}) + \sigma(\mathbf{x}, \boldsymbol{\alpha})\, u \tag{14}$$

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

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

$$f(y; \boldsymbol{\varrho}) = \frac{1}{\sqrt{\varphi^V}} \phi\left(\frac{y - \varphi^M}{\sqrt{\varphi^V}}\right) \left[\epsilon + \frac{(1-\epsilon)\{P[(y - \varphi^M)/\sqrt{\varphi^V}; \boldsymbol{\vartheta}]\}^2}{\int_{-\infty}^{\infty} \{P[u; \vartheta]\}^2 \phi(u) du}\right], \tag{15}$$

with

$$P[u; \boldsymbol{\vartheta}] = 1 + \sum\nolimits_{k=1}^{K} \vartheta_k H_k(u), \tag{16}$$

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

$$\sqrt{k} H_k(u) = u H_{k-1}(u) - \sqrt{k-1} H_{k-2}(u), \tag{17}$$

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

3.2.2 Singularities, reparametrizations and GET test statistic

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

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

under H_0, so that the nullity of the information matrix is 2. Hall (1990) highlighted this problem when he considered tests of normality against semi-nonparametric alternatives in which the ϑ coefficients were in turn functions of some exogenous variable. However, his proposed solution was to ignore the parameters involved in the singularity, focusing instead only on those which could be regularly estimated under the null. Unfortunately, his recipe would leave us with no test in the case of the unconditional model (15)-(16).

If we apply reparametrization (C23)-(C25), we obtain

$$\varphi^M = \varphi^{M\dagger} - 2\sqrt{\varphi^{V\dagger}}\vartheta_1^\dagger \text{ and } \varphi^V = \varphi^{V\dagger} - 2\sqrt{2}\varphi^{V\dagger}\vartheta_2^\dagger,$$

with $\vartheta_1 = \vartheta_1^\dagger$ and $\vartheta_2 = \vartheta_2^\dagger$. By construction, we now have $s_{\vartheta_1^\dagger} = s_{\vartheta_2^\dagger} = 0$. However, additional singularities arise. Specifically,

$$\frac{\partial^2 \ell}{(\partial\vartheta_1^\dagger)^2} + 4\varphi^{V\dagger} s_{\varphi^{V\dagger}} = 0.$$

We can circumvent this problem by implementing a second reparametrization along the lines of (C23)-(C25). In particular, we can define $\varphi^{V\dagger} = \varphi^{V\star} + 2\varphi^{V\star}(\vartheta_1^\star)^2$, with $\varphi^{M\dagger} = \varphi^{M\star}$, $\vartheta_1^\dagger = \vartheta_1^\star$ and $\vartheta_2^\dagger = \vartheta_2^\star$, which achieves $\partial^2\ell/(\partial\vartheta_1^\star)^2 = 0$ but leads to

$$\frac{\partial^3 \ell}{(\partial\vartheta_1^\star)^3} + 2\sqrt{2}\frac{\partial^2 \ell}{\partial\vartheta_1^\star\partial\vartheta_2^\star} + 12\sqrt{\varphi^{V\star}} s_{\varphi^M} = 0.$$

To resolve this new singularity, we apply a third and final reparametrization, namely

$$\varphi^{M\star} = \phi^M + 2\sqrt{\phi^V}\theta_1^3, \quad \varphi^{V\star} = \phi^V, \quad \vartheta_1^\star = \theta_1, \quad \vartheta_2^\star = \theta_2 + \frac{\sqrt{2}}{3}\theta_1^2,$$

after which we obtain:

$$\frac{\partial l}{\partial\phi^M} = \frac{1}{\sqrt{\phi^V}}H_1(u), \quad \frac{\partial l}{\partial\phi^V} = \frac{1}{\sqrt{2}\phi^V}H_2(u), \quad \frac{\partial l}{\partial\theta_1} = \frac{\partial l}{\partial\theta_2} = \frac{\partial^2 l}{(\partial\theta_1)^2} = \frac{\partial^3 l}{(\partial\theta_1)^3} = 0,$$

$$\frac{1}{2}\frac{\partial^2 l}{(\partial\theta_2)^2} = -\sqrt{6}H_4(u) - 2\sqrt{2}H_2(u), \quad \frac{\partial^2 l}{\partial\theta_1\partial\theta_2} = -2\sqrt{3}H_3(u) \text{ and } \frac{1}{4!}\frac{\partial^4 l}{(\partial\theta_1)^4} = \frac{\sqrt{6}}{9}H_4(u) - \frac{13\sqrt{2}}{9}H_2(u).$$

By performing an eighth-order Taylor expansion, we can verify that Assumption 5 holds with

$$\mathcal{S}_\phi = [H_1(u^*)/\sqrt{\phi_2^*}, H_2(u^*)/\sqrt{2}\phi_2^*]', \quad \mathcal{S}_\theta = [H_3(u^*), H_4(u^*)]',$$

$$\boldsymbol{\lambda}_\phi(\boldsymbol{\theta}) = \left[0, -\sqrt{2}\phi_2^*\left(\frac{13\sqrt{2}}{9}\theta_1^4 + 2\sqrt{2}\theta_2^2\right)\right] \text{ and } \boldsymbol{\lambda}_\theta(\boldsymbol{\theta}) = \left(-2\sqrt{3}\theta_1\theta_2, \frac{\sqrt{6}}{9}\theta_1^4 - \sqrt{6}\theta_2^2\right).$$

The derivations above indicate that θ_1 and θ_2 have different orders of identification, and that the second derivative with respect to θ_2 and the fourth derivative with respect to θ_1 are proportional, which means that we need to resort to our Theorem 3. On this basis, we can establish the following result:

Proposition 3 *The difference between the LR test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ in model (15)-(16) based on a random sample of n observations on $\mathbf{y}$ and the following test statistic*

$$GET_n = n\left\{\left[n^{-1}\textstyle\sum_{i=1}^n H_3(\tilde{u}_i)\right]^2 + \left[n^{-1}\textstyle\sum_{i=1}^n H_4(\tilde{u}_i)\right]^2\right\} \tag{18}$$

is $O_p(n^{-1/8})$ when the null is true, where $H_3(\tilde{u}_i)$ and $H_4(\tilde{u}_i)$ are the third- and fourth-order normalized Hermite polynomials of the $\tilde{u}_i'$s, which are the values of the y_i's standardized so that their sample mean and variance are 0 and 1, respectively.

Remarkably, this means that the Jarque and Bera (1980) test, whose asymptotic distribution is a standard χ_2^2 under the null, is asymptotically equivalent to the LR test of normality against SNP densities, although they converge to each other at a much lower rate than in the case of the Pearson family of alternative distributions they considered.

3.2.3 Local power analysis

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

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

$$\lim_{n\to\infty} n^{1/2} \left[\begin{array}{cc} -2\sqrt{3}\theta_{1,n}\theta_{2,n} & \sqrt{6}(\frac{1}{9}\theta_{1,n}^4 - \theta_{2,n}^2) \end{array} \right] = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}' \in \mathbb{R}^2. \tag{19}$$

Under the sequence of DGPs indexed by $\boldsymbol{\theta}_n$, $GET_n \xrightarrow{d} \chi_2^2(\boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}'\boldsymbol{\lambda}_{\boldsymbol{\theta},\infty})$.

To understand this result, it is useful to note that

$$\sqrt{n}E\left\{H_3\left[(y-\varphi^L)/\sqrt{\varphi^V}\right], H_4\left[(y-\varphi^L)/\sqrt{\varphi^V}\right]\right\} = \boldsymbol{\lambda}_{\boldsymbol{\theta},\infty}' + o(1).$$

Unlike in the multivariate regression model with selectivity, though, we can have two different types of local alternatives compatible with (19):

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

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

3.2.4 Simulation evidence

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

Once again, the maximization of the unrestricted likelihood function is highly sensitive to the choice of initial values. As in the previous example, the presence of several first and higher-order derivatives that are identically 0 under the null implies that most numerical algorithms will not move if we start from the restricted parameter estimators. For that reason, we employ the

MATLAB function `fmincon` using three different initial values in which we set $\boldsymbol{\phi}$ to the restricted MLE $\tilde{\boldsymbol{\phi}}$. As for the the two elements of $\boldsymbol{\vartheta}$, we consider $\pm(.1,.1)$ and $\boldsymbol{\theta}^{ET}$, which we obtain by maximizing an eighth-order Taylor expansion of the log-likelihood function (see the proof of Proposition 3 for details). We then define the LR statistic as the maximum of the LR statistics associated to these three sets of initial values.[11]

As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.25, 0.10)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.75, 0.05)$ (H_{a2}), setting $\varphi^M = 0$ and $\varphi^V = 1$ without loss of generality. We then generate 10,000 samples of size n for each design, and compute the restricted and unrestricted parameter estimators together with the GET and LR tests. Panels A and B of Table 2 report the results for samples of size 400 and 1600, respectively. Columns five to seven report rejection rates under H_{a1} at the 1%, 5% and 10% levels, while the last three columns present the rejection rates for H_{a2} at the same levels. As can be seen, our proposed test has similar power to the LR test for both alternatives.

We find a Gaussian rank correlation of 0.90 (0.91) between our proposed test statistic and the LR across Monte Carlo simulations of 400 (1600) observations that satisfy the null displayed in Figure 2, which is in line with the very slow rate of convergence in Proposition 4. However, our results also indicate that the LR takes around 160 and 100 times as much CPU time to compute as GET does for $n = 400$ and $n = 1600$, respectively, considerably slowing down the calculation of simulation-based critical values.

3.2.5 Robustness to the estimation of mean and variance parameters

We now extend our previous results to a situation in which the conditional mean and variance of y are parametric functions of the variable in $\mathbf{x}$, as in (14). In this context, the objective becomes to test whether the standardized innovation u follows a normal distribution rather than an SNP one.

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

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

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

Assumption 8 *(Smoothness of the conditional first two moments) The conditional mean and variance functions $\mu_Y(\mathbf{x}_i, \boldsymbol{\alpha})$ and $\sigma_Y(\mathbf{x}_i, \boldsymbol{\alpha})$ that appear in (14) are such that:*
(8.1) They are eight times continuously differentiable with respect to $\boldsymbol{\alpha}$.
(8.2) For all $\boldsymbol{k} = (k_1, \ldots, k_{d_{\boldsymbol{\alpha}}})' \in N^{d_{\boldsymbol{\alpha}}}$ and $\boldsymbol{\iota}'\boldsymbol{k} = 1, \ldots, 8$, it holds that

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

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

[11] For data simulated under the null hypothesis, we observe that approximately 20% of these three LR statistics have a value which is less than 60% of the largest one irrespectively of whether $n = 400$ or $n = 1600$.

Then, we can prove the following result, which is entirely analogous to Proposition 8 in Amengual et al (2025):

Proposition 5 *Under Assumption 8, replacing the true value of* $\boldsymbol{\alpha}$ *by* $\tilde{\boldsymbol{\alpha}}$*, its restricted maximum likelihood estimator under* H_0*, does not alter the expressions of the GET test in Proposition 3 or its asymptotic distribution under the null or sequences of local alternatives.*

4 Conclusions

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

Our results suggest some additional theoretical developments. For example, Amengual et al (2025) build up on our theorem 3 to derive score-based tests for normality against a finite normal mixture even though strictly speaking it does not cover that model. Similarly, the study of GMM overidentification test statistics in contexts in which not only the expected Jacobian matrix is singular but the expected values of some higher-order Jacobian matrices are singular too would constitute a very interesting topic for further research.

From the empirical point of view, the tests developed in this paper allowed Amengual, Bei and Sentana (2022, 2024) to provide some new insights about the cross-section distribution of city sizes and their growth rates and tests for neglected serial correlation in time series models with and without unobserved components, respectively. Their use in some of the other empirically relevant situations discussed in this paper would also provide a particularly valuable complement to our theoretical results.

References

Amengual, D., Bei, X., Carrasco, M. and Sentana, E. (2025): "Score-type tests for normal mixture models", *Journal of Econometrics* 248, 105717.

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

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

Amengual, D., Bei, X. and Sentana, E. (2024): "Highly irregular serial correlation tests", CEMFI Working Paper 2302, revised January 2024, forthcoming in *Econometrics and Statistics*.

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

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

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

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

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

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

Dovonon, P. and Gonçalves, S. (2017): "Bootstrapping the GMM overidentification test under first-order underidentification," *Journal of Econometrics*, 201, 43–71.

Dovonon, P. and Hall, A. (2018): "The asymptotic properties of GMM and indirect inference under second-order identification", *Journal of Econometrics* 205, 76-111.

Dovonon, P., Hall, A. and Kleibergen, F. (2020): "Inference in second-order identified models", *Journal of Econometrics* 218, 346-372.

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

Dovonon, P. and Renault, E. (2020): "GMM overidentification test with first-order underidentification", mimeo, Concordia University.

Engle, R.F. (1983): "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics", in Intriligator, M. D. and Griliches, Z. (eds.), *Handbook of Econometrics* 2, 796–801, Elsevier.

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

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

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

Han, S. and McCloskey, A. (2019): "Estimation and inference with a (nearly) singular Jacobian", *Quantitative Economics* 10, 1019-1068.

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

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

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

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

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

Kleibergen, F. (2005): "Testing parameters in GMM without assuming that they are identified", *Econometrica* 73, 1103-1124.

Lee, J.H. and Liao, Z. (2018): "On standard inference for GMM with local identification failure of known forms", *Econometric Theory* 34, 790-814.

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

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

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

Newey, W. and McFadden, D. (1994): "Large sample estimation and hypothesis testing", in Engle, R. and McFadden, D. (eds.), *Handbook of Econometrics* 4, 2111–2245, Elsevier.

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.

Sentana, E. (2024) "Finite underidentification", *Journal of Econometrics* 240, 105692.

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

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

Appendices

A Proofs

In this appendix we proceed as follows. We first prove Theorem 3, which is the most general result, and then we use it to prove Theorems 1 and 2 as particular cases. The proofs of the remaining theorems follow the order in which they appear in the main text. Finally, we state and prove all the required lemmas in Supplemental Appendix B, which also contains the proofs of the propositions.

Let

$$LR(\boldsymbol{\rho}) = 2\left[L_n(\boldsymbol{\rho}) - L_n(\boldsymbol{\phi}^*, \mathbf{0})\right] \tag{A1}$$

and

$$LM_n(\boldsymbol{\rho}) = 2\mathcal{S}_n'(\boldsymbol{\phi}^*)\boldsymbol{\lambda}\left(\boldsymbol{\phi},\boldsymbol{\theta}\right) - n\boldsymbol{\lambda}'\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)\mathcal{I}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}\left(\boldsymbol{\phi},\boldsymbol{\theta}\right), \tag{A2}$$

where the definition of $\boldsymbol{\lambda}\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)$, $\mathcal{I}(\boldsymbol{\phi}^*)$ and $\mathcal{S}_n(\boldsymbol{\phi}^*)$ depends on the Assumptions invoked. Let

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

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

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

The stochastic sequence a_n is "bounded in probability", or $O_p(1)$, when $\forall\epsilon > 0$, there exists M such that $\Pr(|a_n| < M) \geq 1-\epsilon$ for all n. In addition, we use $a_n = o_p(R_n)$ if $a_n = b_n R_n$ and $b_n \xrightarrow{p} 0$.

Proof of Theorem 3

Define $LM_n(\boldsymbol{\rho})$ as in (A2), with $\mathcal{S}_n$ and $\boldsymbol{\lambda}\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)$ defined in Assumption 5. By virtue of Lemma 1, we have that $\boldsymbol{\rho}^{LM} \in \Phi\times\Theta$ with probability approaching 1 (w.p.a.1 henceforth), with Θ and Φ satisfying $\boldsymbol{\rho}^* \in \Phi\times\Theta \subseteq \mathbf{P}$, and Φ (resp Θ) contains an open neighborhood of $\boldsymbol{\phi}^*$ (resp $\boldsymbol{\theta}^*$). It is then easy to verify that w.p.a.1

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

where the first equality follows from $\boldsymbol{\rho}^{LM} \in \Phi \times \Theta$ w.p.a.1, and the second one from

$$\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}^*)[n^{-1}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)\boldsymbol{\lambda}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}^{LM}\right)] - \boldsymbol{\lambda}_{\boldsymbol{\phi}}\left(\boldsymbol{\theta}^{LM}\right) \in \{\boldsymbol{\phi} - \boldsymbol{\phi}^* : \boldsymbol{\phi} \in \Phi\} \text{ w.p.a.1.}$$

Similarly, we have that w.p.a.1,

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

As a result,

$$\begin{aligned}
LR &= 2[L_n(\tilde{\boldsymbol{\phi}},\boldsymbol{\theta}) - L_n(\tilde{\boldsymbol{\phi}},\mathbf{0})] \\
&= 2[L_n(\tilde{\boldsymbol{\phi}}_n,\boldsymbol{\theta}) - L_n(\boldsymbol{\phi}^*,\mathbf{0})] - 2[L_n(\tilde{\boldsymbol{\phi}},\mathbf{0}) - L_n(\boldsymbol{\phi}^*,\mathbf{0})] \\
&= \sup_{\boldsymbol{\rho}\in\mathrm{P}}\left\{2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})] - [n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})]\right\} \\
&\quad - \sup_{(\boldsymbol{\phi},\mathbf{0})\in\mathbf{P}}\left\{2[n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi}^*)]'[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})] - [n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]'\mathcal{I}(\boldsymbol{\phi}^*)[n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi},\mathbf{0})]\right\} + o_p(1) \\
&= \sup_{\boldsymbol{\theta}\in\Theta}\left\{2[\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)]'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right. \qquad \text{(A3)} \\
&\quad \left. - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\} + o_p(1),
\end{aligned}$$

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

The last step is to evaluate (A3) at $\tilde{\boldsymbol{\phi}}$ instead of $\boldsymbol{\phi}^*$. Specifically, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}}) &= \frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) + \frac{1}{n}\frac{\partial\mathcal{S}_{\boldsymbol{\theta},n}(\dot{\boldsymbol{\phi}})}{\partial\boldsymbol{\phi}'}\sqrt{n}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*) \\
&= \frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) - \frac{1}{n}\frac{\partial\mathcal{S}_{\boldsymbol{\theta},n}(\dot{\boldsymbol{\phi}})}{\partial\boldsymbol{\phi}'}\left[\frac{1}{n}\frac{\partial\mathcal{S}_{\boldsymbol{\phi},n}(\ddot{\boldsymbol{\phi}})}{\partial\boldsymbol{\phi}'}\right]^{-1}\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*) \\
&= \frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) - \left[\frac{1}{n}\frac{\partial\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*)}{\partial\boldsymbol{\phi}'} + O_p(n^{-\frac{1}{2}})\right]\left[\frac{1}{n}\frac{\partial\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*)}{\partial\boldsymbol{\phi}'} + O_p(n^{-\frac{1}{2}})\right]^{-1}\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*) \\
&= \frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)^{-1}\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*) + O_p(n^{-\frac{1}{2}}) \qquad \text{(A4)}
\end{aligned}$$

The first two equalities follow from the Taylor expansions of $\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})$ and $\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\phi},n}(\tilde{\boldsymbol{\phi}})$ at $\boldsymbol{\phi}^*$, where $\dot{\boldsymbol{\phi}}$ and $\ddot{\boldsymbol{\phi}}$ take values between $\boldsymbol{\phi}^*$ and $\tilde{\boldsymbol{\phi}}$. In turn, the third equation follows from Assumption 5.6, while the last one follows from Assumption 5.5. Moreover, Assumption 5.3 means that $\mathcal{I}(\boldsymbol{\phi})$ is Lipschitz, so that

$$||\mathcal{I}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}(\boldsymbol{\phi}^*)|| = O_p(n^{-\frac{1}{2}}). \qquad \text{(A5)}$$

Combining (A4) and (A5), we get

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\left\{2(\mathcal{S}_{\boldsymbol{\theta},n}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}^*)\mathcal{S}_{\boldsymbol{\phi},n}(\boldsymbol{\phi}^*))'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right. \\
&\qquad \left. - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\phi}^*) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\boldsymbol{\phi}^*)\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\boldsymbol{\phi}^*)]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\} \\
&= \sup_{\boldsymbol{\theta}\in\Theta}\left\{2\mathcal{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})'[\mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}}) - \mathcal{I}_{\boldsymbol{\theta}\boldsymbol{\phi}}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}(\tilde{\boldsymbol{\phi}})\mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})]\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\} + O_p(n^{-\frac{1}{2}}), \quad \text{(A6)}
\end{aligned}$$

where the second equality holds because $\mathbf{0}$ is an interior point of Θ and the maximizer is $o_p(1)$. Together with (A3), (A6) completes the proof of the first part of the theorem.

Next, under Assumption 5.7 and using the same argument, we have that

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

which proofs the second part of the theorem. □

Proof of Theorem 1

As explained above, we make use of Theorem 3 to prove Theorem 1. The first step is to verify Assumption 5. To do so, define $\boldsymbol{\varsigma}_{\boldsymbol{\theta}r,i}(\boldsymbol{\phi}^*) = \mathbf{B}\mathbf{H}_{ri}(\boldsymbol{\phi}^*)$, where

$$\mathbf{H}_{ri}(\boldsymbol{\phi}) = \frac{\partial^r l_i(\boldsymbol{\phi},\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}} - \begin{bmatrix} \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}_r\boldsymbol{\theta}_1}(\boldsymbol{\phi}) \end{bmatrix} \begin{bmatrix} \mathcal{I}_{\boldsymbol{\phi\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\phi}\boldsymbol{\theta}_1}(\boldsymbol{\phi}) \\ \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\phi}}(\boldsymbol{\phi}) & \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\phi}) \end{bmatrix}^{-1} \begin{bmatrix} \partial l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\phi} \\ \partial l_i(\boldsymbol{\phi},\mathbf{0})/\partial\boldsymbol{\theta}_1 \end{bmatrix},$$

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

$$\frac{\partial^r l_i}{\partial\boldsymbol{\theta}_r^{\otimes r}}(\boldsymbol{\phi}^*,\mathbf{0}) = \mathbf{A}_1\frac{\partial l_i}{\partial\boldsymbol{\phi}}(\boldsymbol{\phi}^*,\mathbf{0}) + \mathbf{A}_2\frac{\partial l_i}{\partial\boldsymbol{\theta}_1}(\boldsymbol{\phi}^*,\mathbf{0}) + \mathbf{A}_3\boldsymbol{\varsigma}_{\boldsymbol{\theta}r,i}(\boldsymbol{\phi}^*),$$

where $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ are $r^2\times p$, $r^2\times q_1$ and $r^2\times\dim(\boldsymbol{\varsigma}_{\boldsymbol{\theta}r})$ matrices, respectively. As a consequence, we have that

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

with $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r) = \frac{1}{r!}\boldsymbol{\theta}_r^{\otimes r\prime}\mathbf{A}_1$, $\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r) = \frac{1}{r!}\boldsymbol{\theta}_r^{\otimes r\prime}\mathbf{A}_2$, $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) = \frac{1}{r!}\boldsymbol{\theta}_r^{\otimes r\prime}\mathbf{A}_3$. It is then easy to see that $\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r)$, $\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r)$ and $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)$ are continuous and homogeneous of degree r in $\boldsymbol{\theta}_r$

Next, let $\mathcal{S}_n = (\mathbf{S}'_{\boldsymbol{\phi}n}, \mathbf{S}'_{\boldsymbol{\theta}_1 n}, \mathcal{S}'_{\boldsymbol{\theta}_r n})'$, with

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

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

Further, let

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

denote the asymptotic variance of $n^{-\frac{1}{2}}\mathcal{S}_n(\boldsymbol{\phi})$, which is block diagonal by construction. Let us also define $LM_n(\boldsymbol{\rho})$ as in (A2) with

$$\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) = [\boldsymbol{\phi} - \boldsymbol{\phi}^* + \boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\theta}_r), \boldsymbol{\theta}_1 + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r), \boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)]'.$$

To verify Assumption 5.1 for $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$, the continuity of $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})$ means that we only need to verify that the unique solution to $\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta}) = \mathbf{0}$ is $(\boldsymbol{\phi}^*,\mathbf{0})$ because it is trivial to see that $\boldsymbol{\lambda}(\boldsymbol{\phi}^*,\mathbf{0}) = \mathbf{0}$. First, if $\boldsymbol{\theta}_r = \mathbf{0}$, then it immediately follows that we must have $\boldsymbol{\phi} = \boldsymbol{\phi}^*$ and $\boldsymbol{\theta}_1 = \mathbf{0}$. Consider the case when $\boldsymbol{\theta}_r \neq \mathbf{0}$. By Assumption 3.2, for all $\boldsymbol{\theta}_r \neq \mathbf{0}$, $\boldsymbol{\theta}_r^{\otimes r\prime}\frac{\partial^r l_i(\boldsymbol{\phi}^*,\mathbf{0})}{\partial\boldsymbol{\theta}_r^{\otimes r}}$ is linearly independent of $[\mathbf{s}_{\boldsymbol{\phi},\boldsymbol{i}}(\boldsymbol{\phi}^*),\mathbf{s}_{\boldsymbol{\theta}_1,i}(\boldsymbol{\phi}^*)]'$, which implies that $\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r) \neq \mathbf{0}$ because

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

As for Assumptions 5.2 and 5.3, notice that the covariance kernel of $\mathcal{S}$ is finite by Assumption 2.2, which implies that Assumption 5.2 holds by virtue of the uniform central limit theorem. Thus, $(n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\phi}n}, n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\theta}_1 n})'$ has a full rank asymptotic covariance matrix because of Assumption 3.2, so $n^{-\frac{1}{2}}\mathbf{S}_{\boldsymbol{\theta}_r n}$ does not belong to the linear span of $(n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\phi}n}, n^{-\frac{1}{2}}\mathbf{S}'_{\boldsymbol{\theta}_1 n})'$ by construction. If we combine this result with Assumption 5.2, we have $0 < e_{\min}(\boldsymbol{\phi}^*) < e_{\max}(\boldsymbol{\phi}^*) < \infty$, as desired.

The verification of Assumption 5.4 contains two parts. In the first part, we show that

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

where

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

while in the second part, we show that

$$\sup_{(\boldsymbol{\phi},\boldsymbol{\theta})\in\mathbf{P}:\|(\boldsymbol{\phi},\boldsymbol{\theta})-(\boldsymbol{\phi}^*,\mathbf{0})\|\leq\gamma_n}\frac{h(\boldsymbol{\phi},\boldsymbol{\theta})}{1+n\|\boldsymbol{\lambda}(\boldsymbol{\phi},\boldsymbol{\theta})\|^2} = O(1). \tag{A7}$$

Combining the two parts, we get

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
B_9 &= \frac{1}{r!} n(\boldsymbol{\phi} - \boldsymbol{\phi}^*)' \left(\frac{1}{n} \frac{\partial^{1+r} L_n}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes r \prime}} - E\left[\frac{\partial^{1+r} l_i}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes r \prime}}\right]\right) \boldsymbol{\theta}_r^{\otimes r}, \\
B_{10} &= \sum_{j=r+1}^{2r} \frac{1}{j!} n(\boldsymbol{\phi} - \boldsymbol{\phi}^*)' \left\{\frac{1}{n} \frac{\partial^{1+j} L_n}{\partial \boldsymbol{\phi} \partial \boldsymbol{\theta}_r^{\otimes j \prime}}\right\} \boldsymbol{\theta}_r^{\otimes j} + \sum_{j=1}^{r-1} \frac{1}{j!} \sqrt{n} (\boldsymbol{\phi} - \boldsymbol{\phi}^*)' \left\{\frac{1}{n} \frac{\partial^{1+j} L_n}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_r^{\otimes j \prime}}\right\} \boldsymbol{\theta}_r^{\otimes j},
\end{aligned}$$

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

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

$$\begin{aligned}
B_{14} &= \sum_{j=r+1}^{2r} \frac{1}{j!} n \boldsymbol{\theta}_1' \left\{\frac{1}{n} \frac{\partial^{1+j} L_n}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_r^{\otimes j \prime}}\right\} \boldsymbol{\theta}_r^{\otimes j}, \\
B_{15} &= \sum_{j_1+j_2=3, j_1 \geq 2, j_2 \geq 1}^{8} \frac{1}{j_1! j_2!} n \boldsymbol{\theta}_1^{\otimes j_1 \prime} \left\{\frac{1}{n} \frac{\partial^{j_1+j_2} L_n}{\partial \boldsymbol{\theta}_1^{\otimes j_1} \partial \boldsymbol{\theta}_r^{\otimes j_2 \prime}}\right\} \boldsymbol{\theta}_r^{\otimes j_2},
\end{aligned}$$

$$B_{16} = \sum_{\iota_p' \mathbf{j}_1 + \iota_{q_1}' \mathbf{j}_2 + \iota_{q_1}' \mathbf{j}_3 = 3, \mathbf{j}_{1,2,3} > 0}^{8} \left\{ \frac{1}{n} L_n^{[\mathbf{j}_1, \mathbf{j}_2, \mathbf{j}_3]} \right\} n(\boldsymbol{\phi} - \boldsymbol{\phi}^*)^{\mathbf{j}_1} \boldsymbol{\theta}_1^{\mathbf{j}_2} \boldsymbol{\theta}_r^{\mathbf{j}_3}, , \text{ and}$$

$$B_{17} = \sum_{\iota_p' \mathbf{j}_1 + \iota_{q_1}' \mathbf{j}_2 + \iota_{q_1}' \mathbf{j}_3 = 8} \left(\frac{1}{n} L_n^{[\mathbf{j}_1, \mathbf{j}_2, \mathbf{j}_3]} (\bar{\boldsymbol{\rho}}) - \frac{1}{n} L_n^{[\mathbf{j}_1, \mathbf{j}_2, \mathbf{j}_3]} \right) n(\boldsymbol{\phi} - \boldsymbol{\phi}^*)^{\mathbf{j}_1} \boldsymbol{\theta}_1^{\mathbf{j}_2} \boldsymbol{\theta}_r^{\mathbf{j}_3},$$

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

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

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

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

$$\frac{|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})|}{h_n(\boldsymbol{\phi}, \boldsymbol{\theta})} = O_p(n^{-\frac{1}{2r}}) \tag{A8}$$

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

To verify the second part, let

$$\pi_{\boldsymbol{\phi}} = \max_{\|\boldsymbol{v}\|=1} \|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v})\|, \;\; \pi_{\boldsymbol{\theta}_1} = \max_{\|\boldsymbol{v}\|=1} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{v})\| \;\; \text{and} \;\; \pi_r = \min_{\|\boldsymbol{v}\|=1} \|\boldsymbol{\lambda}_{\boldsymbol{\theta}_r}(\boldsymbol{v})\| > 0, \tag{A9}$$

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

$$\max_{(\boldsymbol{\phi}, \boldsymbol{\theta}) \in \mathbf{P}: \|(\boldsymbol{\phi}, \boldsymbol{\theta}) - (\boldsymbol{\phi}^*, \mathbf{0})\| \leq \gamma_n} \frac{h_n^{\pi}(\boldsymbol{\phi}, \boldsymbol{\theta})}{1 + \|n^{\frac{1}{2}} \boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\|^2} = O(1) \tag{A10}$$

to verify (A7), with

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

where the coefficients $\pi_1 = (2\pi_{\boldsymbol{\phi}} + 1)^{-1}$ and $\pi_2 = (2\pi_{\boldsymbol{\theta}_1} + 1)^{-1}$, which are positive, are only used to simplify the expressions.

For n large enough, we have that

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

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

$$\sup_{(\boldsymbol{\phi}, \boldsymbol{\theta}) \in \mathbf{P}: \|(\boldsymbol{\phi}, \boldsymbol{\theta}) - (\boldsymbol{\phi}^*, \mathbf{0})\| \leq \gamma_n} \frac{h_n^{\pi}(\boldsymbol{\phi}, \boldsymbol{\theta})}{1 + n \|\boldsymbol{\lambda}(\boldsymbol{\phi}, \boldsymbol{\theta})\|^2} = \frac{h_n^{\pi}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)}{1 + n \|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\|^2} \tag{A11}$$

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

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

where the first equality follows directly from (A11) and the second one from the properties of lim sup. Consequently, it is easy to see that if $h_{w_n}^{\pi}(\boldsymbol{\phi}_{w_n},\boldsymbol{\theta}_{w_n}) = O(1)$, then (A7) holds trivially. In turn, if $h_{w_n}^{\pi}(\boldsymbol{\phi}_{w_n},\boldsymbol{\theta}_{w_n}) \neq O(1)$, then we can find a further subsequence $\{u_n\}$ of $\{w_n\}$ such that at least one of the following conditions holds:

$$h_{u_n}^{\pi}(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n}) = u_n\|\boldsymbol{\theta}_{r,u_n}\|^{2r} \to \infty, \tag{A12}$$

$$h_{u_n}^{\pi}(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n}) = \pi_1^2 u_n||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*||^2 \to \infty, \quad \text{or} \tag{A13}$$

$$h_{u_n}^{\pi}(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n}) = \pi_2^2 u_n\|\boldsymbol{\theta}_{1,u_n}\|^2 \to \infty. \tag{A14}$$

Let $\boldsymbol{\theta}_{r,n} = \eta_n\boldsymbol{v}_n$ with $\|\boldsymbol{v}_n\| = 1$ and η_n a scalar. If (A12) holds, then

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

where the first inequality follows from

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

the second one from the definition of $\boldsymbol{\lambda}_{\boldsymbol{\theta}^r}$, and the last inequality from the characterization of π_r in (A9).

If (A13) holds, then

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

so that

$$\pi_1^2||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*||^2 \geq \|\boldsymbol{\theta}_{ru_n}\|^{2r} = \eta_{u_n}^{2r} \Rightarrow \pi_1||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*|| \geq \eta_{u_n}^r, \tag{A15}$$

which in turn yields

$$\|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*+\eta_n^r\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v}_n)\| \geq |\|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*\| - \eta_n^r\|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v}_n)\|| = \|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*\|\left|1-\frac{\eta_n^r}{\|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*\|}\|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v}_n)\|\right|$$
$$\geq \|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*\||1-\pi_1\pi_{\boldsymbol{\phi}}| > \frac{1}{2}\|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*\|, \tag{A16}$$

where the first line follows from the triangle inequality and the second one from $\eta_n^r/\|\boldsymbol{\phi}_n-\boldsymbol{\phi}^*\| \leq \pi_1$ in view of (A15) and $\|\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v}_n)\| \leq \pi_{\boldsymbol{\phi}}$ because of (A9). Then, we have that

$$\frac{h_{u_n}^{\pi}(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})}{1+u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n},\boldsymbol{\theta}_{u_n})||^2} < \frac{\pi_1^2||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*||^2}{||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*+\eta_{u_n}^r\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{v}_{u_n})||^2} \leq \frac{\pi_1^2||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*||^2}{\frac{1}{4}||\boldsymbol{\phi}_{u_n}-\boldsymbol{\phi}^*||^2} = 4\pi_1^2,$$

where the first inequality follows from $u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})||^2 > u_n||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^* + \eta^r_{u_n}\boldsymbol{\lambda}_{\boldsymbol{\phi}}(\boldsymbol{\upsilon}_{u_n})||^2$ and $h^\pi_{u_n}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n}) \leq u_n\pi_1^2||\boldsymbol{\phi}_{u_n} - \boldsymbol{\phi}^*||^2$, while the second one from (A16).

Similarly, if (A14) holds, then we have that

$$\pi_2^2 \|\boldsymbol{\theta}_{1u_n}\|^2 \geq \|\boldsymbol{\theta}_{ru_n}\|^{2r} = \eta^{2r}_{u_n} \text{ implies } \pi_2 \|\boldsymbol{\theta}_{1u_n}\| \geq \eta^r_{u_n}, \tag{A17}$$

whence

$$\begin{aligned} \|\boldsymbol{\theta}_{1u_n} + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_{ru_n})\| &\geq \left|\|\boldsymbol{\theta}_{1u_n}\| - \eta^r_{u_n}\|\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\upsilon}_{u_n})\|\right| \\ &= \|\boldsymbol{\theta}_{1u_n}\| \left|1 - \frac{\eta^r_{u_n}}{\|\boldsymbol{\theta}_{1u_n}\|}\|\boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\upsilon}_{u_n})\|\right| \\ &\geq \|\boldsymbol{\theta}_{1u_n}\| \left|1 - \pi_2\pi_{\boldsymbol{\theta}_1}\right| > \frac{1}{2}\|\boldsymbol{\theta}_{1u_n}\|, \end{aligned} \tag{A18}$$

where the first two inequalities are straightforward, and the third one follows from (A9) and (A17). In addition, we can show that

$$\frac{h^\pi_{u_n}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})}{1 + u_n||\boldsymbol{\lambda}(\boldsymbol{\phi}_{u_n}, \boldsymbol{\theta}_{u_n})||^2} < \frac{\pi_2^2\|\boldsymbol{\theta}_{1u_n}\|^2}{\|\boldsymbol{\theta}_{1u_n} + \boldsymbol{\lambda}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_{ru_n})\|^2} \leq \frac{\pi_2^2\|\boldsymbol{\theta}_{1u_n}\|^2}{\frac{1}{4}\|\boldsymbol{\theta}_{1u_n}\|^2} = 4\pi_2^2, \tag{A19}$$

where the first inequality follows from (A14) and the second one from (A18).

The previous argument also implies that if $h_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) \to \infty$, then $h^\pi_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) \to \infty$ and $n\|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\| \to \infty$, which in turn implies that

$$n^{\frac{1}{2}}\|\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)\| = O(1) \Rightarrow h_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1). \tag{A20}$$

Trivially, Assumption 5.5 follows from Corollary 1 in Rotnitzki et al (2000).

Regarding Assumption 5.6, if $n^{\frac{1}{2}}\boldsymbol{\lambda}(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$, then we have $h_n(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = O(1)$ in view of (A20), which in turn implies $|R_n(\boldsymbol{\phi}, \boldsymbol{\theta})| = O_p(n^{-\frac{1}{2r}})$ thanks to (A8).

Consequently, since Theorem 3 implies that

$$LR = \sup_{\boldsymbol{\theta}}\left\{2\mathbf{S}_{\boldsymbol{\theta},n}(\tilde{\boldsymbol{\phi}})'\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - n\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})' V_{\boldsymbol{\theta}\boldsymbol{\theta}}(\tilde{\boldsymbol{\phi}})\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\} + O_p(n^{-\frac{1}{2r}}),$$

where

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

and $\boldsymbol{\lambda}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = [\boldsymbol{\theta}_1' + \boldsymbol{\lambda}'_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_r), \boldsymbol{\lambda}'_{\boldsymbol{\theta}_r}(\boldsymbol{\theta}_r)]'$, after rearranging terms we obtain

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

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

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

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

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

After noticing that

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

and

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

we can finally verify that Theorem 1 holds. □

Proof of Theorem 2

In this proof, we use the notation $\mathbf{j}_q < C$ (resp. $\mathbf{j}_q > C$, $\mathbf{j}_q \leq C$, $\mathbf{j}_q \geq C$) to indicate that there exists some $\mathbf{j}'_q \in C$ such that $\mathbf{j}_q < \mathbf{j}'_q$ (resp. $\mathbf{j}_q > \mathbf{j}'_q$, $\mathbf{j}_q \leq \mathbf{j}'_q$, $\mathbf{j}_q \geq \mathbf{j}'_q$). To simplify the notation, we only give the proof for $\boldsymbol{\rho} = (\theta_1, \theta_2)$ so that the set C is defined as $C = \{(i_1, j_1), ..., (i_K, j_K)\}$. However, we could easily generalize it to models in which $\dim(\boldsymbol{\theta}) \geq 2$ and there additional parameters $\boldsymbol{\phi}$. Define $LM_n(\boldsymbol{\theta})$ as in (A2) with $\mathcal{I}$ defined in Theorem 2, and

$$\boldsymbol{\lambda}'(\boldsymbol{\theta}) = \begin{pmatrix} \theta_1^{i_1}\theta_2^{j_1} & \theta_1^{i_2}\theta_2^{j_2} & \ldots & \theta_1^{i_K}\theta_2^{j_K} \end{pmatrix}, \quad \mathcal{S}'_n = \begin{pmatrix} L_n^{[i_1,j_1]} & L_n^{[i_2,j_2]} & \ldots & L_n^{[i_K,j_K]} \end{pmatrix}.$$

We first verify Assumptions 5.4 and 5.7. Assumption 5.5 follows from Lemma 6, while the other parts of Assumption 5 hold trivially.

Considering a $2r^{th}$-order Taylor expansion of $L_n(\boldsymbol{\theta}_n)$ around $\mathbf{0}$ in terms of $(\theta_{1n}, \theta_{2n})$, omitting $\mathbf{0}$ as an argument and the subscript n from $\boldsymbol{\theta}_n$ for simplification, then we have $L_n^{[i,j]} = 0$ for terms (i, j) such that $(i, j) < C$ because of the definition of C. Further, Lemma 6 implies that under Assumption 5, we have $l^{[i,j]} = f^{[i,j]}/f$ evaluated at the null for $(i,j) \in C$, and hence $E(l^{[i,j]}) = 0$ for $(i,j) \in C$. Note that in the Taylor expansion, the corresponding term is

$$\left(n^{-\frac{1}{2}} \frac{\partial^{i+j} L_n}{\partial \theta_1^i \partial \theta_2^j}\right) n^{\frac{1}{2}} \theta_1^i \theta_2^j,$$

which belongs to the first summand of $LM_n(\boldsymbol{\theta}_n)$.

For those pairs (i,j) such that $(i,j) > C$, if $l^{[i,j]} \neq 0$ and $E(l^{[i,j]}) = 0$, then the corresponding term in the Taylor expansion is again

$$\left(n^{-\frac{1}{2}} \frac{\partial^{i+j} L_n}{\partial \theta_1^i \partial \theta_2^j}\right) n^{\frac{1}{2}} \theta_1^i \theta_2^j. \tag{A21}$$

Since $(i,j) > C$, we can find $(i',j') \in C$ and $(i',j') < (i,j)$ such that the associated term

$$\left(n^{-\frac{1}{2}} \frac{\partial^{i'+j'} L_n}{\partial \theta_1^{i'} \partial \theta_2^{j'}}\right) n^{\frac{1}{2}} \theta_1^{i'} \theta_2^{j'}$$

dominates the (i,j) term because $\theta_1, \theta_2 = o_p(1)$, which means that (A21) is $o_p(1 + n||\boldsymbol{\lambda}(\boldsymbol{\theta})||^2)$.

For those pairs (i,j) such that $(i,j) > C$, if $E(l^{[i,j]}) \neq 0$, then Lemma 6 implies that

$$\begin{aligned} E[l^{[i,j]}] &= E\left[\sum_{1\le h \le i+j} (-1)^{h+1} \sum_{\substack{s=1:h \\ p_s[(i,j),h]}} \prod_{a=1}^{s} \frac{1}{m_a!}\left(\frac{f^{[\boldsymbol{k}_a]}}{f}\right)^{m_a}\right] \\ &= E\left[-\sum_{\substack{s=1:2 \\ p_s[(i,j),2]}} \prod_{a=1}^{s} \frac{1}{m_a!}\left(\frac{f^{[\boldsymbol{k}_a]}}{f}\right)^{m_a} + \sum_{2<h \le i+j} (-1)^{h+1} \sum_{\substack{s=1:h \\ p_s[(i,j),h]}} \prod_{a=1}^{s} \frac{1}{m_a!}\left(\frac{f^{[\boldsymbol{k}_a]}}{f}\right)^{m_a}\right], \end{aligned}$$

where $p_s[(i,j),h]$ is defined in (B19). The first equality is a direct consequence of Lemma 6, while the second follows from: i) splitting $\{1 \le h \le i+j\}$ into $\{1 \le h \le 2\}$ and $\{2 < h \le i+j\}$, and ii) when $h = 1$,

$$(-1)^{h+1} \sum_{\substack{s=1:h \\ p_s[(i,j),h]}} E\left[\prod_{a=1}^{s} \frac{1}{m_a!}\left(\frac{f^{[\boldsymbol{k}_a]}}{f}\right)^{m_a}\right] = \sum_{\substack{s=1:1 \\ p_s[(i,j),1]}} E\left(\frac{f^{[i,j]}}{f}\right) = 0.$$

In this context, the law of large numbers implies that the $(i,j)^{th}$ term in the Taylor expansion will be given by

$$\begin{aligned} (n^{-1}L_n^{[i,j]}) n\theta_1^i\theta_2^j = &- \sum_{\substack{s=1:2 \\ p_s[(i,j),2]}} E\left[\prod_{a=1}^{s} \frac{1}{m_a!}\left(\frac{f^{[\boldsymbol{k}_a]}}{f}\right)^{m_a}\right] n\theta_1^i\theta_2^j \\ &+ \sum_{2<h\le i+j} (-1)^{h+1} \sum_{\substack{s=1:h \\ p_s[(i,j),h]}} E\left[\prod_{a=1}^{s} \frac{1}{m_a!}\left(\frac{f^{[\boldsymbol{k}_a]}}{f}\right)^{m_a}\right] n\theta_1^i\theta_2^j \\ &+ O_p(n^{-\frac{1}{2}}(1 + n||\boldsymbol{\lambda}(\boldsymbol{\theta})||^2)). \end{aligned}$$

If $h = 2$ and $s = 1$ then $m_1 = 2$. More generally, if i or j are odd, then $p_1[(i,j),2] = \emptyset$, while if i,j are both even instead, then $p_1[(i,j),2] = \{[2;(\frac{i}{2},\frac{j}{2})]\}$ (see (B19)). Consequently, when $(\frac{i}{2},\frac{j}{2}) \in C$, then the corresponding term is

$$-\frac{1}{2}E\left[\left(\frac{f^{[\frac{i}{2},\frac{j}{2}]}}{f}\right)^2\right]n\theta_1^i\theta_2^j = -\frac{1}{2}Var(l^{[\frac{i}{2},\frac{j}{2}]})n\theta_1^i\theta_2^j, \tag{A22}$$

which belongs to the second summand of $LM_n(\boldsymbol{\theta})$. In turn, if $(\frac{i}{2},\frac{j}{2}) \notin C$, then either (i) $(\frac{i}{2},\frac{j}{2}) > C$, which means that $\exists(i',j') \in C$ such that $(i',j') < (\frac{i}{2},\frac{j}{2})$, in which case the LHS of (A22) is dominated by $-\frac{1}{2}Var(l^{[i',j']})n\theta_1^{2i'}\theta_2^{2j'}$; or (ii) $(\frac{i}{2},\frac{j}{2}) < C$ and, therefore, the LHS of (A22) must be equal to zero because $l^{[\frac{i}{2},\frac{j}{2}]} = 0$.

Consider next the case in which $h = 2$, $s = 2$, $m_1 = m_2 = 1$, and $(i,j) = \boldsymbol{k}_1 + \boldsymbol{k}_2$. If $\boldsymbol{k}_1, \boldsymbol{k}_2 \in C$, then the corresponding term is

$$-E\left(\frac{f^{[\boldsymbol{k}_1]}}{f}\frac{f^{[\boldsymbol{k}_2]}}{f}\right)n\theta_1^i\theta_2^j = -Cov(l^{[\boldsymbol{k}_1]}, l^{[\boldsymbol{k}_2]})n\theta_1^i\theta_2^j, \tag{A23}$$

which also belongs to the second summand of $LM_n(\boldsymbol{\theta})$. If either $\boldsymbol{k}_1 < C$ or $\boldsymbol{k}_2 < C$, then the LHS of (A23) is equal to zero. Next, we look at the cases in which $\boldsymbol{k}_1 \geq C$ and $\boldsymbol{k}_2 > C$ or $\boldsymbol{k}_2 \geq C$ and $\boldsymbol{k}_1 > C$. Specifically, if we can find a pair $(i',j') \in C$ such that $\boldsymbol{k}_1 \geq (i',j')$ and another pair $(i'',j'') \in C$ such that $\boldsymbol{k}_2 \geq (i'',j'')$ so that $\boldsymbol{k}_2 > (i'',j'')$ if $\boldsymbol{k}_1 = (i',j')$ and vice versa, then the LHS of (A23) is dominated by the largest of $n\theta_1^{2i'}\theta_2^{2j'}$ and $n\theta_1^{2i''}\theta_2^{2j''}$. Consequently,

$$\begin{aligned}\left|E\left(\frac{f^{[\boldsymbol{k}_1]}}{f}\frac{f^{[\boldsymbol{k}_2]}}{f}\right)n\theta_1^i\theta_2^j\right| &= o_p(1 + n\theta_1^{i'+i''}\theta_2^{j'+j''}) = o_p(1 + n(\theta_1^{2i'}\theta_2^{2j'} + \theta_1^{2i''}\theta_2^{2j''})) \\ &= o_p(1 + n\left\|\boldsymbol{\lambda}(\boldsymbol{\theta})\right\|^2),\end{aligned} \tag{A24}$$

where $\boldsymbol{k}' = (i',j') \leq \boldsymbol{k}_1$ and $\boldsymbol{k}'' = (i'',j'') \leq \boldsymbol{k}_2$.

Finally, consider $h \geq 3$. In this context, either there exists a j such that $\boldsymbol{k}_j < C$, in which case $E\left[\prod_{j=1}^s \frac{1}{m_j!}\left(\frac{f^{[\boldsymbol{k}_j]}}{f}\right)^{m_j}\right] = 0$, or $\boldsymbol{k}_j \geq C$ for all j, in which case $E\left[\prod_{j=1}^s \frac{1}{m_j!}\left(\frac{f^{[\boldsymbol{k}_j]}}{f}\right)^{m_j}\right]n\theta_1^i\theta_2^j$ will be dominated by the second summand of $LM_n(\boldsymbol{\theta})$, as in (A24).

The remainder terms, which correspond to all those indices that satisfy $(i+j) = 2r$, are such that

$$|\delta^{[i,j]}| = |n^{-1}(L_n^{[i,j]}(\bar{\boldsymbol{\theta}}) - L_n^{[i,j]})||n\theta_1^i\theta_2^j| = o_p(\max\{1, n\theta_1^i\theta_2^j\}) = o_p(1 + n||\boldsymbol{\lambda}(\boldsymbol{\theta})||^2), \tag{A25}$$

because $|n^{-1}(L_n^{[i,j]}(\bar{\boldsymbol{\theta}}) - L_n^{[i,j]})| \leq ||\bar{\boldsymbol{\theta}}||n^{-1}\sum_i g(\mathbf{y}_i) = o_p(1)$, since $||\bar{\boldsymbol{\theta}}|| = o_p(1)$ and $|n^{-1}\sum_i g(\mathbf{y}_i)| = O_p(1)$ by Assumption 2.3. But given that (A25) contains the last terms in the $2r^{th}$-order Taylor expansion of of $L(\boldsymbol{\theta}_n)$ around $\mathbf{0}$, Assumption 5.4 holds.

Let us now turn to verifying Assumption 5.7, for which we further assume that $n^{1/2}\left\|\boldsymbol{\lambda}(\boldsymbol{\theta})\right\| = O_p(1)$. We then have $\theta_1 = O_p(n^{-1/2r_1})$ and $\theta_2 = O_p(n^{-1/2r_2})$ because $(r_1, 0) \in C$ and $(0, r_2) \in C$, which has important implications for the different terms of the expansion. First, notice that we do not make any approximation for the leading terms with $(i+j) \leq 2r$ in the first summand of LM_n. In addition, we can write those (i,j)-th terms that are not included in the first two summands of LM_n as $O_p(1)\theta_1^{k_1}\theta_2^{k_2}$ with $k_1 + k_2 \geq 1$, which implies that they are $O_p(n^{-1/r})$. As for the rest of the leading terms, namely those whose (i,j)-th term belongs to the second summand of LM_n, we can approximate $\frac{1}{n}L_n^{[i,j]}$ by its expectation, where the convergence rate is $O_p(n^{-1/2})$ as shown by Rotnitzky at al (2000). Finally, we can easily see that Assumption 2.6 implies that $\delta^{[i,j]} = O_p(n^{-1/r})$ for the remainder terms. Therefore, Assumption 5.7 holds, as desired. □

Proof of Theorem 4

Treating the strictly exogenous covariates X_i as fixed in repeated samples, we can modify the testing procedure in Section 2.3 as follows:

1. For a fixed $\{X\}_{i=1}^n$, simulate a sample $\{\mathbf{y}_i^{(s)}\}_{i=1}^n$ from $f[\mathbf{y}_i|X_i;(\tilde{\boldsymbol{\phi}},0)]$
2. Compute $GET^{(s)}$ using the simulated sample $\{X_i, \mathbf{y}_i^{(s)}\}_{i=1}^n$

The proof contains three steps.

Step 1. Consider a sequence $\boldsymbol{\phi}_n \in \mathcal{N}_{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}_n \to \boldsymbol{\phi}^*$. Under the sequence of DGPs $(\boldsymbol{\phi}_n, 0)$ a first order Taylor expansion of $\mathcal{S}_{\phi,n}(\tilde{\boldsymbol{\phi}})$ delivers

$$o_p(1) = \frac{1}{\sqrt{n}}\mathcal{S}_{\phi,n}(\tilde{\boldsymbol{\phi}}) = \frac{1}{\sqrt{n}}\mathcal{S}_{\phi,n}(\boldsymbol{\phi}^*) + \frac{1}{n}\frac{\partial \mathcal{S}_{\phi,n}(\dot{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\sqrt{n}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*)$$

$$\Rightarrow \sqrt{n}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}^*) = -\left[\frac{1}{n}\frac{\partial \mathcal{S}_{\phi,n}(\dot{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\right]^{-1}\frac{1}{\sqrt{n}}\mathcal{S}_{\phi,n}(\boldsymbol{\phi}^*) + o_p(1)$$

Then,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\mathcal{S}_{\theta,n}(\tilde{\boldsymbol{\phi}}) &= \frac{1}{\sqrt{n}}\mathcal{S}_{\theta,n}(\boldsymbol{\phi}_n) + \frac{1}{n}\frac{\partial \mathcal{S}_{\theta,n}(\ddot{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\sqrt{n}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_n) \\
&= \frac{1}{\sqrt{n}}\mathcal{S}_{\theta,n}(\boldsymbol{\phi}_n) - \left[\frac{1}{n}\frac{\partial \mathcal{S}_{\theta,n}(\ddot{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\right]\left[\frac{1}{n}\frac{\partial \mathcal{S}_{\phi,n}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right]^{-1}\frac{1}{\sqrt{n}}\mathcal{S}_{\boldsymbol{\phi},\boldsymbol{n}}(\boldsymbol{\phi}_n^*) \\
&= \frac{1}{\sqrt{n}}\mathcal{S}_{\theta,n}(\boldsymbol{\phi}_n) - \left[\frac{1}{n}\frac{\partial \mathcal{S}_{\theta,n}(\boldsymbol{\phi}_n)}{\partial \boldsymbol{\phi}'} + o_p(1)\right]\left[\frac{1}{n}\frac{\partial \mathcal{S}_{\phi,n}(\boldsymbol{\phi}_n)}{\partial \phi'} + o_p(1)\right]^{-1}\frac{1}{\sqrt{n}}\mathcal{S}_{\phi}(\boldsymbol{\phi}_n) \\
&\xrightarrow{d} \mathcal{S}^{\perp}(\boldsymbol{\phi}^*) \qquad \text{(A26)}
\end{aligned}$$

where
$$\mathcal{S}^{\perp}(\boldsymbol{\phi}^*) = \mathcal{S}_{\theta}(\boldsymbol{\phi}^*) - \mathcal{I}_{\theta\phi}(\boldsymbol{\phi}^*)\mathcal{I}_{\phi\phi}(\boldsymbol{\phi}^*)^{-1}\mathcal{S}_{\phi}(\boldsymbol{\phi}^*),$$

the third line follows from Assumption 6.2 and the fact that $\sqrt{n}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_n) = O_p(1)$, and the last line follows from the continuous mapping theorem and Assumption 6.1.

Step 2. Let BL_1 denotes the set of Lipschitz functions that are bounded by 1 in absolute value and have Lipschitz constant bounded by 1, and let

$$\mathcal{N}_{\phi,n} = \left\{ \phi \in \mathcal{N}_{\phi} : |\phi - \phi^*| \leq \frac{\ln n}{\sqrt{n}} \right\}.$$

Expression (A26) implies that

$$\lim_n \sup_{\phi \in \mathcal{N}_{\phi,n}} \sup_{h \in BL_1} \left| E\left[h\left(\frac{1}{\sqrt{n}} \mathcal{S}_n(\phi, 0) \right); (\phi, 0) \right] - E\left[h\left(\mathcal{S}(\phi^*) \right) \right] \right| = 0. \tag{A27}$$

Then, noticing that

$$\begin{aligned}
&\Pr\left(\lim_n \sup_{h \in BL_1} \left| E^{(s)}\left[h\left(\frac{1}{\sqrt{n}} \mathcal{S}_n^{(s)}(\tilde{\phi}^{(s)}) \right) \middle| \{X_i, \mathbf{y}_i\}_{i=1}^n \right] - E\{h[\mathcal{S}^{\perp}(\phi^*)]\} \right| = 0 \right) \\
&= \Pr\left(\lim_n \sup_{h \in BL_1} \left| E^{(s)}\left[h\left(\frac{1}{\sqrt{n}} \mathcal{S}_n^{(s)}(\tilde{\phi}^{(s)}) \right) \middle| (\tilde{\phi}, \mathbf{0}) \right] - E\{h[\mathcal{S}^{\perp}(\phi^*)]\} \right| = 0 \right) \\
&\geq \Pr\left(\lim_n \sup_{\phi \in \mathcal{N}_{\phi,n}} \sup_{h \in BL_1} \left| E^{(s)}\left[h\left(\frac{1}{\sqrt{n}} \mathcal{S}_n^{(s)}(\tilde{\phi}^{(s)}) \right); (\phi, \mathbf{0}) \right] - E\{h[\mathcal{S}^{\perp}(\phi^*)]\} \right| + \mathbf{1}[\tilde{\phi} \notin \mathcal{N}_{\phi,n}] = 0 \right) \\
&= \Pr\left(\lim_n \sup_{\phi \in \mathcal{N}_{\phi,n}} \sup_{h \in BL_1} \left| E\left[h\left(\frac{1}{\sqrt{n}} \mathcal{S}_n(\tilde{\phi}) \right); (\phi, \mathbf{0}) \right] - E\{h[\mathcal{S}^{\perp}(\phi^*)]\} \right| = 0 \right) = 1,
\end{aligned}$$

where the first equality follows from the fact that $\{\mathbf{y}_i^{(s)}\}$ depends on $\{X_i, \mathbf{y}_i\}$ only through $\tilde{\phi}$, the middle inequality is straightforward, the second equality follows from $\limsup \Pr(\tilde{\phi} \notin \mathcal{N}_{\phi,n}) = 0$, and the last one from (A27). Therefore, we have

$$\frac{1}{\sqrt{n}} \mathcal{S}_n^{(s)}(\tilde{\phi}^{(s)}) \bigg| \{X_i, \mathbf{y}_i\}_{i=1}^n \xrightarrow{d} \mathcal{S}^{\perp}(\phi^*). \tag{A28}$$

In addition, given that $\tilde{\phi}^{(s)} | \{X_i, \mathbf{y}_i\}_{i=1}^n \xrightarrow{a.s.} \phi^*$ and $\mathcal{I}(\cdot)$ is continuous, it follows that

$$\mathcal{I}(\tilde{\phi}^{(s)}) \Big| \{X_i, \mathbf{y}_i\}_{i=1}^n \xrightarrow{p} \mathcal{I}(\phi^*). \tag{A29}$$

Step 3. By (A28), (A29), and a proof similar to Theorem 3, we have (6).

Finally, the proof of Theorem 4.2 follows from van der Vaart (1996) Lemma 23.3, by changing $(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ to GET_n and $(\hat{\theta}_n^* - \hat{\theta})/\hat{\sigma}_n^*$ to $GET_n^{(s)}$. □

For the sake of completeness, the following primitive assumption implies Assumption 6.

Assumption 9 *(Bootstrap validity for Theorem 1)*

1. *There is an open set $\mathcal{N}_{\phi}$ such that $\phi^* \in \mathcal{N}_{\phi}$ and*

$$\sup_{\phi \in \mathcal{N}_{\phi}} E\left[\left\| \left(\frac{\partial^2 \ell(\phi, \mathbf{0})}{\partial \phi \partial \phi'}, \frac{\partial^2 \ell(\phi, \mathbf{0})}{\partial \phi \partial \theta_1'}, \frac{\partial^{r+1} \ell(\phi, \mathbf{0})}{\partial \phi \partial \theta_r^{\otimes r \prime}} \right) \right\|^{1+\varepsilon}; (\phi, \mathbf{0}) \right] < \infty$$

$$\sup_{\phi \in \mathcal{N}_{\phi}} E\left[\|(s_{\phi}(\phi), s_{\theta_1}(\phi), s_{\theta_r}(\phi))\|^{2+\varepsilon}; (\phi, \mathbf{0}) \right] < \infty$$

for some $\varepsilon > 0$.

2. *When $\boldsymbol{\iota}_{p+q}' \mathbf{j} = 2r$ there is some function $g(\mathbf{y})$ satisfying $\sup_{\phi \in \mathcal{N}_{\phi}} E\left[g^2(\mathbf{y}); (\phi, 0) \right] < \infty$ such that with probability 1, $\left| L_n^{[j]}(\boldsymbol{\rho}) - L_n^{[j]}(\boldsymbol{\rho}^{\dagger}) \right| \leq \|\boldsymbol{\rho} - \boldsymbol{\rho}^{\dagger}\| \sum_i g(\mathbf{y}_i)$ for all $\boldsymbol{\rho}$ and $\boldsymbol{\rho}^{\dagger}$ in $\mathcal{N}$.*

Proof of Theorem 5

By Le Cam's first Lemma (see Lemma 6.4 of van der Vaart (1998)), contiguity holds if under P_0, $dP_{\boldsymbol{\phi},\boldsymbol{\theta}_n}/dP_0 \xrightarrow{d} U$ with $E(U)=1$. Letting $L_n(\boldsymbol{\phi}^*,\boldsymbol{\theta}_n)$ denote the joint log-likelihood of the observations, Assumption 5 allows us to write

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

Therefore, under H_0,

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

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

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

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

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

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

Finally, given Assumption 7, we can then prove that under $P_{\boldsymbol{\theta}_n}$,

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

where

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

as desired. □

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

	Null hypothesis			Alternative hypotheses					
				H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel A: $n = 400$									
LR_{naive}	2.6	11.6	20.8	17.6	39.7	54.2	17.3	39.6	53.9
LR	0.9	4.9	10.4	9.1	25.2	37.1	9.1	25.2	36.9
GET	1.0	5.0	10.2	8.5	23.2	35.1	8.6	23.9	35.9
GMM	1.0	5.1	10.1	7.6	22.0	32.5	7.8	22.4	33.3
GMM_{asy}	1.1	4.9	9.8	8.1	21.3	32.0	8.2	21.9	32.7
Panel B: $n = 1600$									
LR_{naive}	2.0	9.2	17.9	77.9	91.4	95.3	78.1	91.3	95.6
LR	0.9	4.8	9.6	68.0	86.6	91.6	68.9	86.4	91.7
GET	0.8	5.1	9.7	62.2	82.7	88.8	62.7	83.1	89.5
GMM	1.0	5.2	10.0	57.9	79.3	87.5	58.5	79.2	87.6
GMM_{asy}	1.0	5.2	10.0	57.6	79.5	87.5	58.2	79.4	87.5

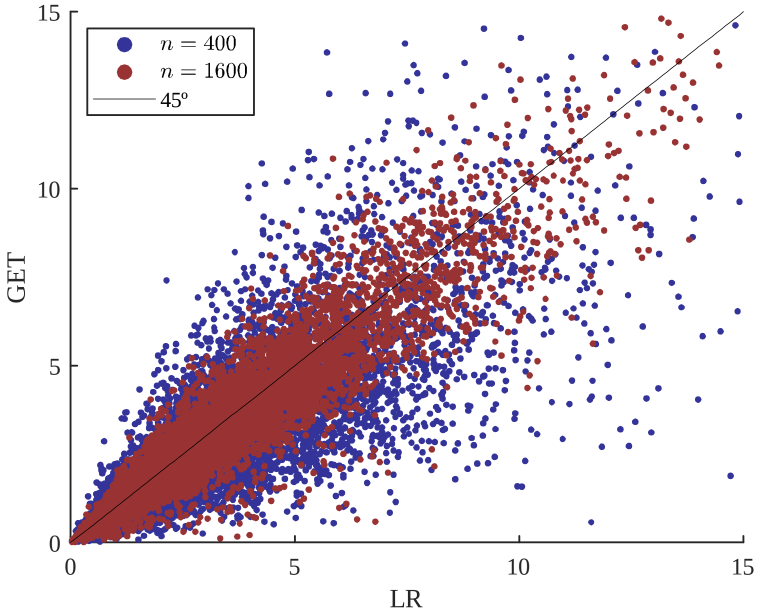
Notes: Results based on 10,000 samples. GET and LR are defined in section 3.1. GMM refers to the J-tests based on the influence functions underlying GET. Finite sample critical values are computed using the parametric bootstrap, as described in Section 2.3. LR_{naive} uses the 0.99, 0.95, and 0.9 quantiles of a χ^2_2 as critical values, while GMM_{asy} uses the 0.99, 0.95, 0.9 quantiles of a χ^2_4. DGPs: $w = x_1 = 1$ and $x_2 \sim N(0,1)$, $\boldsymbol{\varphi}_k^M = (0,1)$, $\boldsymbol{\varphi}^D = \boldsymbol{\iota}_2$, $\varphi^S = 1$ and $\varphi^L = 0.25$. As alternative hypotheses, we consider $\boldsymbol{\vartheta}' = (0.57, 0.57)$ (H_{a1}) and $\boldsymbol{\vartheta}' = (0.80, 0)$ (H_{a2}); see section 3.1 for the parametrization.

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

	Null hypothesis			Alternative hypotheses					
				H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel A: $n = 400$									
LR	*1.0*	*5.0*	*10.0*	10.6	26.8	39.4	25.0	37.5	45.2
GET	*1.0*	*5.0*	*10.0*	8.8	27.7	39.5	30.2	40.4	46.6
Panel B: $n = 1600$									
LR	*1.0*	*5.0*	*10.0*	64.3	83.1	89.7	64.7	76.4	82.2
GET	*1.0*	*5.0*	*10.0*	59.5	83.5	89.7	67.8	78.2	82.3

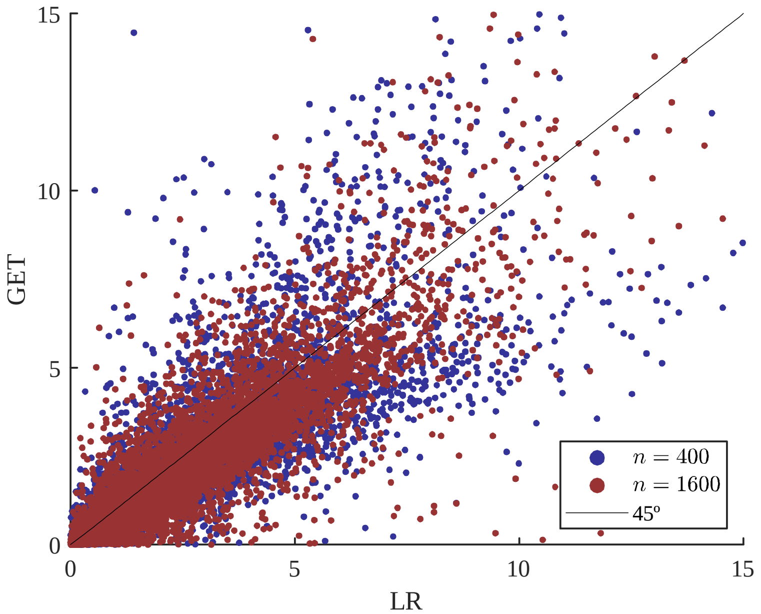
Notes: Results based on 10,000 samples. GET and LR are defined in section 3.2. Finite sample critical values are computed by simulation. DGPs: $\varphi^M = 0$, $\varphi^V = 1$, $\boldsymbol{\vartheta}' = (0.25, 0.10)$ for H_{a1}, and $\boldsymbol{\vartheta}' = (0.75, 0.05)$ for H_{a2}.

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



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

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



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