Electronic Journal of Statistics
Vol. 14 (2020) 4207–4243
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1770

Optimal pseudo-Gaussian and rank-based random coefficient detection in multiple regression

Abdelhadi Akharif[1], **Mohamed Fihri**[2], **Marc Hallin**[3], **and Amal Mellouk**[4]

[1] *Laboratoire de Mathématiques et Applications*
Faculté des Sciences et Techniques, Tangier, Morocco
e-mail: aakharif@gmail.com

[2] *Laboratory of Mathematics, Computing and Applications - Information Security (LabMiA-SI)*
Department of Mathematics, Faculty of Sciences,
Mohammed-V University in Rabat, Morocco
e-mail: m.fihri@um5s.net.ma

[3] *ECARES and Département de Mathématique*
Université libre de Bruxelles, Brussels, Belgium
e-mail: mhallin@ulb.ac.be

[4] *Centre Régional des Métiers de l'Education et de la Formation, Tangier, Morocco*
e-mail: melloukamal@hotmail.com

Abstract: Random coefficient regression (RCR) models are the regression versions of random effects models in analysis of variance and panel data analysis. Optimal detection of the presence of random coefficients (equivalently, optimal testing of the hypothesis of constant regression coefficients) has been an open problem for many years. The simple regression case has been solved recently and the multiple regression case is considered here. The latter poses several theoretical challenges: *(a)* a nonstandard ULAN structure, with log-likelihood gradients vanishing at the null; *(b)* cone-shaped alternatives under which traditional optimality concepts are no longer adequate; *(c)* nuisance parameters that are not identified under the null but have a significant impact on local powers. We propose a new (local and asymptotic) concept of optimality for this problem and, for specified error densities, derive parametrically optimal procedures. A suitable modification of the Gaussian version of the latter is shown to qualify as a pseudo-Gaussian test. The asymptotic performances of those pseudo-Gaussian tests, however, are quite poor under skewed and heavy-tailed densities. We therefore also construct rank-based tests, possibly based on data-driven scores, the asymptotic relative efficiencies of which are remarkably high with respect to their pseudo-Gaussian counterparts.

MSC2020 subject classifications: Primary 62J05, 62G10; secondary 62F03.
Keywords and phrases: Aligned rank test, cone-shaped alternative, local asymptotic normality, multiple regression model, pseudo-Gaussian test, random coefficient, unidentified nuisance, testing for homogeneity, mixture models.

Received January 2020.

Contents

1. Introduction

1.1. Random coefficients regression models

Random coefficients regression (RCR) models are the regression versions of random effects models in analysis of variance and panel data analysis. RCR models have been considered as early as Wald (1947), and received some interest in the seventies, with contributions by Hildreth and Houck (1968), Swamy (1971), Raj (1975) or Breusch and Pagan (1979), to quote only a few; see also the surveys and monographs by Raj and Ullah (1981), Nicholls and Pagan (1985), and Newbold and Bos (1985). The ultimate objective in RCR models is the estimation of the distribution of the random regression parameters; that delicate problem has been addressed, among others, by Mallet (1986), Beran and Hall (1992), Beran and Millar (1994), Beran et al. (1996); goodness-of-fit testing is considered by Delicado and Romo (1999, 2004).

An important issue in this context is the detection of the random nature of regression coefficients. Before entering the complexities of RCR modeling, indeed, one should make sure that traditional regression models with constant coefficients are to be rejected in favor of the more involved random coefficients approach. Surprisingly, the problem of optimal random coefficients detection until very recently (Fihri, Akharif, Mellouk, and Hallin 2017) had never been considered in full generality. Newbold and Bos (1985), on heuristic grounds, propose a Gaussian Lagrange Multiplier test, and Ramanathan and Rajarshi (1992) a locally most powerful (under logistic densities) signed-rank test of the Wilcoxon type. Adopting a more general Le Cam approach, Fihri et al. (2020) derive locally asymptotically optimal, optimal pseudo-Gaussian, and optimal rank-based tests for the same problem; based on local power evaluations, they also provide evidence of the poor performance of pseudo-Gaussian procedures in the presence of non-Gaussian (skew or leptokurtic densities), and recommend the use of rank-based tests with data-driven scores (Section 8.4) which significantly outperform the pseudo-Gaussian one.

All those methods, including the analysis by Fihri et al. (2020), unfortunately, are limited to simple regression models. Desirable as it is, an extension to multiple regression is far from obvious, though, with three major sources of complexity:

(a) the ULAN structure of the problem remains, as in simple RCR, of a non-standard type, with central sequences (first-order quadratic mean derivatives) involving second-order derivatives of the density;
(b) alternatives, which were one-sided in the simple regression context, are now cone-shaped, so that the usual tests (such as the Neyman $C(\alpha)$, Lagrange multiplier, or Rao efficient score tests), based on quadratic statistics are inappropriate;
(c) the correlation matrix of the random regression coefficients, which is not identified under the null, has a significant impact on local powers.

While *(a)*, thanks to the fact that quadratic mean differentiability reduces to partial quadratic mean differentiability, can be taken care of along the same lines as in Fihri et al. (2020), *(b)* and *(c)*, which are specific to the multiple regression case, are more challenging and require new solutions. Inspired by a directional concept recently proposed by Novikov (2011), we propose an adequate optimality concept, then derive locally asymptotically optimal pseudo-Gaussian and rank-based tests for the hypothesis of constant coefficients in the multiple regression setting.

The problem considered here is a particular case of the more general problem of testing for homogeneity in mixture models: see Gu (2016) and Gu et al. (2018) for a recent account and a survey of the literature. The dominant approaches in that literature are based on Neyman's $C(\alpha)$ (Lagrange multiplier) tests[1] and likelihood ratio tests. The $C(\alpha)$ tests[2] are particularly attractive due to the fact that, unlike the likelihood ratio tests, they do not require the (unpleasant) estimation of the model under the alternative. Unfortunately, $C(\alpha)$ tests are quite inefficient against cone-shaped alternatives.

1.2. The model

The model considered throughout is the random coefficients multiple regression (RCR) model

$$Y_j = \mu + \boldsymbol{\beta}_j' \mathbf{x}_j + \varepsilon_j, \quad \text{with} \quad \boldsymbol{\beta}_j = \mathbf{b} + \mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}\,\boldsymbol{\xi}_j, \quad j = 1, \ldots, n, \tag{1.1}$$

where

- Y_j is the observed response, $\mathbf{x}_j = \big(x_{1j}, \ldots, x_{pj}\big)'$, $j = 1, \ldots, n$ a non-stochastic vector of exogenous regressors;
- μ and $\mathbf{b} = (b_1, \ldots, b_p)'$ are real-valued parameters (the null values of the regression coefficients);
- ε_j, $j = 1, \ldots, n$ is an *i.i.d.* sequence of error terms with mean zero, variance σ^2, and probability density function $z \mapsto f(z) := (1/\sigma) f_1(z/\sigma)$;
- $\boldsymbol{\xi}_j := (\xi_{1j}, \ldots, \xi_{pj})'$, $j = 1, \ldots, n$ is an (unobserved) n-tuple of independent copies of some random vector $\boldsymbol{\xi}$ with mean $\mathbf{0}$, identity covariance, and probability density h;
- $\boldsymbol{\beta}_j = \big(\beta_{1j}, \ldots, \beta_{pj}\big)'$, $j = 1, \ldots, n$ are random regression coefficients, that is, n independent copies of a random vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ with mean $\mathbf{b} = \big(b_1, \ldots, b_p\big)'$ and covariance matrix $\mathbf{P}'\boldsymbol{\Lambda}^2\mathbf{P}$ ($\boldsymbol{\Lambda}^2$ diagonal with diagonal elements $\lambda_1^2, \ldots, \lambda_p^2$; $\mathbf{P}$ orthogonal); write $\boldsymbol{\lambda}^2$ and $\boldsymbol{\lambda}$ for $(\lambda_1^2, \ldots, \lambda_p^2)'$ and $(\lambda_1, \ldots, \lambda_p)'$, respectively;
- $\boldsymbol{\xi}_j$ and ε_k are independent for all $j, k = 1, \ldots, n$.

Model (1.1) clearly reduces to the classical multiple regression model

$$Y_j = \mu + \sum_{i=1}^{p} b_i x_{ij} + \varepsilon_j, \qquad j = 1, \ldots, n \tag{1.2}$$

with constant coefficients μ and b_i ($i = 1, \ldots, p$) if and only if $\boldsymbol{\lambda}^2 = \mathbf{0}$. The detection problem we are addressing thus reduces to testing the null hypothesis that $\boldsymbol{\lambda}^2$ is zero against the alternative that it has at least one strictly positive component, that is, belongs to the *blunt* (i.e., deprived of the origin) positive orthant $\mathcal{C}^+$ of $\mathbb{R}^p$; μ, $\mathbf{b}$, σ^2, $\mathbf{P}$, f_1, and h throughout are nuisance parameters. Moreover, $\mathbf{P}$ and h are not identified under the null hypothesis.

[1] In Le Cam's LAN terminology, those tests are *locally asymptotically most stringent* provided, however, that the null value of the parameter belongs to the interior of the parameter space—which is not the case here.

[2] The Newbold and Bos (1985) tests belong to that type.

1.3. A nonstandard ULAN property

In order to construct optimal tests, we first establish (Section 2) the uniform local asymptotic normality (ULAN), with respect to μ, $\mathbf{b}$, σ^2, and $\boldsymbol{\lambda}^2$, in the vicinity of $\boldsymbol{\lambda}^2 = \mathbf{0}$, of the family of distributions associated with (1.1) (under specified f_1, h, and $\mathbf{P}$). Unfortunately, the log-likelihood gradient with respect to $\boldsymbol{\lambda}^2$ vanishes at $\boldsymbol{\lambda}^2 = \mathbf{0}$. Quadratic mean differentiability of $f^{1/2}$, however, still holds, with second-order derivatives entering the quadratic mean gradient, under the same atypical (but mild) regularity assumptions as in the simple regression case.

ULAN characterizes the form of limiting local experiments, and allows us to deal, in a "classical" way, with the nuisance parameters μ, $\mathbf{b}$, and σ^2. Two serious difficulties arise, however, which are specific to the problem considered here: (i) as far as $\boldsymbol{\lambda}^2$ is concerned, ULAN is "cone-shaped"—namely, only those local perturbations of $\boldsymbol{\lambda}^2$ that belong to $\mathcal{C}^+$ are meaningful—so that the classical maximin and stringency arguments, leading to quadratic test statistics with asymptotic Neyman $C(\alpha)$ structure, do not apply; (ii) the correlation matrix $\mathbf{P}$ of the random regression parameter, which plays an essential role under the alternative, is not identified, hence cannot be estimated, under the null. We therefore consider the local and asymptotic version of an extension of the *locally asymptotically directionally maximin* tests proposed by Novikov (2011), closely related to the local and asymptotic version of the *most stringent somewhere most powerful* ideas described by Kudo (1963) and Schaafsma and Smid (1966).

1.4. Pseudo-Gaussian and rank-based tests

While the proposed directionally maximin tests no longer involve the unspecified h and $\mathbf{P}$, their form and validity still strongly depends on the (standardized) noise density f_1. Assuming f_1 to be known, as a rule, is not reasonable, and the true nature of the problem thus is semiparametric. We therefore propose pseudo-Gaussian and rank-based procedures that remain valid under a broad class of densities f_1. The resulting tests yield the same asymptotic relative efficiencies (AREs) as the locally asymptotically one-sided optimal ones in Fihri et al. (2020), to which they reduce in the particular case of simple regression. Those ARE values (with respect to the pseudo-Gaussian procedure)—especially, those of rank tests based on data-driven scores—can be quite high, demonstrating the poor performances of pseudo-Gaussian tests under non-Gaussian densities and, more particularly, the skewed and heavy-tailed ones. Rank-based tests accordingly are highly recommended in this context.

1.5. Outline of the paper

The paper is organized as follows. Section 2.1 provides the main definitions and assumptions. The uniform local asymptotic normality (ULAN), with respect to μ, $\mathbf{b}$, σ^2, and $\boldsymbol{\lambda}^2$, in the vicinity of $\boldsymbol{\lambda}^2 = \mathbf{0}$, of the family of distributions associated with (1.1) is established (under specified f_1) in Section 2.2. Sections 3.1 and 3.2 extend (still, for specified f_1) the concept of *locally asymptotically directionally maximin* tests (Novikov, 2011). The parametric version of those tests is obtained in Sections 3.3 and 3.4. Sections 4 and 5 propose pseudo-Gaussian and rank-based alternatives that remain valid irrespective of f_1. Asymptotic relative efficiencies are investigated in Section 6. Proofs, some technical details, a simulation study, and an empirical illustration are given in Sections 8 and 9. Section 10 concludes.

2. Uniform local asymptotic normality

2.1. Notation and main assumptions

The null hypothesis we are interested in is the traditional multiple linear regression model, that is, the hypothesis under which $\lambda_i^2 = 0$ in (1.1) for all $i = 1, \dots, p$. The orthogonal matrix $\mathbf{P}$ and the density h playing no role under the null, we denote by $\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};f_1} =: \mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f_1}$ the null distribution of $\mathbf{Y}^{(\mathbf{n})} := \big(Y_1^{(n)}, \dots, Y_n^{(n)}\big)'$. The same distribution under the (cone-shaped) alternative that $\lambda_i^2 > 0$ for at least one value of $i = 1, \dots, p$ is denoted by

$$\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h} =: \mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}.$$

Throughout, we consider the class of nonvanishing standardized densities

$$\mathcal{F}_0 := \Big\{ f_1 : f_1(z) > 0 \text{ for all } z \in \mathbb{R}, \int_{-1}^{1} f_1(z)dz = 0.5 = \int_{-\infty}^{0} f_1(z)dz \Big\}.$$

For $f_1 \in \mathcal{F}_0$, the median and the median absolute deviation are 0 and 1, respectively; hence, for f such that $f_1 \in \mathcal{F}_0$, 0 is the median and σ the median absolute deviation. This standardization, contrary to the usual one based on the mean and the standard deviation, avoids moment assumptions; an identification constraint, it has no impact on subsequent results. For instance, the standardized version of the normal density takes the form

$$f_1(z) = \phi_1(z) := \sqrt{a/2\pi}\exp(-az^2/2) \text{ with } a \approx 0.4549.$$

The main technical tool in our derivation is the *Uniform Local Asymptotic Normality* (ULAN) property, with respect to $(\mu, \mathbf{b}, \sigma^2, \boldsymbol{\lambda}^2)$, of the families (here and in the sequel, $\mathcal{C}^+$ denotes the (blunt) positive orthant in $\mathbb{R}^p$)

$$\mathcal{P}^{(n)}_{\mathbf{P};f_1,h} := \Big\{ \mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h} : (\mu, \mathbf{b}) \in \mathbb{R}^{p+1}, \sigma^2 > 0, \boldsymbol{\lambda}^2 \in \{\mathbf{0}\} \cup \mathcal{C}^+ \Big\}$$

in the vicinity of $(\mu, \mathbf{b}, \sigma^2, \mathbf{0}) =: (\boldsymbol{\vartheta}, \mathbf{0})$ for given f_1, h, and $\mathbf{P}$.

As in Fihri et al. (2020), this ULAN property is not the standard one based on a second-order Taylor expansion of log-likelihoods—traditional log-likelihood derivatives with respect to $\boldsymbol{\lambda}^2$ indeed vanish at $\boldsymbol{\lambda}^2 = \mathbf{0}$. It rather exploits the quadratic mean differentiability (with respect to $\boldsymbol{\lambda}^2$) of $f^{1/2}$, with quadratic mean derivatives involving the second-order derivatives of f. Mild as they are, the required regularity conditions, therefore, take atypical forms (see (A.1) and (C.2) below).

Assumption (A)

(A.1) $f_1 \in \mathcal{F}_0$; the mapping $z \mapsto f_1(z)$ is twice continuously differentiable on $\mathbb{R}$, with first and second derivatives $\dot{f}_1$ and $\ddot{f}_1$, respectively; letting $\varphi_{f_1} := -\dot{f}_1/f_1$ and $\psi_{f_1} := \ddot{f}_1/f_1$, assume that

$$\mathcal{I}_\varphi(f_1) := \int_{\mathbb{R}} \varphi^2_{f_1}(z) f_1(z) dz < \infty, \quad \mathcal{I}_\psi(f_1) := \int_{\mathbb{R}} \psi^2_{f_1}(z) f_1(z) dz < \infty,$$
$$\mathcal{J}_\varphi(f_1) := \int_{\mathbb{R}} z^2 \varphi^2_{f_1}(z) f_1(z) dz < \infty, \text{ and } \mathcal{K}_{\varphi\varphi}(f_1) := \int_{\mathbb{R}} z \varphi^2_{f_1}(z) f_1(z) dz < \infty.$$

When considering rank statistics, we moreover need an additional very mild assumption:

(A.2) φ_{f_1} and ψ_{f_1} are differences of monotone functions (that is, have bounded variation).

Under (A.1), it follows from Cauchy-Schwarz that we also have

$$\mathcal{I}_{\varphi\psi}(f_1) := \int_{\mathbb{R}} \varphi_{f_1}(z) \psi_{f_1}(z) f_1(z) dz < \infty$$

and

$$\mathcal{K}_{\varphi\psi}(f_1) := \int_{\mathbb{R}} z \varphi_{f_1}(z) \psi_{f_1}(z) f_1(z) dz < \infty.$$

Denote by $\mathcal{F}_A$ the set of all densities satisfying Assumptions (A).

Turning to the asymptotic behavior of covariates, let

$$\mathbf{C}_1^{(n)} := n^{-1} \sum_{j=1}^n \mathbf{x}_j \mathbf{x}_j'$$

and, for any orthogonal matrix $\mathbf{P}$,

$$\mathbf{C}_{\mathbf{P};2}^{(n)} := n^{-1} \sum_{j=1}^n \big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2} \big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2\prime} \quad \text{and} \quad \mathbf{C}_{\mathbf{P};3}^{(n)} := n^{-1} \sum_{j=1}^n \mathbf{x}_j \big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2\prime}.$$

Also define $\bar{\mathbf{x}}^{(n)} := n^{-1} \sum_{j=1}^n \mathbf{x}_j$, $\overline{\|\mathbf{x}\|^2}^{(n)} := n^{-1} \sum_{j=1}^n \|\mathbf{x}_j\|^2$, $\overline{\|\mathbf{x}\|^4}^{(n)} := n^{-1} \sum_{j=1}^n \|\mathbf{x}_j\|^4$, and $\overline{\big(\mathbf{P}\mathbf{x}\big)^{\odot 2}}^{(n)} := n^{-1} \sum_{j=1}^n \big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}$. Here and in the sequel, $\mathbf{M}^{\odot k} := \big(M_{ij}^k\big)$

stands for the kth Hadamard power of an arbitrary matrix $\mathbf{M} = (\mathbf{M_{ij}})$, and $\mathbf{M}^{1/2}$ denotes the symmetric square root of a symmetric positive definite matrix $\mathbf{M}$.

Assumption (B)

(B.1) The limits $\lim_{n\to\infty} \mathbf{C}_1^{(n)} =: \mathbf{C}_1$, $\lim_{n\to\infty} \mathbf{C}_{\mathbf{P};2}^{(n)} =: \mathbf{C}_{\mathbf{P};2}$, and $\lim_{n\to\infty} \mathbf{C}_{\mathbf{P};3}^{(n)} =: \mathbf{C}_{\mathbf{P};3}$ exist; $\mathbf{C}_1$ and $\mathbf{C}_{\mathbf{P};2}$ are positive definite.

It follows that $\mathbf{C}_1^{(n)}$and $\mathbf{C}_{\mathbf{P};2}^{(n)}$ are positive definite for n large enough; letting

$$\mathbf{K}_1^{(n)} := \left(\mathbf{C}_1^{(n)}\right)^{-1/2} \text{ and } \mathbf{K}_{\mathbf{P};2}^{(n)} := \left(\mathbf{C}_{\mathbf{P};2}^{(n)}\right)^{-1/2},$$

note that $\lim_{n\to\infty} \mathbf{K}_1^{(n)} =: \mathbf{K}_1 = \mathbf{C}_1^{-1/2}$ and $\lim_{n\to\infty} \mathbf{K}_{\mathbf{P};2}^{(n)} =: \mathbf{K}_{\mathbf{P};2}$ also exist, and that $\mathbf{K}_1^{(n)}$ and $\mathbf{K}_{\mathbf{P};2}^{(n)}$ have full rank (n large enough);

(B.2) the limits $\lim_{n\to\infty} \bar{\mathbf{x}}^{(n)} =: \boldsymbol{\mu}^{\mathbf{x}}$, $\lim_{n\to\infty} \overline{\|\mathbf{x}\|^2}^{(n)} =: \boldsymbol{\mu}^{\|\mathbf{x}\|^2}$, $\lim_{n\to\infty} \overline{\|\mathbf{x}\|^4}^{(n)} =: \boldsymbol{\mu}^{\|\mathbf{x}\|^4}$, and $\lim_{n\to\infty} \overline{(\mathbf{Px})^{\odot 2}}^{(n)} =: \boldsymbol{\mu}^{(\mathbf{Px})^{\odot 2}}$ exist and are finite;

(B.3) the classical *Noether conditions* hold, namely,

$$\lim_{n\to\infty} \frac{\max_{1\le j\le n}(x_{ij} - \overline{x}_i^{(n)})^2}{\sum_{j=1}^n (x_{ij} - \overline{x}_i^{(n)})^2} = 0 \quad \text{with} \quad \overline{x}_i^{(n)} := n^{-1}\sum_{j=1}^n x_{ij},\ i = 1, \ldots, p.$$

In (B.1) and (B.2), the existence of finite limits (the value of which remains unknown anyway) is assumed as a convenience; this can be dispensed with by stating asymptotic results along appropriate subsequences. None of the assumptions listed under (B) has any finite-sample implication.

Assumption (C)

(C.1) $\int \boldsymbol{\xi}\ h(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \mathbf{0}$ and $\int \boldsymbol{\xi}\boldsymbol{\xi}'\ h(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \mathbf{I}_{p\times p}$.

(C.2) Denoting by h_i, $i = 1, \ldots, p$ the marginal densities of h and letting

$$\mathcal{I}_{i\psi}^x(f_1; y) := \begin{cases} \dfrac{1}{y^2}\displaystyle\int_{z=-\infty}^{\infty} \frac{\left[\int_{w=0}^{y}\int \ddot{f}_1(z - xvw)x^2v^2h_i(v)dvdw\right]^2}{\int f_1(z - xyv)h_i(v)dv}\, dz & y > 0 \\ x^4\mathcal{I}_\psi(f_1) & y = 0 \end{cases} \tag{2.1}$$

$$= \begin{cases} \dfrac{1}{y^2}\displaystyle\int_{z=-\infty}^{\infty} \frac{\left[\int \dot{f}_1(z - xyv)xvh_i(v)dv\right]^2}{\int f_1(z - xyv)h_i(v)dv} dz, & y > 0 \\ x^4\mathcal{I}_\psi(f_1), & y = 0, \end{cases}$$

the couple (f_1, h) is such that the function $y \mapsto \mathcal{I}_{i\psi}^x(f_1; y)$ is continuous from the right at $y = 0$ for all x and $i = 1, \ldots, p$.

Such assumptions go back to Hájek (1972). Noting that (C.2) involves the *couple* (f_1, h), define, for $f_1 \in \mathcal{F}_A$ satisfying (C.1), the class of densities $\mathcal{F}_{C|f_1} := \Big\{h|\,(f_1, h) \text{ satisfies (C.2)}\Big\}$; the quantity $\mathcal{I}^x_{i\psi}(f_1; 0)$, as we shall see, is closely related to the Fisher information for $\boldsymbol{\lambda}^2$.

2.2. ULAN

In this section, ULAN with respect to $(\mu, \mathbf{b}, \sigma^2, \boldsymbol{\lambda}^2)'$ of the family $\mathcal{P}^{(n)}_{\mathbf{P};f_1,h}$ is established, for specified $\mathbf{P}$ and f_1, in the vicinity of $(\mu, \mathbf{b}, \sigma^2, \mathbf{0}) = (\boldsymbol{\vartheta}, \mathbf{0})$. For this, consider local sequences $(\boldsymbol{\vartheta}, \mathbf{0}) + n^{-1/2}\boldsymbol{\nu}^{(n)}_{\mathbf{P}}\boldsymbol{\tau}^{(n)}$ of perturbations of $(\boldsymbol{\vartheta}, \mathbf{0})$, where

$$\boldsymbol{\nu}^{(n)}_{\mathbf{P}} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \mathbf{K}^{(n)}_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \mathbf{K}^{(n)}_{\mathbf{P};2} \end{pmatrix} \tag{2.2}$$

and $\boldsymbol{\tau}^{(n)} := \big(\tau^{(n)}_1, \boldsymbol{\tau}^{(n)\prime}_2, \tau^{(n)}_3, \boldsymbol{\tau}^{(n)\prime}_4\big)' \in \mathbb{R}^{p+2} \times \mathbb{R}^p_+$ such that $\sup_n \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\tau}^{(n)} < \infty$. Incorporating $\mathbf{K}^{(n)}_1$ and $\mathbf{K}^{(n)}_{\mathbf{P};2}$ in the contiguity rates simplifies asymptotic statements; since $\mathbf{K}^{(n)}_1$ and $\mathbf{K}^{(n)}_{\mathbf{P};2}$ converge, under (B.1), to finite full-rank matrices, contiguity rates in (2.2) are of the traditional root-n magnitude.

Writing $Z^{(n)}_j$ for the standardized residuals

$$Z^{(n)}_j(\boldsymbol{\vartheta}) = Z^{(n)}_j(\mu, \mathbf{b}, \sigma^2) := \sigma^{-1}\Big(Y_j - \mu - \sum_{i=1}^p b_i x_{ij}\Big), \quad j = 1, \dots, n \tag{2.3}$$

(note that, under the null hypothesis, $Z^{(n)}_j$ coincides with $\sigma^{-1}\varepsilon_j$), we have the following result.

Proposition 2.1 (ULAN). *Let Assumptions (A)–(C) hold. Fix $f_1 \in \mathcal{F}_A$ and $h \in \mathcal{F}_{C|f_1}$. Then, the family $\mathcal{P}^{(n)}_{\mathbf{P};f_1,h}$ is ULAN, with $(2p+2)$-dimensional central sequence*

$$\boldsymbol{\Delta}^{(n)}_{\mathbf{P};f_1}(\boldsymbol{\vartheta}) := \begin{pmatrix} \Delta^{(n)}_{f_1;1}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Delta}^{(n)}_{f_1;2}(\boldsymbol{\vartheta}) \\ \Delta^{(n)}_{f_1;3}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Delta}^{(n)}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta}) \end{pmatrix} = n^{-1/2} \begin{pmatrix} \frac{1}{\sigma}\sum_{j=1}^n \varphi_{f_1}(Z_j) \\ \frac{1}{\sigma}\sum_{j=1}^n \varphi_{f_1}(Z_j)\mathbf{K}^{(n)}_1\mathbf{x}_j \\ \frac{1}{2\sigma^2}\sum_{j=1}^n (Z_j\varphi_{f_1}(Z_j) - 1) \\ \frac{1}{2\sigma^2}\sum_{j=1}^n \psi_{f_1}(Z_j)\mathbf{K}^{(n)}_{\mathbf{P};2}\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2} \end{pmatrix} \tag{2.4}$$

and $(2p+2) \times (2p+2)$ *information matrix*

$$\boldsymbol{\Gamma}_{\mathbf{P};f_1}(\boldsymbol{\vartheta})$$
$$:= \begin{pmatrix} \frac{\mathcal{I}_{\varphi}(f_1)}{\sigma^2} & \frac{\mathcal{I}_{\varphi}(f_1)}{\sigma^2}\boldsymbol{\mu}^{\mathbf{x}\prime}\mathbf{K}_1 & \frac{\mathcal{K}_{\varphi\varphi}(f_1)}{2\sigma^3} & \frac{\mathcal{I}_{\varphi\psi}(f_1)}{2\sigma^3}\boldsymbol{\mu}^{(\mathbf{Px})^{\odot 2}\prime}\mathbf{K}_{\mathbf{P};2} \\ \frac{\mathcal{I}_{\varphi}(f_1)}{\sigma^2}\mathbf{K}_1\boldsymbol{\mu}^{\mathbf{x}} & \frac{\mathcal{I}_{\varphi}(f_1)}{\sigma^2}\mathbf{I}_{p\times p} & \frac{\mathcal{K}_{\varphi\varphi}(f_1)}{2\sigma^3}\mathbf{K}_1\boldsymbol{\mu}^{\mathbf{x}} & \frac{\mathcal{I}_{\varphi\psi}(f_1)}{2\sigma^3}\mathbf{K}_1\mathbf{C}_{\mathbf{P};3}\mathbf{K}_{\mathbf{P};2} \\ \frac{\mathcal{K}_{\varphi\varphi}(f_1)}{2\sigma^3} & \frac{\mathcal{K}_{\varphi\varphi}(f_1)}{2\sigma^3}\boldsymbol{\mu}^{\mathbf{x}\prime}\mathbf{K}_1 & \frac{\mathcal{J}_{\varphi}(f_1)-1}{4\sigma^4} & \frac{\mathcal{K}_{\varphi\psi}(f_1)}{4\sigma^4}\boldsymbol{\mu}^{(\mathbf{Px})^{\odot 2}\prime}\mathbf{K}_{\mathbf{P};2} \\ \frac{\mathcal{I}_{\varphi\psi}(f_1)}{2\sigma^3}\mathbf{K}_{\mathbf{P};2}\boldsymbol{\mu}^{(\mathbf{Px})^{\odot 2}} & \frac{\mathcal{I}_{\varphi\psi}(f_1)}{2\sigma^3}\mathbf{K}_{\mathbf{P};2}\mathbf{C}'_{\mathbf{P};3}\mathbf{K}_1 & \frac{\mathcal{K}_{\varphi\psi}(f_1)}{4\sigma^4}\mathbf{K}_{\mathbf{P};2}\boldsymbol{\mu}^{(\mathbf{Px})^{\odot 2}} & \frac{\mathcal{I}_{\psi}(f_1)}{4\sigma^4}\mathbf{I}_{p\times p} \end{pmatrix}. \tag{2.5}$$

More precisely, for any sequence of the form

$$(\boldsymbol{\vartheta}^{(n)},\mathbf{0}) := (\mu^{(n)},\mathbf{b}^{(n)},\sigma^{2(n)},\mathbf{0})$$

such that $\mu^{(n)}-\mu$, $\left(\mathbf{K}_1^{(n)}\right)^{-1}\left(\mathbf{b}^{(n)}-\mathbf{b}\right)$, *and* $\sigma^{2(n)}-\sigma^2$ *are* $O\left(n^{-1/2}\right)$, *and for any bounded sequence* $\boldsymbol{\tau}^{(n)} := \left(\tau_1^{(n)},\boldsymbol{\tau}_2^{(n)},\tau_3^{(n)},\boldsymbol{\tau}_4^{(n)}\right)$ *in* $\mathbb{R}^{p+2}\times\mathcal{C}^+$, *we have, under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^{(n)},\mathbf{0};f_1}$,

$$\begin{aligned}\Lambda^{(n)}_{(\boldsymbol{\vartheta}^{(n)},\mathbf{0})+n^{-1/2}\boldsymbol{\nu}^{(n)}_{\mathbf{P}}\boldsymbol{\tau}^{(n)}/(\boldsymbol{\vartheta}^{(n)},\mathbf{0}),\mathbf{P};f_1,h} &:= \log\frac{d\mathrm{P}^{(n)}_{(\boldsymbol{\vartheta}^{(n)},\mathbf{0})+n^{-1/2}\boldsymbol{\nu}^{(n)}_{\mathbf{P}}\boldsymbol{\tau}^{(n)},\mathbf{P};f_1,h}}{d\mathrm{P}^{(n)}_{(\boldsymbol{\vartheta}^{(n)},\mathbf{0});f_1}} \\ &= \boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Delta}^{(n)}_{\mathbf{P};f_1}(\boldsymbol{\vartheta}^{(n)}) - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\boldsymbol{\Gamma}_{\mathbf{P};f_1}(\boldsymbol{\vartheta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1),\end{aligned} \tag{2.6}$$

and $\boldsymbol{\Delta}^{(n)}_{\mathbf{P};f_1}(\boldsymbol{\vartheta}^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}\big(\mathbf{0}_{2p+2},\boldsymbol{\Gamma}_{\mathbf{P};f_1}(\boldsymbol{\vartheta})\big)$, *as* $n\to\infty$.

See Section 7.1 for a proof.

3. Locally asymptotically optimal parametric tests

3.1. Locally asymptotically directionally maximin *tests*

This section is about the local asymptotic optimality, in general multiparameter LAN families, of tests with cone-shaped alternatives involving nuisances that are not identified under the null.

Let $\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} : \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ be some LAN family, with p-dimensional parameter $\boldsymbol{\theta}$, root-n contiguity rates, central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$, and information matrix $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$. Recall that $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$, under sequences of parameter values of the form $\boldsymbol{\theta}+n^{-1/2}\boldsymbol{\tau}$ (local alternatives), converges in distribution (as $n\to\infty$) to a Gaussian random vector $\boldsymbol{\Delta}\sim\mathcal{N}\big(\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau},\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\big)$. Let $\varkappa(\boldsymbol{\Delta})$ be some optimal test in the *Gaussian shift model* $\left\{\mathcal{N}\left(\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\tau},\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\right)|\boldsymbol{\tau}\in\mathbb{R}^p\right\}$ ($\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ specified) describing some (hypothetical) observation $\boldsymbol{\Delta}$: then, the sequence of tests $\varkappa(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}))$ enjoys a local (at $\boldsymbol{\theta}$) and asymptotic form of the same optimality property: if $\varkappa(\boldsymbol{\Delta})$ is uniformly most powerful, $\varkappa(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}))$ is *locally asymptotically most powerful*; if $\varkappa(\boldsymbol{\Delta})$ is most stringent, $\varkappa(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}))$ is *locally asymptotically most stringent*, etc. See Le Cam (1986) or Le Cam and Yang (2000) for details and a justification.

Next, consider the Gaussian shift experiment $\{\mathcal{N}(\mathbf{\Gamma}\boldsymbol{\tau}, \mathbf{\Gamma}) \,|\, \boldsymbol{\tau} \in \mathbb{R}^p\}$ with observation $\mathbf{\Delta}$. Fix $\mathbf{u} \in \mathcal{S}_{p-1}$, where $\mathcal{S}_{p-1}$, as usual, denotes the unit sphere in $\mathbb{R}^p$, and consider the test $\varkappa_{\mathbf{u}}(\mathbf{\Delta})$ rejecting the null hypothesis $\boldsymbol{\tau} = \mathbf{0}$ in favor of $\boldsymbol{\tau} = \tau\mathbf{u}$, $\tau > 0$ whenever $\mathbf{u}'\mathbf{\Delta} > z_\alpha \left(\mathbf{u}'\mathbf{\Gamma}\mathbf{u}\right)^{1/2}$ (with z_α the standard normal quantile of order $1-\alpha$). That test is uniformly most powerful at probability level α: call it *directionally most powerful*. The power of $\varkappa_{\mathbf{u}}(\mathbf{\Delta})$ against $\boldsymbol{\tau} = \tau\mathbf{v}$ (that is, against $\mathbf{\Delta} \sim \mathcal{N}(\tau\mathbf{\Gamma}\mathbf{v}, \mathbf{\Gamma})$), where $\tau > 0$ and $\mathbf{v} \in \mathcal{S}_{p-1}$, is

$$1 - \Phi\big(z_\alpha - \tau\mathbf{u}'\mathbf{\Gamma}\mathbf{v}/(\mathbf{u}'\mathbf{\Gamma}\mathbf{u})^{1/2}\big)$$

where Φ, as usual, stands for the standard normal distribution function. At $\tau = 0$, the derivative with respect to τ of this power—a directional derivative in direction $\mathbf{v}$—is $\phi(z_\alpha)\mathbf{u}'\mathbf{\Gamma}\mathbf{v}/(\mathbf{u}'\mathbf{\Gamma}\mathbf{u})^{1/2}$ which, for given α and $\mathbf{u}$, is a monotone increasing function of $\mathbf{u}'\mathbf{\Gamma}\mathbf{v}$, hence of the cosine $\cos(\mathbf{v}, \mathbf{\Gamma}\mathbf{u})$.

In the same Gaussian shift, consider testing $\boldsymbol{\tau} = \mathbf{0}$ against the multidirectional alternative $\boldsymbol{\tau} \in \mathcal{C}$, where $\mathcal{C}$ is some given *blunt* closed convex half-cone with vertex at the origin (recall that a *blunt* closed cone is a closed cone deprived of its vertex). For any $\mathbf{u} \in \mathcal{C} \cap \mathcal{S}_{p-1}$, denote by $\mathbf{v}_{\mathbf{u}}$ the least favorable direction for $\varkappa_{\mathbf{u}}$ within $\mathcal{C}$, namely,

$$\mathbf{v}_{\mathbf{u}} = \mathrm{argmin}_{\mathbf{v}\in\mathcal{C}\cap\mathcal{S}_{p-1}} \mathbf{u}'\mathbf{\Gamma}\mathbf{v} = \mathrm{argmin}_{\mathbf{v}\in\mathcal{C}\cap\mathcal{S}_{p-1}} \cos(\mathbf{v}, \mathbf{\Gamma}\mathbf{u}).$$

Such $\mathbf{v}_{\mathbf{u}}$ exists, as $\mathbf{v} \mapsto \mathbf{u}'\mathbf{\Gamma}\mathbf{v}$ is continuous and $\mathcal{C} \bigcap \mathcal{S}_{p-1}$ is compact; it is not unique, though, and can be chosen as any unit vector at the intersection of $\mathcal{C}$ and the surface of the cone of revolution with axis $\mathbf{\Gamma}\mathbf{u}$ circumscribing $\mathcal{C}$ (that is, the smallest cone of revolution with given axis $\mathbf{\Gamma}\mathbf{u}$ containing $\mathcal{C}$).

Now, let $\mathbf{u}^* \in \mathcal{C} \bigcap \mathcal{S}_{p-1}$ be such that the power at $\boldsymbol{\tau} = \tau\mathbf{v}_{\mathbf{u}^*}$ of $\varkappa_{\mathbf{u}^*}$ be maximal for any given $\tau > 0$, that is,

$$1 - \Phi\big(z_\alpha - \tau\mathbf{u}^{*\prime}\mathbf{\Gamma}\mathbf{v}_{\mathbf{u}^*}/(\mathbf{u}^{*\prime}\mathbf{\Gamma}\mathbf{u}^*)^{1/2}\big) \geq 1 - \Phi\big(z_\alpha - \tau\mathbf{u}'\mathbf{\Gamma}\mathbf{v}_{\mathbf{u}}/(\mathbf{u}'\mathbf{\Gamma}\mathbf{u})^{1/2}\big) \quad (3.1)$$

for all $\mathbf{u} \in \mathcal{C} \bigcap \mathcal{S}_{p-1}$ and $\tau > 0$ or, equivalently,

$$\mathbf{u}^* = \mathrm{argmax}_{\mathbf{u}\in\mathcal{C}\cap\mathcal{S}_{p-1}} \min_{\mathbf{v}\in\mathcal{C}\cap\mathcal{S}_{p-1}} \cos(\mathbf{v}, \mathbf{\Gamma}\mathbf{u}). \quad (3.2)$$

Again, due to continuity and the compactness of $\mathcal{C} \bigcap \mathcal{S}_{p-1}$, such $\mathbf{u}^*$ exists, and is such that $\mathbf{\Gamma}\mathbf{u}^*$ is the axis of the cone of revolution circumscribing $\mathcal{C}$—provided that this axis belongs to $\mathcal{C}$. If it does not, then $\mathbf{\Gamma}\mathbf{u}^*$ should be chosen as the axis of the cone of revolution circumscribing the subpolyhedral cone $\mathcal{C}_J$ spanned by a subset of J vertices of $\mathcal{C}$ and such that *(i)* this axis belongs to $\mathcal{C}_J$, and *(ii)* the angle of this axis with any of the vertices of $\mathcal{C}$ that do not belong to $\mathcal{C}_J$ is less than the J equal angles between this axis and the J vertices spanning $\mathcal{C}_J$: see Abelson and Tukey (1963) for details and an algorithmic solution. That $\mathbf{u}^*$ clearly does not depend on α nor τ.

The test $\varkappa^*(\mathbf{\Delta}) := \varkappa_{\mathbf{u}^*}(\mathbf{\Delta})$ is called *directionally maximin*, as it maximizes, for any given $\tau > 0$, the minimum power over the class of all directionally optimal tests with size α. Accordingly, with the notation previously adopted for LAN families (and with $\mathbf{\Gamma} = \mathbf{\Gamma}_{\boldsymbol{\theta}}$), the sequence of tests $\varkappa^*(\mathbf{\Delta}_{\boldsymbol{\theta}}^{(n)})$ is *locally* *asymptotically directionally maximin* at $\boldsymbol{\theta}$ and asymptotic probability level α—as well as any asymptotically equivalent sequence of tests.

This intuitively appealing concept of *directionally maximin* test has been proposed, for $\mathcal{C} = \mathcal{C}^+$, by Novikov (2011) in a broader context, where densities are not necessarily Gaussian, and optimality is described in terms of derivatives of power functions. Gaussian models are treated as a particular case in Section 3 (same reference); the characterization of the solution in terms of cones of revolution, which is not provided there, follows from previous results by Abelson and Tukey (1963) and Schaafsma and Smid (1966).

The concept of *most stringent somewhere most powerful* test was introduced by Kudo (1963) and Schaafsma and Smid (1966). Unlike Novikov's, Schaafsma and Smid's concept is restricted to Gaussian densities. On the other hand, the hypotheses considered by Schaafsma and Smid are of a more general form. In the notation adopted here, their null hypothesis is $\mathbf{H}'\boldsymbol{\Gamma}\boldsymbol{\tau} = \mathbf{0}$, where $\mathbf{H}$ is some given $p \times q$ matrix of rank $1 \leq q \leq p$, while under the alternative the q components of $\mathbf{H}'\boldsymbol{\Gamma}\boldsymbol{\tau}$ are non-negative, with one of them at least strictly positive (a blunt cone with apex at the origin, thus). For each direction in that cone, one-sided uniformly most powerful directional tests could be derived similarly as above. The terminology "most stringent" is used instead of "maximin" because, when applied to a cone of $\boldsymbol{\tau}$ values, the maximin argument leads to a trivial maximin power α (corresponding to $\|\boldsymbol{\tau}\| \to \mathbf{0}$)—something Novikov avoids by considering directions whereas the stringency argument recurs to Wald's concept of *minimal regret*. This is, however, a formal terminological detail: the two concepts, in the Gaussian case, coincide—and so do their local and asymptotic versions in LAN families (see Akharif and Hallin (2003) for an application in the context of random coefficients autoregressive models). For the sake of simplicity, we adopt Novikov's *directionally maximin* terminology.

3.2. Gaussian shift with nuisance

In the present context, however, we need yet a slight extension of the optimality concept just described, due to the fact that, besides the parameter of interest $\boldsymbol{\tau}$, the limiting Gaussian shifts in the ULAN experiments of Proposition 2.1 also involve a nuisance parameter $\mathbf{P} \in \boldsymbol{\Pi}$, say (here, the class $\boldsymbol{\Pi}_p$ of $p \times p$ orthogonal matrices), which is not identified under $\boldsymbol{\tau} = \mathbf{0}$. Accordingly, *directionally most powerful*, in the sequel, does not reduce to (locally asymptotically) *most powerful along some* $\boldsymbol{\tau}$, but is to be understood as (locally asymptotically) *most powerful along some* $\boldsymbol{\tau}$ *for some value* $\mathbf{P}$ *of the nuisance*; the directional maximin argument then is applied to the set of possible values of $(\boldsymbol{\tau}, \mathbf{P})$. More precisely, denote by $\varkappa_{\mathbf{u},\mathbf{P}}$ the directionally most powerful test for $\boldsymbol{\tau} = \mathbf{0}$ in favor of $\boldsymbol{\tau} = \tau\mathbf{u}$, $\tau > 0$ under the value $\mathbf{P}$ of the nuisance, and let $\beta_{\mathbf{u},\mathbf{P}}(\tau\mathbf{v}, \mathbf{Q})$ be the power of that test under the alternative characterized by the parameter and nuisance values $\tau\mathbf{v}$, $\tau > 0$ and $\mathbf{Q}$, respectively. Assuming that it exists, denote by $\dot{\beta}_{\mathbf{u},\mathbf{P}}(\mathbf{v}, \mathbf{Q})$ the value at $\tau = 0$ of the derivative with respect to τ of this power.

The test $\varkappa^* := \varkappa_{\mathbf{u}^*,\mathbf{P}^*}$ will be called *directionally maximin* for the cone alternative $\mathcal{C}$ if

$$(\mathbf{u}^*, \mathbf{P}^*) = \mathrm{argmax}_{(\mathbf{u},\mathbf{P})\in(\mathcal{C}\cap\mathcal{S}_{p-1})\times\mathbf{\Pi}} \mathrm{min}_{(\mathbf{v},\mathbf{Q})\in(\mathcal{C}\cap\mathcal{S}_{p-1})\times\mathbf{\Pi}} \ \dot{\beta}_{\mathbf{u},\mathbf{P}}(\mathbf{v}, \mathbf{Q}).$$

Depending on the role of $\mathbf{P}$ in the Gaussian shift, such $\varkappa^*$ may or may not exist. Proposition 3.1 shows that, for the specific problem considered here, such tests do exist and admit an explicit form.

3.3. Optimal parametric tests: specified parameters

To start with, let us assume that the parameters μ, $\mathbf{b}$, and σ^2 (to be considered as nuisance parameters in Section 3.4) are specified, as well as the densities $f_1 \in \mathcal{F}_A$ and $h \in \mathcal{F}_{C|f_1}$. We are thus testing the sequence of simple null hypotheses $\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};f_1}\}$ against the parametric alternatives $\bigcup_{\boldsymbol{\lambda}^2,\,\mathbf{P}} \{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}\}$ where $\boldsymbol{\lambda}^2$ and $\mathbf{P}$ range over $\mathcal{C}^+$ and the set $\mathbf{\Pi}_p$ of $p \times p$ full-rank orthogonal matrices, respectively. The following result shows that, quite remarkably, a very simple maximin element exists within the collection of directionally most powerful tests for this problem, based on the test statistic

$$T^{(n)}_{f_1}(\boldsymbol{\vartheta}) := \Big(\mathcal{I}_\psi(f_1)\sum_{j=1}^n \|\mathbf{x}_j\|^4\Big)^{-1/2} \sum_{j=1}^n \psi_{f_1}(Z_j)\|\mathbf{x}_j\|^2. \tag{3.3}$$

More precisely, we prove the following result.

Proposition 3.1. *Let Assumptions (B) hold, fix $f_1 \in \mathcal{F}_A$, and assume that $h \in \mathcal{F}_{C|f_1}$. Then,*

- (i) *the test statistic $T^{(n)}_{f_1}(\boldsymbol{\vartheta})$ defined in (3.3) is asymptotically normal, with mean zero and variance one under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f_1}$;*
- (ii) *under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},n^{-1/2}\mathbf{K}^{(n)}_{\mathbf{P};2}\boldsymbol{\tau}_4,\mathbf{P};f_1,h}$, and along subsequences such that the limit exists, $T^{(n)}_{f_1}(\boldsymbol{\vartheta})$ is asymptotically normal with variance one, but with mean*

$$\mu_{f_1} := \big(\mathcal{I}^{1/2}_\psi(f_1)/2\sigma^2\big) \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \|\mathbf{x}_j\|^2 \big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2\prime} \mathbf{K}^{(n)}_{\mathbf{P};2}\boldsymbol{\tau}_4; \tag{3.4}$$

- (iii) *the sequence of tests rejecting the null hypothesis $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f_1}$ whenever $T^{(n)}_{f_1}$ exceeds the standard normal $(1-\alpha)$-quantile z_α is locally asymptotically directionally maximin at asymptotic size α, against alternatives of the form*

$$\bigcup_{\boldsymbol{\lambda}^2\in\mathcal{C}^+} \bigcup_{\mathbf{P}\in\mathbf{\Pi}_p} \bigcup_{h\in\mathcal{F}_{C|f_1}} \left\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}\right\}.$$

See Section 7.2 for the proof.

3.4. Optimal parametric tests: unspecified parameters

The test described in Proposition 3.1 settles the efficiency bounds, but is of theoretical interest only, as in practice neither the nuisance parameters μ, $\mathbf{b}$, and σ^2 nor the density f_1 are known. First, let us deal with the parametric nuisances μ, $\mathbf{b}$, and σ^2, under specified f_1. The null hypothesis then is

$$\bigcup_{\mu\in\mathbb{R}}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\left\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};f_1}\right\},$$

with alternative

$$\bigcup_{\mu\in\mathbb{R}}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\bigcup_{\mathbf{P}\in\mathbf{\Pi}_p}\bigcup_{h\in\mathcal{F}_{C|f_1}}\bigcup_{\boldsymbol{\lambda}^2\in\mathcal{C}^+}\left\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}\right\}. \tag{3.5}$$

The problem thus is about testing a p-tuple of linear constraints (viz., $\boldsymbol{\lambda}^2=\mathbf{0}$) on a parameter $(\boldsymbol{\vartheta},\boldsymbol{\lambda}^2)=(\mu,\mathbf{b},\sigma^2,\boldsymbol{\lambda}^2)$ ranging over $\mathbb{R}^{1+p}\times\mathbb{R}_0^+\times(\{\mathbf{0}\}\cup\mathcal{C}^+)$. Proposition 3.2 below shows that an optimal test for this can be based on a test statistic of the form $T^{(n)*}_{f_1}(\hat{\boldsymbol{\vartheta}}^{(n)})$, where $\hat{\boldsymbol{\vartheta}}^{(n)}=(\hat{\mu}^{(n)},\hat{\mathbf{b}}^{(n)},\hat{\sigma}^{2(n)},\mathbf{0})'$, as an estimator of $\boldsymbol{\vartheta}$, is $n^{1/2}$*-consistent* and *locally asymptotically discrete* (see Section 7.3 for a definition), and

$$\begin{aligned}T^{(n)*}_{f_1}(\boldsymbol{\vartheta}) := \;& \left(\Gamma^{(n)*}_{f_1}(\boldsymbol{\vartheta})\right)^{-1/2}\Bigg[\frac{1}{2\sigma^2\sqrt{n}}\sum_{j=1}^n\psi_{f_1}(Z_j)\|\mathbf{x}_j\|^2\\
&\times\frac{1}{2\sigma^3}\begin{pmatrix}\mathcal{I}_{\varphi\psi}(f_1)\overline{\|\mathbf{x}\|^2}^{(n)}\\ \mathcal{I}_{\varphi\psi}(f_1)\frac{1}{n}\sum_{j=1}^n\|\mathbf{x}_j\|^2\mathbf{x}_j'\mathbf{K}_1^{(n)}\\ \frac{1}{2\sigma}\mathcal{K}_{\varphi\psi}(f_1)\overline{\|\mathbf{x}\|^2}^{(n)}\end{pmatrix}'\\
&\times\begin{pmatrix}\Gamma^{(n)}_{f_1;11}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;12}(\boldsymbol{\vartheta}) & \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta})\\ \boldsymbol{\Gamma}^{(n)\prime}_{f_1;12}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;22}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;23}(\boldsymbol{\vartheta})\\ \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)\prime}_{f_1;23}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;33}(\boldsymbol{\vartheta})\end{pmatrix}^{-1}\begin{pmatrix}\Delta^{(n)}_{f_1;1}(\boldsymbol{\vartheta})\\ \boldsymbol{\Delta}^{(n)}_{f_1;2}(\boldsymbol{\vartheta})\\ \Delta^{(n)}_{f_1;3}(\boldsymbol{\vartheta})\end{pmatrix}\Bigg],\end{aligned} \tag{3.6}$$

(which does not depend on $\boldsymbol{\lambda}^2$ nor $\mathbf{P}$), with

$$\begin{aligned}\Gamma^{(n)*}_{f_1}(\boldsymbol{\vartheta}) \;:=\;& \frac{\mathcal{I}_\psi(f_1)}{4\sigma^4}\overline{\|\mathbf{x}\|^4}^{(n)}-\frac{1}{4\sigma^6}\begin{pmatrix}\mathcal{I}_{\varphi\psi}(f_1)\overline{\|\mathbf{x}\|^2}^{(n)}\\ \mathcal{I}_{\varphi\psi}(f_1)\frac{1}{n}\sum_{j=1}^n\|\mathbf{x}_j\|^2\mathbf{x}_j'\mathbf{K}_1^{(n)}\\ \frac{\mathcal{K}_{\varphi\psi}(f_1)}{2\sigma}\overline{\|\mathbf{x}\|^2}^{(n)}\end{pmatrix}'\\
&\times\begin{pmatrix}\Gamma^{(n)}_{f_1;11}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;12}(\boldsymbol{\vartheta}) & \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta})\\ \boldsymbol{\Gamma}^{(n)\prime}_{f_1;12}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;22}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;23}(\boldsymbol{\vartheta})\\ \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)\prime}_{f_1;23}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;33}(\boldsymbol{\vartheta})\end{pmatrix}^{-1}\begin{pmatrix}\mathcal{I}_{\varphi\psi}(f_1)\overline{\|\mathbf{x}\|^2}^{(n)}\\ \mathbf{K}_1^{(n)}\frac{\mathcal{I}_{\varphi\psi}(f_1)}{n}\sum_{j=1}^n\mathbf{x}_j\|\mathbf{x}_j\|^2\\ \frac{\mathcal{K}_{\varphi\psi}(f_1)}{2\sigma}\overline{\|\mathbf{x}\|^2}^{(n)}\end{pmatrix}.\end{aligned}$$

Writing $\Gamma_{f_1}^{(n)*}$ and $\widehat{\Gamma}_{f_1}^{(n)*}$, respectively, for $\Gamma_{f_1}^{(n)*}(\boldsymbol{\vartheta})$ and $\Gamma_{f_1}^{(n)*}(\hat{\boldsymbol{\vartheta}}^{(n)})$, the continuous mapping theorem then implies, since the mapping $\boldsymbol{\vartheta} \mapsto \Gamma_{f_1}(\boldsymbol{\vartheta})$ is continuous, that $\widehat{\Gamma}_{f_1}^{(n)*} - \Gamma_{f_1}^{(n)*}$ is $o_{\mathrm{P}}(1)$.

The properties of $T_{f_1}^{(n)*}(\hat{\boldsymbol{\vartheta}}^{(n)})$-based tests are summarized as follows; see Section 7.3 for Assumptions (D) and a proof.

Proposition 3.2. *Let Assumptions (B) and (C) hold. Assume that $\hat{\boldsymbol{\vartheta}}^{(n)}$ satisfies Assumptions (D). Fix $f_1 \in \mathcal{F}_A$ and $h \in \mathcal{F}_{C|f_1}$. Then, for any $\boldsymbol{\vartheta}$ of the form $(\mu, \mathbf{b}, \sigma^2, \mathbf{0})'$,*

(i) $T_{f_1}^{(n)*}(\hat{\boldsymbol{\vartheta}}^{(n)}) = T_{f_1}^{(n)*}(\boldsymbol{\vartheta}) + o_{\mathrm{P}}(1)$ is asymptotically normal, with mean zero under $\mathrm{P}_{\boldsymbol{\vartheta};f_1}^{(n)}$, mean

$$\mu_T^* := \lim_{n\to\infty} \Big(\Gamma_{f_1}^{(n)*}(\boldsymbol{\vartheta})\Big)^{-1/2} \left[\frac{\mathcal{I}_\psi(f_1)}{4n\sigma^4} \sum_{j=1}^n \|\mathbf{x}_j\|^2 \big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2\prime} \mathbf{K}_{\mathbf{P};2}^{(n)} \right. \tag{3.7}$$
$$- \frac{1}{2\sigma^3} \left(\frac{\mathcal{I}_{\varphi\psi}(f_1)}{n} \sum_{j=1}^n \|\mathbf{x}_j\|^2 \; \frac{\mathcal{I}_{\varphi\psi}(f_1)}{n} \sum_{j=1}^n \|\mathbf{x}_j\|^2 \mathbf{x}_j' \mathbf{K}_1^{(n)} \; \frac{\mathcal{K}_{\varphi\psi}(f_1)}{2n\sigma} \sum_{j=1}^n \|\mathbf{x}_j\|^2 \right)$$
$$\left. \times \begin{pmatrix} \Gamma_{f_1;11}^{(n)}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}_{f_1;12}^{(n)}(\boldsymbol{\vartheta}) & \Gamma_{f_1;13}^{(n)}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Gamma}_{f_1;12}^{(n)\prime}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}_{f_1;22}^{(n)}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}_{f_1;23}^{(n)}(\boldsymbol{\vartheta}) \\ \Gamma_{f_1;13}^{(n)}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}_{f_1;23}^{(n)\prime}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}_{f_1;33}^{(n)}(\boldsymbol{\vartheta}) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\Gamma}_{\mathbf{P};f_1;14}^{(n)}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Gamma}_{\mathbf{P};f_1;24}^{(n)}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Gamma}_{\mathbf{P};f_1;34}^{(n)}(\boldsymbol{\vartheta}) \end{pmatrix} \right] \boldsymbol{\tau}_4$$

under $\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),n^{-1/2}\mathbf{K}_{\mathbf{P};2}^{(n)}\boldsymbol{\tau}_4,\mathbf{P};f_1,h}$ (and along subsequences such that the limit exists), and variance one under both;

(ii) the sequence of tests rejecting $\bigcup_{\mu\in\mathbb{R}^p} \bigcup_{\mathbf{b}\in\mathbb{R}^p} \bigcup_{\sigma^2\in\mathbb{R}^+} \big\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};f_1}\big\}$ whenever the test statistic $T_{f_1}^{(n)*}(\hat{\boldsymbol{\vartheta}}^{(n)})$ exceeds the $(1-\alpha)$ standard normal quantile z_α is locally asymptotically directionally maximin at asymptotic level α against alternatives of the form

$$\bigcup_{\mu\in\mathbb{R}^p} \bigcup_{\mathbf{b}\in\mathbb{R}^p} \bigcup_{\sigma^2\in\mathbb{R}^+} \bigcup_{\mathbf{P}\in\boldsymbol{\Pi}_p} \bigcup_{h\in\mathcal{F}_{C|f_1}} \bigcup_{\boldsymbol{\lambda}^2\in\mathcal{C}^+} \Big\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}\Big\}.$$

Since the model, under the null hypothesis, reduces to the traditional regression model, the traditional MLE is a natural option for $\hat{\boldsymbol{\vartheta}}^{(n)}$—provided that f can be assumed to have finite variance. A finite-variance density f, however, is not required here, so that a more robust $\hat{\boldsymbol{\vartheta}}^{(n)}$, such as the LAD estimator, may be preferable.

Now, the resulting $T_{f_1}^{(n)*}$-based tests, typically, are valid under a correctly specified density only—that is, under $g_1 = f_1$ where g_1 stands for the actual density. An important exception is the case of a Gaussian f_1, though: in the next section, we show that an adequately modified version of the Gaussian test enjoys the pseudo-Gaussian property, and remains valid under non-Gaussian g_1.

4. Pseudo-Gaussian test

The Gaussian version of $T_{f_1}^{(n)*}(\boldsymbol{\vartheta})$ is easily obtained by letting $f_1 = \phi_1$ (ϕ_1 the standard normal density; see Section 2.1) in (3.7), yielding

$$T_{\phi_1}^{(n)*}(\boldsymbol{\vartheta}) = \left(\Gamma_{\phi_1}^{(n)*}\right)^{-1/2} \frac{a^2}{2\sigma^2\sqrt{n}} \sum_{j=1}^{n} \left(Z_j^{(n)}\right)^2 \left(\|\boldsymbol{x}_j\|^2 - \overline{\|\boldsymbol{x}\|^2}^{(n)}\right). \tag{4.1}$$

with $Z_j^{(n)} = Z_j^{(n)}(\boldsymbol{\vartheta})$ as in (2.3). If, however, a pseudo-Gaussian test is to be performed, the standardization constant in (4.1) is no longer correct under $g_1 \neq \phi_1$: letting

$$V_{\|\mathbf{x}\|^2}^{(n)} := \frac{1}{n}\sum_{j=1}^{n} \left(\|\boldsymbol{x}_j\|^2 - \overline{\|\boldsymbol{x}\|^2}^{(n)}\right)^2, \quad \mu(g_1) := \int z g_1(z) dz,$$

and

$$\mu_j(g_1) := \int (z - \mu(g_1))^j g_1(z) dz, \quad j = 2, 3, 4,$$

the variance $\Gamma_{\phi_1}^{(n)*}$ in (4.1) is to be replaced with

$$\Gamma_{\phi_1;g_1}^{(n)*} := \frac{a^4}{4\sigma^4}\left(\mu_4(g_1) - \mu_2^2(g_1)\right) V_{\|\mathbf{x}\|^2}^{(n)}$$

(provided that $\mu_4(g_1) < \infty$) or any consistent estimator thereof. Putting

$$\bar{Z}^{(n)} := \frac{1}{n}\sum_{i=1}^{n} Z_i^{(n)} \quad \text{and} \quad m_j^{(n)} := \frac{1}{n}\sum_{i=1}^{n}\left(Z_i^{(n)} - \bar{Z}^{(n)}\right)^j,$$

and assuming that g_1 is in the class $\mathcal{F}_A^{(4)}$ of densities in $\mathcal{F}_A$ with $\mu_4(g_1) < \infty$,

$$\widehat{\Gamma}_{\phi_1}^{(n)*} := \frac{a^4}{4\sigma^4}\left(m_4^{(n)} - (m_2^{(n)})^2\right) V_{\|\mathbf{x}\|^2}^{(n)}$$

is such a consistent estimator of $\Gamma_{\phi_1;g_1}^{(n)*}$. Denote by $T_{\phi_1}^{(n)\bullet}(\boldsymbol{\vartheta})$ the statistic resulting from substituting $\widehat{\Gamma}_{\phi_1}^{(n)*}$ for $\Gamma_{\phi_1}^{(n)*}$ in (4.1). Then, for any $g_1 \in \mathcal{F}_A^{(4)}$ and $\hat{\boldsymbol{\vartheta}}^{(n)}$ satisfying Assumptions (D), $T_{\phi_1}^{(n)\bullet}(\hat{\boldsymbol{\vartheta}}^{(n)})$ under $\mathrm{P}_{\boldsymbol{\vartheta};g_1}^{(n)}$ is asymptotically standard normal, hence qualifies as a pseudo-Gaussian test statistic. We thus have shown the following.

Proposition 4.1. *Let Assumptions (B) and (C) hold; assume $h \in \mathcal{F}_{C|\phi_1}$ and that $\hat{\boldsymbol{\vartheta}}^{(n)}$ satisfies Assumptions (D) under $\mathrm{P}_{\boldsymbol{\vartheta},\mathbf{0};g_1,h}^{(n)}$ for any $g_1 \in \mathcal{F}_A^{(4)}$. Then, for any $\boldsymbol{\vartheta} = (\mu, \mathbf{b}, \sigma^2, \mathbf{0})'$,*

(i) $T_{\phi_1}^{(n)\bullet}(\hat{\boldsymbol{\vartheta}}^{(n)}) = T_{\phi_1}^{(n)\bullet}(\boldsymbol{\vartheta}) + o_{\mathrm{P}}(1)$ is asymptotically normal, with mean zero under $\mathrm{P}_{\boldsymbol{\vartheta};f_1}^{(n)}$, mean

$$\begin{aligned}\mu_T^\bullet = \mu_T^\bullet(\mathbf{P}; g_1) &:= \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{j=1}^{n}\left(\|\boldsymbol{x}_j\|^2 - \overline{\|\boldsymbol{x}\|^2}^{(n)}\right)\left(\mathbf{P}\mathbf{x}_j\right)^{\odot 2\prime}}{\sigma^2\sqrt{\left(\mu_4(g_1) - \mu_2^2(g_1)\right) V_{\|\mathbf{x}\|^2}}} \mathbf{K}_{\mathbf{P};2}^{(n)}\boldsymbol{\tau}_4 \\ &= \sigma^2\mu_T^\bullet(\mathbf{P}; g)\end{aligned} \tag{4.2}$$

with $V_{\|\mathbf{x}\|^2} := \lim_{n\to\infty} V^{(n)}_{\|\mathbf{x}\|^2} = \mu^{\|\mathbf{x}\|^4} - \left(\mu^{\|\mathbf{x}\|^2}\right)^2$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},n^{-1/2}\mathbf{K}^{(n)}_{\mathbf{P};2}\boldsymbol{\tau}_4;g_1,h}$ *(along subsequences such that the limit exists), and variance one under both;*

(ii) the sequence of tests rejecting $\bigcup_{\mu\in\mathbb{R}^p}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\bigcup_{g_1\in\mathcal{F}_A^{(4)}}\left\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};g_1}\right\}$ *whenever* $T^{(n)\bullet}_{\phi_1}(\hat{\boldsymbol{\vartheta}}^{(n)})$ *exceeds the* $(1-\alpha)$ *standard normal quantile* z_α *is locally asymptotically directionally maximin at asymptotic level* α *against the Gaussian alternative*

$$\bigcup_{\mu\in\mathbb{R}^p}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\bigcup_{\mathbf{P}\in\boldsymbol{\Pi}_p}\bigcup_{h\in\mathcal{F}_{C|f_1}}\bigcup_{\boldsymbol{\lambda}^2\in\mathcal{C}^+}\left\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};\phi_1,h}\right\}.$$

In view of Proposition 4.1, the test based on $T^{(n)\bullet}_{\phi_1}$, at first sight, looks quite reasonable. Unfortunately, its non-Gaussian local asymptotic powers (especially under skew and heavy-tailed g) can be extremely poor; see Section 9.1 and Tables 9.1 and 9.2 for empirical evidence of this fact.

5. Rank-based tests

Pseudo-Gaussian tests are common practice in semiparametric problems with unspecified error or innovation densities where Gaussian central sequences enjoy the so-called *Fisher consistency property*—remaining centered under a broad class of densities g. Typically, the Gaussian is the only density satisfying, in this context, that property, which implies that it is *least favorable*. A classical alternative to pseudo-Gaussian methods is provided by the semiparametric theory developed, e. g., in Bickel et al. (1993). The tests described there are semiparametrically optimal under any g, but their implementation is rather heavy, as it requires tangent space projection, kernel estimation of g_1, $\dot{g}_1$, and $\ddot{g}_1$, and, in most cases, sample splitting.

Rank tests, as we shall see, offer an intermediate and easily implementable solution. Just as the pseudo-likelihood ones, rank tests are based on the choice of a reference density f_1. The crucial difference is that this reference density now needs not satisfy the Fisher consistency property; it even can be data-driven (Section 8.4), as long as it only depends on the order statistic of residuals—kernel estimation is possible but not required, thus, and sample splitting is useless. Moreover, μ and σ^2 have no impact on residual ranks, whence the notation $Z^{(n)}_j(\mathbf{b})$ for $Z^{(n)}_j(\boldsymbol{\vartheta})$ ($R^{(n)}_j(\mathbf{b})$, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b})$, etc).

Throughout this section, we tacitly assume that f also satisfies Assumption (A.2). Let $R^{(n)}_j = R^{(n)}_j(\mathbf{b})$ be the rank of $Z^{(n)}_j$ among $Z^{(n)}_1(\mathbf{b}),\ldots,Z^{(n)}_n(\mathbf{b})$, and write $\mathbf{R}^{(n)}(\mathbf{b})$ for $\left(R^{(n)}_1(\mathbf{b}),\ldots,R^{(n)}_n(\mathbf{b})\right)$.

Define

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b}) := \begin{pmatrix}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f;2}(\mathbf{b})\\ \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f;4}(\mathbf{b})\end{pmatrix}$$

$$:= \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^{n} \Big[\varphi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) - \overline{\varphi}_f^{(n)} \Big] \mathbf{K}_1^{(n)} \mathbf{x}_j \\ \frac{1}{2} \sum_{j=1}^{n} \Big[\psi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) - \overline{\psi}_f^{(n)} \Big] \mathbf{K}_{\mathbf{P};2}^{(n)} (\mathbf{P}\mathbf{x}_j)^{\odot 2} \end{pmatrix} \tag{5.1}$$

$$= \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^{n} \varphi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) \mathbf{K}_1^{(n)} \Big[\mathbf{x}_j - \bar{\mathbf{x}}^{(n)} \Big] \\ \frac{1}{2} \sum_{j=1}^{n} \psi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) \mathbf{K}_{\mathbf{P};2}^{(n)} \Big[(\mathbf{P}\mathbf{x}_j)^{\odot 2} - \overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)} \Big] \end{pmatrix}$$

$$= \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^{n} \varphi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) \mathbf{K}_1^{(n)} \mathbf{x}_j \\ \frac{1}{2} \sum_{j=1}^{n} \psi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) \mathbf{K}_{\mathbf{P};2}^{(n)} (\mathbf{P}\mathbf{x}_j)^{\odot 2} \end{pmatrix} + o_{\mathrm{P}}(1),$$

with

$$\overline{\varphi}_f^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \varphi_f \Big(F^{-1} \Big(\frac{i}{n+1} \Big) \Big) \quad \text{and} \quad \overline{\psi}_f^{(n)} := \frac{1}{n} \sum_{i=1}^{n} \psi_f \Big(F_1^{-1} \Big(\frac{i}{n+1} \Big) \Big).$$

The last equality in (5.1) follows from the square-integrability of φ_f and ψ_f, which implies (see, e.g., Appendix B2 in Hallin and La Vecchia (2017))

$$\sum_{j=1}^{n} \varphi_f \Big(F^{-1} \Big(\frac{R_j^{(n)}}{n+1} \Big) \Big) = \sum_{j=1}^{n} \varphi_f \Big(F^{-1} \Big(\frac{j}{n+1} \Big) \Big) = o(n^{1/2})$$

(a similar result holds for ψ_f). It also shows that $\underset{\sim}{\boldsymbol{\Delta}}_{\mathbf{P};f}^{(n)}(\mathbf{b})$ actually constitutes the *approximate score* version of the conditional expectations

$$\begin{pmatrix} \mathrm{E}_f \big[\boldsymbol{\Delta}_{\mathbf{P};f_1;2}^{(n)} \big| \mathbf{R}^{(n)}(\mathbf{b}) \big] \\ \mathrm{E}_f \big[\boldsymbol{\Delta}_{\mathbf{P};f_1;4}^{(n)} \big| \mathbf{R}^{(n)}(\mathbf{b}) \big] \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{1}{\sigma} \sum_{j=1}^{n} \mathrm{E}_f \Big[\varphi_{f_1} \big(Z_j^{(n)}(\mathbf{b}) \big) \Big| \mathbf{R}^{(n)}(\mathbf{b}) \Big] \\ \qquad \times \mathbf{K}_1^{(n)} \mathbf{x}_j \\ \frac{1}{2\sigma^2} \sum_{j=1}^{n} \mathrm{E}_f \Big[\psi_{f_1} \big(Z_j^{(n)}(\mathbf{b}) \big) \Big| \mathbf{R}^{(n)}(\mathbf{b}) \Big] \\ \qquad \times \mathbf{K}_{\mathbf{P};2}^{(n)} (\mathbf{P}\mathbf{x}_j)^{\odot 2} \end{pmatrix} \tag{5.2}$$

of the $\mathbf{b}$ and $\boldsymbol{\lambda}^2$ blocks of the central sequence $\boldsymbol{\Delta}_{\mathbf{P};f_1}^{(n)}(\boldsymbol{\vartheta}) = \boldsymbol{\Delta}_{\mathbf{P};f_1}^{(n)}(\mu, \mathbf{b}, \sigma^2)$ conditional on $\mathbf{R}^{(n)}(\mathbf{b})$ (which constitute the *exact score* version of the same). The same expectations, for the μ and σ^2 blocks, is zero, hence safely can be omitted.

Classical asymptotic representation results for linear rank statistics (see, e.g., Chapter V of Hájek and Šidák (1967), or the very clear exposition by Lombard

(1986)) entail, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$, the asymptotic equivalence of the exact score version (5.2), the approximate score version (5.1), and

$$\underline{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f,g}(\mathbf{b}) := \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^{n} \varphi_f\big(F^{-1}\circ G_1\big(Z^{(n)}_j(\mathbf{b})\big)\big)\mathbf{K}^{(n)}_1\Big[\mathbf{x}_j - \bar{\mathbf{x}}^{(n)}\Big] \\ \frac{1}{2}\sum_{j=1}^{n} \psi_f\big(F^{-1}\circ G_1\big(Z^{(n)}_j(\mathbf{b})\big)\big)\mathbf{K}^{(n)}_{\mathbf{P};2}\Big[(\mathbf{P}\mathbf{x}_j)^{\odot 2} - \overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)}\Big] \end{pmatrix} \tag{5.3}$$

(F and G_1 the distribution functions associated with f and g_1, respectively).

Now, Hallin and Werker (2003) establish the asymptotic equivalence, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$, of the conditional expectation (5.2) and the *semiparametrically efficient central sequence* at $(\boldsymbol{\vartheta},\mathbf{0})$ and f (namely, the tangent space projection of $\big(\boldsymbol{\Delta}^{(n)\prime}_{\mathbf{P};f;2}(\boldsymbol{\vartheta}), \boldsymbol{\Delta}^{(n)\prime}_{\mathbf{P};f;4}(\boldsymbol{\vartheta})\big)'$); see their Example 4.2 for the case of traditional regression. It follows that both $\underline{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f,f}(\mathbf{b})$ and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b})$ are versions of that semiparametrically efficient central sequence; being rank-based, however, the latter is distribution-free. The first and third components of $\boldsymbol{\Delta}^{(n)}_{\mathbf{P};f}(\boldsymbol{\vartheta})$ are not taken into account since their conditional expectations, as explained, are zero.

Note that the $\mathbf{b}$-components of $\underline{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f,f}(\mathbf{b})$ and $\boldsymbol{\Delta}^{(n)\prime}_{\mathbf{P};f}(\boldsymbol{\vartheta})$ asymptotically coincide iff $\bar{\mathbf{x}}^{(n)}$ is $o(1)$. But, since $\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)}$ under Assumption (B.1) cannot be $o(1)$, the same equivalence is impossible for the $\boldsymbol{\lambda}^2$-components: the detection problem we are considering, thus, is nonadaptive with respect to the non-specification of f, irrespective of the regression design.

Under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$, the $F^{-1}\circ G_1\big(Z^{(n)}_j(\mathbf{b})\big)$'s are i.i.d. with density f. This implies that $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b})$ is asymptotically normal, with mean zero and covariance (limits in $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$-probability)

$$\underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f} := \begin{pmatrix} \underset{\sim}{\boldsymbol{\Gamma}}_{f;22} & \underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f;24} \\ \underset{\sim}{\boldsymbol{\Gamma}}'_{\mathbf{P};f;24} & \underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f;44} \end{pmatrix} = \lim_{n\to\infty} \begin{pmatrix} \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f;22} & \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f;24} \\ \underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f;24} & \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f;44} \end{pmatrix} =: \lim_{n\to\infty} \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f} \tag{5.4}$$

where $\underset{\sim}{\boldsymbol{\Gamma}}_{f;22} := \mathcal{I}_{\varphi}(f)\Big(\mathbf{I}_{p\times p} - \mathbf{K}_1\boldsymbol{\mu}^{\mathbf{x}}\boldsymbol{\mu}^{\mathbf{x}\prime}\mathbf{K}_1\Big) = \lim_{n\to\infty}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f;22}$,

$$\underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f;24} := \frac{\mathcal{I}_{\phi\psi}(f)}{2}\mathbf{K}_1\Big(\mathbf{C}_{\mathbf{P};3} - \boldsymbol{\mu}^{\mathbf{x}}\boldsymbol{\mu}^{(\mathbf{P}\mathbf{x})^{\odot 2}\prime}\Big)\mathbf{K}_{\mathbf{P};2} = \lim_{n\to\infty}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f;24},$$

and $\underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f;44} := \frac{\mathcal{I}_{\psi}(f)}{4}\Big(\mathbf{I}_{p\times p} - \mathbf{K}_{\mathbf{P};2}\boldsymbol{\mu}^{(\mathbf{P}\mathbf{x})^{\odot 2}}\boldsymbol{\mu}^{(\mathbf{P}\mathbf{x})^{\odot 2}\prime}\mathbf{K}_{\mathbf{P};2}\Big) = \lim_{n\to\infty}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f;44}$; the empirical counterpart $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f}$ of $\underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f}$ (which cannot be computed from the sample) is obtained by replacing, in an obvious fashion, $\mathbf{K}_1$ and $\mathbf{K}_{\mathbf{P};2}$ with $\mathbf{K}^{(n)}_1$ and $\mathbf{K}^{(n)}_{\mathbf{P};2}$, $\mathbf{C}_{\mathbf{P};3}$ with $\mathbf{C}^{(n)}_{\mathbf{P};3}$, $\boldsymbol{\mu}^{\mathbf{x}}$ and $\boldsymbol{\mu}^{(\mathbf{P}\mathbf{x})^{\odot 2}}$ with $\bar{\mathbf{x}}^{(n)}$ and $\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)}$, respectively.

5.1. *The rank-based test statistic: specified* P

To start with, fix some value for $\mathbf{P}$. Let f be some chosen reference density (under which optimality is to be achieved), the actual density being g. The qualification of $\underset{\sim}{\mathbf{\Delta}}_f^{(n)}(\mathbf{b})$ as a semiparametrically efficient central sequence implies that $\underset{\sim}{\mathbf{\Gamma}}_{\mathbf{P};f}$ is the information matrix settling the values of semiparametric efficiency bounds in the contiguous vicinity of $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$, while testing procedures reaching those bounds are obtained, *mutatis mutandis*, by treating $\underset{\sim}{\mathbf{\Delta}}_f^{(n)}$ as we did $\mathbf{\Delta}_{f_1}^{(n)*}$ in Section 3. Accordingly, a test achieving local and asymptotic optimality under density f and for given $\mathbf{P}$, in the presence of an unspecified $\mathbf{b}$, is to be based on $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\hat{\mathbf{b}}^{(n)})$, where

(a) $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\mathbf{b})$ denotes the residual of the regression of $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f;4}^{(n)}(\mathbf{b})$ with respect to $\underset{\sim}{\mathbf{\Delta}}_{f;2}^{(n)}(\mathbf{b})$ in the covariance structure $\underset{\sim}{\mathbf{\Gamma}}_{\mathbf{P};f}$ given in (5.4) (equivalently, its consistent empirical counterpart $\underset{\sim}{\mathbf{\Gamma}}_{\mathbf{P};f}^{(n)}$), and

(b) $\hat{\mathbf{b}}^{(n)}$is an estimator of $\mathbf{b}$ satisfying (with $\boldsymbol{\nu}^{(n)}=\mathbf{K}_1^{(n)}$) Assumptions (D).

Since $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\mathbf{b})$, by construction, is asymptotically uncorrelated, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$, with $\underset{\sim}{\mathbf{\Delta}}_{f;2}^{(n)}(\mathbf{b})$, it is asymptotically insensitive, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$ and contiguous alternatives, to the replacement of $\mathbf{b}$ with $\hat{\mathbf{b}}^{(n)}$—a particular case of Lemma 8.1. Thus, $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\hat{\mathbf{b}}^{(n)})$, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$ and contiguous alternatives, is asymptotically equivalent to $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\mathbf{b})$ and enjoys, under density f, the same semiparametric optimality properties.

While taking care of the asymptotic validity and optimality under density f of tests based on $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\hat{\mathbf{b}}^{(n)})$, this does not entail their validity under density $g\neq f$. Therefore, rather than $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\hat{\mathbf{b}}^{(n)})$, consider $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)*}(\hat{\mathbf{b}}^{(n)})$, where

$$\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)*}(\mathbf{b}) := \underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f;4}^{(n)}(\mathbf{b}) - \underset{\sim}{\widehat{\mathbf{\Gamma}}}_{\mathbf{P};f;24}^{(n)\prime}\big(\underset{\sim}{\widehat{\mathbf{\Gamma}}}_{f;22}^{(n)}\big)^{-1}\underset{\sim}{\mathbf{\Delta}}_{f;2}^{(n)}(\mathbf{b})$$

results from substituting the estimators

$$\begin{aligned}\underset{\sim}{\widehat{\mathbf{\Gamma}}}_{f;22}^{(n)} &:= \frac{1}{n}\mathcal{I}_{\varphi}^{(n)}(f)\mathbf{K}_1^{(n)}\sum_{j=1}^{n}\Big[\mathbf{x}_j-\bar{\mathbf{x}}^{(n)}\Big]\Big[\mathbf{x}_j-\bar{\mathbf{x}}^{(n)}\Big]'\mathbf{K}_1^{(n)}\\ &= \mathcal{I}_{\varphi}^{(n)}(f)\big[\mathbf{I}_p-\mathbf{K}_1^{(n)}\bar{\mathbf{x}}^{(n)}\bar{\mathbf{x}}^{(n)\prime}\mathbf{K}_1^{(n)}\big]\end{aligned}\tag{5.5}$$

and

$$\begin{aligned}\underset{\sim}{\widehat{\mathbf{\Gamma}}}_{\mathbf{P};f;24}^{(n)} &:= \frac{1}{2n}\mathcal{I}_{\psi\varphi}^{(n)}(f)\mathbf{K}_1^{(n)}\sum_{j=1}^{n}\Big[\mathbf{x}_j-\bar{\mathbf{x}}^{(n)}\Big]\Big[(\mathbf{P}\mathbf{x}_j)^{\odot 2}-\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)}\Big]'\mathbf{K}_{\mathbf{P};2}^{(n)}\\ &= \frac{1}{2}\mathcal{I}_{\psi\varphi}^{(n)}(f)\mathbf{K}_1^{(n)}\big[\mathbf{C}_{\mathbf{P};3}^{(n)}-\bar{\mathbf{x}}^{(n)}\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)\prime}\big]\mathbf{K}_{\mathbf{P};2}^{(n)},\end{aligned}\tag{5.6}$$

for $\underset{\sim}{\mathbf{\Gamma}}_{f;22}^{(n)}$ and $\underset{\sim}{\mathbf{\Gamma}}_{\mathbf{P};f;24}^{(n)}$, in the definition of $\underset{\sim}{\mathbf{\Delta}}_{\mathbf{P};f}^{(n)\dagger}(\mathbf{b})$; $\mathcal{I}_{\varphi}^{(n)}(f)$ and $\mathcal{I}_{\psi\varphi}^{(n)}(f)$ in (5.5) and (5.6) denote consistent, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$, estimators of

$$\mathcal{I}_{\varphi}(f,g) := \int_0^1 \varphi_f(F^{-1}(u))\varphi_g(G^{-1}(u))du$$

and

$$\mathcal{I}_{\psi\varphi}(f,g) := \int_0^1 \psi_f(F^{-1}(u))\varphi_g(G^{-1}(u))du,$$

respectively; such estimators are provided in Section 8.2.

Since, for $g = f$, $\mathcal{I}_\varphi(f,f) = \mathcal{I}_\varphi(f)$ and $\mathcal{I}_{\psi\varphi}(f,f) = \mathcal{I}_{\psi\varphi}(f)$, the difference between $\widehat{\underset{\sim}{\boldsymbol{\Gamma}}}^{(n)}_{\mathbf{P};f;ij}$ and $\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{\mathbf{P};f;ij}$ is $o_{\mathrm{P}}(1)$, under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$, for $(i,j) = (2,2)$ and $(2,4)$. Hence, still under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)\dagger}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})$ too is $o_{\mathrm{P}}(1)$ (hence also under contiguous alternatives); $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})$ thus retains, under density f, the optimality properties of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)\dagger}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})$ and $\underset{\sim}{\boldsymbol{\Delta}}^{(n)\dagger}_{\mathbf{P};f}(\mathbf{b})$.

More generally, $\widehat{\underset{\sim}{\boldsymbol{\Gamma}}}^{(n)}_{f;22}$ and $\widehat{\underset{\sim}{\boldsymbol{\Gamma}}}^{(n)}_{\mathbf{P};f,;24}$ are converging, in $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$-probability, to

$$\underset{\sim}{\boldsymbol{\Gamma}}_{f,g;22} := \mathcal{I}_\varphi(f,g)\Big(\mathbf{I}_{p\times p} - \mathbf{K}_1\boldsymbol{\mu}^{\mathbf{x}}\boldsymbol{\mu}^{\mathbf{x}\prime}\mathbf{K}_1\Big)$$

and

$$\underset{\sim}{\boldsymbol{\Gamma}}_{\mathbf{P};f,g;24} := \frac{1}{2}\mathcal{I}_{\varphi\psi}(f,g)\mathbf{K}_1\Big(\mathbf{C}_{\mathbf{P};3} - \boldsymbol{\mu}^{\mathbf{x}}\boldsymbol{\mu}^{(\mathbf{P}\mathbf{x})^{\odot 2}\prime}\Big)\mathbf{K}_{\mathbf{P};2},$$

respectively. This (see Lemma 8.1) implies that the difference

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\mathbf{b})$$

is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$—not just under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$. Hence, irrespective of $\mathbf{b}$ and g, substituting $\hat{\mathbf{b}}^{(n)}$ for $\mathbf{b}$ has no asymptotic impact on $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\mathbf{b})$. Therefore, tests reaching, for specified $\mathbf{P}$ and density f, asymptotic optimality against given directions $\mathbf{u}$ in the $\boldsymbol{\lambda}^2$ space (see Section 3.3) while remaining asymptotically valid under any density g can be based on adequate projections of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})$. Those tests play the role of directionally (i.e., for some given choice of $\mathbf{P}$ and $\mathbf{u}$) optimal tests.

5.2. The rank-based test statistic: unspecified P

Now, just as in Section 3.3, $\mathbf{P}$ and $\mathbf{u}$ are to be selected in order to obtain a directionally maximin test. The same reasoning as in the proof of Proposition 3.1 leads to projecting $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})$ onto a unit vector $\mathbf{u}$ such that $\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u} = \mathbf{1}_p$. Letting

$$\mathbf{W}^{(n)}_{\|\mathbf{x}\|^2;\mathbf{x}} := \frac{1}{n}\sum_{j=1}^{n}\Big[\|\boldsymbol{x}_j\|^2 - \frac{1}{n}\sum_{l=1}^{n}\|\boldsymbol{x}_l\|^2\Big]\mathbf{x}_j$$

and

$$\mathbf{W}^{(n)}_{\|\mathbf{x}\|^2;(\mathbf{P}\mathbf{x})^{\odot 2}} := \frac{1}{n}\sum_{j=1}^{n}\Big[\|\boldsymbol{x}_j\|^2 - \frac{1}{n}\sum_{l=1}^{n}\|\boldsymbol{x}_l\|^2\Big]\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2},$$

this yields, after some easy algebra,

$$\begin{aligned}\mathbf{u}'\underset{\sim}{\mathbf{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) =& \frac{1}{2\sqrt{n}}\sum_{j=1}^{n}\psi_f\Big(F^{-1}\Big(\frac{R_j^{(n)}}{n+1}\Big)\Big)\Big(\|\boldsymbol{x}_j\|^2-\frac{1}{n}\sum_{l=1}^{n}\|\boldsymbol{x}_l\|^2\Big)\\ &-\frac{\mathcal{I}^{(n)}_{\psi\varphi}(f)}{2\sqrt{n}\mathcal{I}^{(n)}_{\varphi}(f)}\mathbf{W}^{(n)\prime}_{\|\mathbf{x}\|^2;\mathbf{x}}\Big[\mathbf{C}^{(n)}_1-\bar{\mathbf{x}}^{(n)}\bar{\mathbf{x}}^{(n)\prime}\Big]^{-1}\\ &\times\sum_{j=1}^{n}\varphi_f\Big(F^{-1}\Big(\frac{R_j^{(n)}}{n+1}\Big)\Big)\Big[\mathbf{x}_j-\bar{\mathbf{x}}^{(n)}\Big]\\ =:&\ \underset{\sim}{T}^{(n)*}_{\mathbf{1}_p;f}(\hat{\mathbf{b}}^{(n)}),\end{aligned}$$

which no longer depends on $\mathbf{P}$ and has asymptotic variance $\lim_{n\to\infty}\underset{\sim}{\Gamma}^{(n)*}_{f;\mathbf{1}_p}$, with

$$\begin{aligned}\underset{\sim}{\Gamma}^{(n)*}_{\mathbf{1}_p;f}:=&\frac{\mathcal{I}_\psi(f)}{4}V^{(n)}_{\|\mathbf{x}\|^2}-\frac{\mathcal{I}^{(n)}_{\psi\varphi}(f)}{4\mathcal{I}^{(n)}_{\varphi}(f)\big)^2}\big(2\mathcal{I}^{(n)}_{\varphi}(f)\mathcal{I}_{\varphi\psi}(f)-\mathcal{I}^{(n)}_{\psi\varphi}(f)\mathcal{I}_{\varphi}(f)\big)\\ &\times\mathbf{W}^{(n)\prime}_{\|\mathbf{x}\|^2;\mathbf{x}}\Big[\mathbf{C}^{(n)}_1-\bar{\mathbf{x}}^{(n)}\bar{\mathbf{x}}^{(n)\prime}\Big]^{-1}\mathbf{W}^{(n)}_{\|\mathbf{x}\|^2;\mathbf{x}}.\end{aligned}$$

The test statistic we are proposing is the standardized version

$$\underset{\sim}{T}^{(n)}_{f}(\hat{\mathbf{b}}^{(n)}):=\Big(\underset{\sim}{\Gamma}^{(n)*}_{\mathbf{1}_p;f}\Big)^{-1/2}\underset{\sim}{T}^{(n)*}_{\mathbf{1}_p;f}(\hat{\mathbf{b}}^{(n)}) \tag{5.7}$$

of $\underset{\sim}{T}^{(n)*}_{\mathbf{1}_p;f}(\hat{\boldsymbol{\beta}}^{(n)})$. Summing up, we established the following results (see Section 8.3 for a proof).

Proposition 5.1. *Let Assumptions (B) and (C) hold, denote by $\hat{\mathbf{b}}^{(n)}$ an estimator satisfying Assumptions (D), and fix f such that $f_1\in\mathcal{F}_A$. Then,*

(i) $\underset{\sim}{T}^{(n)}_{f}(\hat{\mathbf{b}}^{(n)})=\underset{\sim}{T}^{(n)}_{f}(\mathbf{b})+o_{\mathrm{P}}(1)$ *is asymptotically normal, with mean zero under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$, *mean*

$$\begin{aligned}\underset{\sim}{\mu}^*_{\mathbf{P};f,g}:=\ &\underset{\sim}{\Gamma}^{*-1/2}_{\mathbf{1}_p;f,g}\Big\{\frac{\mathcal{I}_\psi(f,g)}{4}\lim_{n\to\infty}\mathbf{W}^{(n)}_{\|\mathbf{x}\|^2;(\mathbf{Px})^{\odot 2}}-\frac{\mathcal{I}_{\varphi\psi}(f,g)\mathcal{I}_{\psi\varphi}(f,g)}{4\mathcal{I}_{\varphi}(f,g)}\\ &\times\lim_{n\to\infty}\mathbf{W}^{(n)\prime}_{\|\mathbf{x}\|^2;\mathbf{x}}\Big[\mathbf{C}_1-\boldsymbol{\mu}^{\mathbf{x}}\boldsymbol{\mu}^{\mathbf{x}\prime}\Big]^{-1}\Big[\mathbf{C}_{\mathbf{P};3}-\boldsymbol{\mu}^{\mathbf{x}}\boldsymbol{\mu}^{(\mathbf{Px})^{\odot 2}\prime}\Big]\Big\}\mathbf{K}_{\mathbf{P};2}\boldsymbol{\tau}_4\end{aligned}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},n^{-1/2}\mathbf{K}^{(n)}_{\mathbf{P};2}\boldsymbol{\tau}_4;g,h}$ *(along subsequences such that* $\lim_{n\to\infty}\mathbf{W}^{(n)}_{\|\mathbf{x}\|^2;\mathbf{x}}$ *and* $\lim_{n\to\infty}\mathbf{W}^{(n)}_{\|\mathbf{x}\|^2;(\mathbf{Px})^{\odot 2}}$ *exist), and variance one under both;*

(ii) the sequence of tests rejecting $\bigcup_{\mu\in\mathbb{R}}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\bigcup_{g}\big\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};g}\big\}$ *whenever the test statistic* $\underset{\sim}{T}^{(n)}_{f}(\hat{\mathbf{b}}^{(n)})$ *exceeds the* $(1-\alpha)$ *standard normal quantile* z_α *is locally asymptotically directionally maximin at asymptotic level* α *against alternatives of the form*

$$\bigcup_{\mu\in\mathbb{R}}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\bigcup_{\mathbf{P}\in\mathbf{\Pi}_p}\bigcup_{h\in\mathcal{F}_{C|f_1}}\bigcup_{\boldsymbol{\lambda}^2\in\mathcal{C}^+}\Big\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f,h}\Big\}.$$

6. Asymptotic relative efficiencies (AREs)

The asymptotic relative efficiencies of the rank-based tests developed in the previous section, with respect to the pseudo-Gaussian tests of Section 4 are easily obtained as the ratios of the standardized asymptotic shifts $\underset{\sim}{\mu}^*_{\mathbf{P};f,g}$ and $\mu^\bullet_T = \mu^\bullet_T(\mathbf{P};g_1)$ obtained in Propositions 4.1 and 5.1, respectively. Those shifts, unfortunately, depend on the unspecified $\mathbf{P}$, and the perturbation $\boldsymbol{\tau}_4$ does not cancel out when taking ratios—unless $\mathbf{P}$ and $\boldsymbol{\tau}_4$ are such that $\mathbf{K}_{\mathbf{P};2}\boldsymbol{\tau}_4 = \mathbf{1}_p$. Table 6.1 is listing, for various scores and under various densities, some of those ARE values.

Inspection of Table 6.1 reveals the dramatic gains achieved by considering ranks in this context. The van der Waerden test which, under Gaussian alternatives, is asymptotically equivalent to the pseudo-Gaussian one, reaches a huge ARE of 360% under Student densities with 5 degrees of freedom! ARE values as high as 569% are attained under Student t_5 by the Wilcoxon test—meaning that the pseudo-Gaussian test requires five times more observations than the Wilcoxon test in order to achieve the same large-sample performance.

All AREs in the van der Waerden row of the table are larger than one. This might be the empirical indication that the celebrated Chernoff-Savage property[3] holds for this problem. Due to the complicated form of the asymptotic shifts $\underset{\sim}{\mu}^*_{\mathbf{P};f,g}$ and $\mu^\bullet_T = \mu^\bullet_T(\mathbf{P};g_1)$ in Propositions 4.1 and 5.1, however, a theoretical confirmation is hard to obtain, and we were not able to prove nor disprove the property. Whether the Chernoff-Savage property holds in this context, thus, remains unknown.

7. Proofs for Sections 2 and 3

7.1. Proof of Proposition 2.1

The result is obtained by checking that the six conditions of Lemma 2 in Swensen (1985) are satisfied. As usual, the only delicate one is the quadratic mean differentiability of the square root of the density, computed at the residual—here, the quadratic mean differentiability, at any $(\mu, \mathbf{b}, \sigma^2, \mathbf{0})$ and for all $(y, \mathbf{x})$ in $\mathbb{R} \times \mathbb{R}^p$, of $(\mu, \mathbf{b}, \sigma^2, \boldsymbol{\lambda}^2) \longmapsto q^{1/2}_{\mu,\mathbf{b},\sigma^2,\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}(y)$, where

$$q_{\mu,\mathbf{b},\sigma^2,\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}(y) := \frac{1}{\sigma}\int_{\mathbb{R}^p} f_1\Big(\frac{1}{\sigma}\Big(y - \mu - \sum_{i=1}^p b_i x_i - \sum_{i=1}^p \big(\mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}\,\boldsymbol{\xi}\big)_i x_i\Big)\Big) h(\boldsymbol{\xi}) d\boldsymbol{\xi}.$$

[3]Recall that Chernoff and Savage (1958) established the surprising fact that, in (univariate) two-sample location models, the ARE of the van der Waerden (i.e., normal-score) rank test with respect to its Gaussian competitor (the Student test) is strictly larger than one under any density but the Gaussian one under which, of course, it equals one. That property immediately extends to linear models (regression, ANOVA, etc.). Further and less obvious extensions were obtained later on to ARMA models (Hallin, 1994; Hallin and Tribel, 2000), then to elliptical location and VARMA models (Hallin and Paindaveine, 2002a,b). The same property was established in Paindaveine (2006) for elliptical rank tests of shape and in Hallin and Paindaveine (2008) for the the van der Waerden version of Wilks' test of independence between random vectors.

Table 6.1

AREs, under Student ($\nu = 5, 8,$ and 10 degrees of freedom), logistic (ℓ_1), normal (ϕ_1) and skew-normal ($s\mathcal{N}(\delta)$, $\delta = 2, 3$) densities, for alternatives such that $\mathbf{K}_{\mathbf{P};2}\boldsymbol{\tau}_4 = \mathbf{1}_p$, of various rank tests based on Student, Wilcoxon, van der Waerden, and skew-normal scores, with respect to the pseudo-Gaussian test.

g_1 / score f_1	t_5	t_8	t_{10}	ℓ_1	ϕ_1	$s\mathcal{N}(2)$	$s\mathcal{N}(3)$
t_5	5. 8333	1.7818	1.3749	1.6906	0.6187	0.7711	0.8287
t_8	5.5878	1.8601	1.4836	1.7178	0.7858	1.0243	1.2753
t_{10}	5.3709	1.8482	1.4932	1.6934	0.8440	1.1184	1.4603
ℓ_1 (Wilcoxon)	5.6947	1.8452	1.4601	1.7317	0.7599	0.9853	1.1932
ϕ_1 (van der Waerden)	3.6089	1.4617	1.2603	1.3159	1.0000	1.4278	2.2396
$s\mathcal{N}(2)$	2.5411	1.0765	0.9435	0.9640	0.8068	1.7698	3.4093
$s\mathcal{N}(3)$	1.3393	0.6573	0.6042	0.5725	0.6207	1.6721	3.6088

In view of Theorem 2.1 in Lind and Roussas (1972) (independently rediscovered by Garel and Hallin (1995), Lemma 2.1), the existence of partial derivatives in quadratic mean is sufficient for quadratic mean differentiability to hold. Therefore, quadratic mean differentiability for arbitrary p follows from quadratic mean differentiability for $p = 1$—a result which is established in Lemma A.1 of Fihri et al. (2020), where we refer to for details. Proposition 2.1 follows. □

7.2. *Proof of Proposition 3.1*

The main difficulty in the proof of Proposition 3.1 is with the role of the matrix $\mathbf{P}$, which is not identified, hence cannot be estimated under the null hypothesis $\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};f_1}\}$ (nor can h; but the latter, as we shall see, does not play any role).

In order to obtain (locally asymptotically) *directionally most powerful* tests, a one-dimensional subhypothesis has to be selected within the alternative

$$\bigcup_{\boldsymbol{\lambda}^2,\,\mathbf{P}} \left\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}\right\}$$

by specifying a value of $\mathbf{P}$ and a direction $\mathbf{u}\ \boldsymbol{\lambda}^2/\|\boldsymbol{\lambda}^2\|$. A locally asymptotically (uniformly) most powerful test against this sub-alternative consists in rejecting whenever

$$T^{(n)}_{\mathbf{P},\mathbf{u};f_1}(\boldsymbol{\vartheta}) := \Big(n\mathcal{I}_\psi(f_1)\Big)^{-1/2}\Big(\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u}\Big)'\sum_{j=1}^{n}\psi_{f_1}(Z_j)\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2} \tag{7.1}$$

(with $\boldsymbol{\vartheta} = (\mu, \mathbf{b}, \sigma^2)$) exceeds the $(1-\alpha)$ standard normal quantile z_α.

It follows from Le Cam's Third Lemma that the asymptotic power, under alternatives of the form $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},n^{-1/2}\mathbf{K}^{(n)}_{\mathbf{Q};2}\boldsymbol{\tau}_4,\mathbf{P};f_1,h}$ and along appropriate subsequences (such that the limits exist), of the directionally most powerful test based on (7.1) is

$$1-\Phi\Big(z_\alpha-\lim_{n\to\infty}(2n\sigma^2)^{-1/2}\mathcal{I}_\psi^{1/2}(f_1)\mathbf{u}'\mathbf{K}^{(n)}_{\mathbf{P};2}\sum_{j=1}^n\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\big(\mathbf{Q}\mathbf{x}_j\big)^{\odot 2\prime}\mathbf{K}^{(n)}_{\mathbf{Q};2}\boldsymbol{\tau}_4\Big).$$

For $\boldsymbol{\tau}_4=\tau\mathbf{v}$, $\tau\geq 0$ and $\mathbf{v}\in\mathcal{S}_{\mathbf{p-1}}$, the derivative with respect to τ of that power, at $\tau=0$, is (up to a positive multiplicative constant) the limit of

$$\mathbf{u}'\mathbf{K}^{(n)}_{\mathbf{P};2}\sum_{j=1}^n\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\big(\mathbf{Q}\mathbf{x}_j\big)^{\odot 2\prime}\mathbf{K}^{(n)}_{\mathbf{Q};2}\mathbf{v}, \tag{7.2}$$

that is, the limit of the scalar product $\langle\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u},\sum_{j=1}^n\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\big(\mathbf{Q}\mathbf{x}_j\big)^{\odot 2\prime}\mathbf{K}^{(n)}_{\mathbf{Q};2}\mathbf{v}\rangle$. Note that $\mathbf{K}^{(n)}_{\mathbf{P};2}\mathcal{C}^+=\mathbf{K}^{(n)}_{\mathbf{Q};2}\mathcal{C}^+=\mathcal{C}^+$; similarly, for any choice of $\mathbf{P}$, we have

$$\sum_{j=1}^n\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\big(\mathbf{Q}\mathbf{x}_j\big)^{\odot 2\prime}\mathcal{C}^+=\mathcal{C}^+.$$

For given n, $\mathbf{u}$, $\mathbf{P}$, $\mathbf{Q}$, and τ, the moduli of $\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u}$ and

$$\sum_{j=1}^n\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\big(\mathbf{Q}\mathbf{x}_j\big)^{\odot 2\prime}\mathbf{K}^{(n)}_{\mathbf{Q};2}\mathbf{v}$$

(both ranging over $\mathcal{C}^+$) are constant, so that (7.2) is minimal when the angle between them is maximal. Irrespective of $\mathbf{Q}$, the orthogonal matrices $\mathbf{P}$ and the unit vectors $\mathbf{u}$ minimizing that maximal angle are those for which $\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u}=\mathbf{1}_p$, where $\mathbf{1}_p:=(1,\ldots,1)'$ (an infinite number of solutions). Although the values of (7.2) at those solutions depend on $\mathbf{Q}$, the resulting tests do not, with a test statistic of the form

$$\frac{\mathbf{1}_p'}{\sqrt{n\mathcal{I}_\psi(f_1)}}\sum_{j=1}^n\psi_{f_1}(Z_j)\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}=\frac{1}{\sqrt{n\mathcal{I}_\psi(f_1)}}\sum_{j=1}^n\psi_{f_1}(Z_j)\mathbf{1}_p'\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}. \tag{7.3}$$

Now, since $\mathbf{P}$ is orthogonal, $\mathbf{1}_p'\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}=\|\mathbf{P}\mathbf{x}_j\|^2=\|\mathbf{x}_j\|^2$, and (7.3) thus reduces, after standardization, to the test statistic $T^{(n)}_{f_1}(\boldsymbol{\vartheta})$ defined in (3.3), which no longer depends on $\mathbf{P}$. This establishes part *(iii)* of the Proposition; parts *(i)* and *(ii)* then are straightforward. □

7.3. Proof of Proposition 3.2

Before turning to the proof of Proposition 3.2, let us recall some classical facts on rate-optimal locally asymptotically discrete estimators—a classical concept when dealing with the estimation of nuisance parameters in ULAN families. Denoting by $\hat{\boldsymbol{\eta}}^{(n)}$ a sequence of estimators of $\boldsymbol{\eta}$ in a sequence of experiments $\{\mathrm{P}^{(n)}_{\boldsymbol{\eta}}\mid\boldsymbol{\eta}\in\boldsymbol{\Theta}\}$ with contiguity rates $n^{-1/2}(\boldsymbol{\nu}^{(n)})$, recall that $\hat{\boldsymbol{\eta}}^{(n)}$ is called $n^{1/2}(\boldsymbol{\nu}^{(n)})^{-1}$-*consistent* and *locally asymptotically discrete* if it satisfies the following assumptions.

Assumption (D) Under $\mathrm{P}_{\boldsymbol{\eta}}^{(n)}$, as $n \to \infty$,

(D.1) $(\boldsymbol{\nu}^{(n)})^{-1}\big(\hat{\boldsymbol{\eta}}^{(n)} - \boldsymbol{\eta}\big) = O_{\mathrm{P}}(n^{-1/2})$;

(D.2) the number of possible values of $\hat{\boldsymbol{\eta}}^{(n)}$ in balls with $O(n^{-1/2}\boldsymbol{\nu}^{(n)})$ radius centered at $\boldsymbol{\eta}$ is bounded as $n \to \infty$.

Assumption (D.2) is quite mild, as any estimator $\hat{\boldsymbol{\eta}}^{(n)} = (\hat{\boldsymbol{\eta}}_1^{(n)}, \ldots, \hat{\boldsymbol{\eta}}_q^{(n)})'$ satisfying (D.1) can be discretized into

$$\hat{\boldsymbol{\eta}}_{\#}^{(n)} := \boldsymbol{\nu}^{(n)}(cn^{1/2})^{-1} \begin{pmatrix} \mathrm{sign}(\hat{\boldsymbol{\eta}}_1^{(n)})\lceil cn^{1/2}\|\hat{\boldsymbol{\eta}}_1^{(n)}\|\rceil \\ \vdots \\ \mathrm{sign}(\hat{\boldsymbol{\eta}}_q^{(n)})\lceil cn^{1/2}\|\hat{\boldsymbol{\eta}}_q^{(n)}\|\rceil \end{pmatrix}$$

satisfying both (D.1) and (D.2) ($c > 0$ is some arbitrary constant tuning the discretization; $\lceil\|\boldsymbol{\eta}\|\rceil$, as usual, stands for the smallest integer larger than or equal to $\|\boldsymbol{\eta}\|$). Such discretization is needed in the statement of asymptotic results, although it has no practical consequences for fixed n (as c can be chosen arbitrarily large). Note that $\boldsymbol{\nu}^{(n)}$ being $O(1)$ as $n \to \infty$, it safely can be omitted in (D.1) and (D.2).

The problem of testing the null hypothesis $\bigcup_{\mu\in\mathbb{R}}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\big\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\mathbf{0};f_1}\big\}$ against the alternative

$$\bigcup_{\mu\in\mathbb{R}}\bigcup_{\mathbf{b}\in\mathbb{R}^p}\bigcup_{\sigma^2\in\mathbb{R}^+}\bigcup_{\mathbf{P}\in\boldsymbol{\Pi}_p}\bigcup_{h\in\mathcal{F}_{C|f_1}}\bigcup_{\boldsymbol{\lambda}^2\in\mathcal{C}^+}\Big\{\mathrm{P}^{(n)}_{(\mu,\mathbf{b},\sigma^2),\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}\Big\}$$

is thus a classical problem of testing linear restrictions on the parameter $(\boldsymbol{\vartheta}, \boldsymbol{\lambda}^2)$ under ULAN. Locally and asymptotically optimal (at $(\boldsymbol{\vartheta}, \mathbf{0})$) inference should be based (see Chapter 11.9 in Le Cam (1986)) on the residual

$$\boldsymbol{\Delta}^{(n)*}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta}) = \boldsymbol{\Delta}^{(n)}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta}) - \frac{1}{2\sigma^3}\begin{pmatrix} \mathcal{I}_{\varphi\psi}(f_1)\mathbf{K}^{(n)}_{\mathbf{P};2}\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)} \\ \frac{\mathcal{I}_{\varphi\psi}(f_1)}{n}\mathbf{K}^{(n)}_{\mathbf{P};2}\sum_{j=1}^n\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\mathbf{x}_j'\mathbf{K}_1^{(n)} \\ \frac{1}{2\sigma}\mathcal{K}_{\varphi\psi}(f_1)\mathbf{K}^{(n)}_{\mathbf{P};2}\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)} \end{pmatrix}'$$
$$\times\begin{pmatrix} \Gamma^{(n)}_{f_1;11}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;12}(\boldsymbol{\vartheta}) & \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Gamma}^{(n)\prime}_{f_1;12}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;22}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;23}(\boldsymbol{\vartheta}) \\ \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)\prime}_{f_1;23}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;33}(\boldsymbol{\vartheta}) \end{pmatrix}^{-1}\begin{pmatrix} \Delta^{(n)}_{f_1;1}(\boldsymbol{\vartheta}) \\ \boldsymbol{\Delta}^{(n)}_{f_1;2}(\boldsymbol{\vartheta}) \\ \Delta^{(n)}_{f_1;3}(\boldsymbol{\vartheta}) \end{pmatrix} \quad (7.4)$$

of the regression on $\big(\Delta^{(n)}_{f_1;1}(\boldsymbol{\vartheta}), \boldsymbol{\Delta}^{(n)}_{f_1;2}(\boldsymbol{\vartheta}), \boldsymbol{\Delta}^{(n)}_{f_1;3}(\boldsymbol{\vartheta})\big)$ of $\boldsymbol{\Delta}^{(n)}_{f_1;4}(\boldsymbol{\vartheta})$, in the metric induced by $\boldsymbol{\Gamma}_{f_1}(\boldsymbol{\vartheta})$ or any sequence $\boldsymbol{\Gamma}^{(n)}_{f_1}(\boldsymbol{\vartheta})$ converging in probability to $\boldsymbol{\Gamma}_{f_1}(\boldsymbol{\vartheta})$; the purpose of that projection is to neutralize the impact of local perturbations $\tau_1^{(n)}$, $\boldsymbol{\tau}_2^{(n)}$, and $\tau_3^{(n)}$ of μ, $\mathbf{b}$, and σ^2. From there, the reasoning runs along the same lines as in the proof of Proposition 3.1, and we only briefly sketch it.

For any $\mathbf{u} \in \mathcal{S}_{p-1}$, a (locally and asymptotically) directionally most powerful $\boldsymbol{\Delta}^{(n)*}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta})$-based test statistic against $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\boldsymbol{\lambda}^2,\mathbf{P};f_1,h}$ with specified $\mathbf{P}$ and $\frac{\boldsymbol{\lambda}^2}{\|\boldsymbol{\lambda}^2\|} = \mathbf{u}$ can, parallel to (7.1), be based on the standardized version of

$$\mathbf{u}'\boldsymbol{\Delta}^{(n)*}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta}) = \left(\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u}\right)' \left[\frac{1}{2\sigma^2\sqrt{n}}\sum_{j=1}^{n}\psi_{f_1}(Z_j)\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}\right.$$
$$-\frac{1}{2\sigma^3}\begin{pmatrix}\mathcal{I}_{\varphi\psi}(f_1)\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)}\\ \mathcal{I}_{\varphi\psi}(f_1)\frac{1}{n}\sum_{j=1}^{n}(\mathbf{P}\mathbf{x}_j)^{\odot 2}\mathbf{x}_j'\mathbf{K}^{(n)}_1\\ \frac{1}{2\sigma}\mathcal{K}_{\varphi\psi}(f_1)\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)}\end{pmatrix}'$$
$$\left.\times\begin{pmatrix}\Gamma^{(n)}_{f_1;11}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;12}(\boldsymbol{\vartheta}) & \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta})\\ \boldsymbol{\Gamma}^{(n)'}_{f_1;12}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;22}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;23}(\boldsymbol{\vartheta})\\ \Gamma^{(n)}_{f_1;13}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)'}_{f_1;23}(\boldsymbol{\vartheta}) & \boldsymbol{\Gamma}^{(n)}_{f_1;33}(\boldsymbol{\vartheta})\end{pmatrix}^{-1}\begin{pmatrix}\Delta^{(n)}_{f_1;1}(\boldsymbol{\vartheta})\\ \boldsymbol{\Delta}^{(n)}_{f_1;2}(\boldsymbol{\vartheta})\\ \Delta^{(n)}_{f_1;3}(\boldsymbol{\vartheta})\end{pmatrix}\right].$$

The local (at $(\boldsymbol{\vartheta}, \mathbf{0})$) problem is thus the same as in the specified parameter case of Section 3.3, with, however, $\boldsymbol{\Delta}^{(n)*}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta})$ replacing $\boldsymbol{\Delta}^{(n)}_{\mathbf{P};f_1;4}(\boldsymbol{\vartheta})$. As in the proof of Proposition 3.1, the power is maximal for $(\mathbf{u}, \mathbf{P})$ such that $\mathbf{K}^{(n)}_{\mathbf{P};2}\mathbf{u} = \mathbf{1}_p$, where $\mathbf{1}_p$ stands for $(1, \ldots, 1)'$. Since, again, $\mathbf{1}_p'\big(\mathbf{P}\mathbf{x}_j\big)^{\odot 2}$ reduces to $\|\mathbf{P}\mathbf{x}_j\|^2 = \|\mathbf{x}_j\|^2$, this yields (similar to (7.3)) the test statistic $T^{(n)*}_{f_1}(\boldsymbol{\vartheta})$ defined in (3.7).

The asymptotic distribution of $T^{(n)*}_{f_1}(\boldsymbol{\vartheta})$ is normal, with mean 0 and variance one under the null. Under alternatives of the form $\mathrm{P}^{(n)}_{(\boldsymbol{\vartheta},\mathbf{0})+n^{-1/2}\boldsymbol{\nu}^{(n)}_{\mathbf{P}}\boldsymbol{\tau},\mathbf{P};f_1,h}$, it is still normal, still with variance one, but (applying Le Cam's third lemma along any subsequence such that the limit exists) with mean μ^*_T. Now, the test statistic $T^{(n)*}_{f_1}(\boldsymbol{\vartheta})$ still depends on the unspecified value of $\boldsymbol{\vartheta}$. An estimator $\hat{\boldsymbol{\vartheta}}^{(n)*}$of $\boldsymbol{\vartheta}$ can be plugged in provided, however, that it is $n^{1/2}(\boldsymbol{\nu}^{(n)})^{-1}$-*consistent* and *asymptotically discrete* (see Assumptions (D) in Section 7.3), yielding $T^{(n)*}_{f_1}(\hat{\boldsymbol{\vartheta}}^{(n)*}) = T^{(n)*}_{f_1}(\boldsymbol{\vartheta}) + o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f_1}$ and contiguous alternatives; this follows from ULAN, which implies the asymptotic linearity (with respect to $(\boldsymbol{\vartheta}, \boldsymbol{\lambda}^2)$, under density f) of central sequences, combined with Lemma 4.2 in Kreiss (1987). The asymptotic distribution, under the null and contiguous alternatives, of $T^{(n)*}_{f_1}(\hat{\boldsymbol{\vartheta}}^{(n)*})$ thus is the same as that of $T^{(n)*}_{f_1}(\boldsymbol{\vartheta})$, from which it inherits the same asymptotic optimality properties. □

8. Asymptotic linearity and cross-information quantities

8.1. Asymptotic linearity

The following lemma shows that, under Assumptions (D), substituting $\hat{\mathbf{b}}^{(n)}$ for $\mathbf{b}$ in the rank-based central sequence $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}$ has no asymptotic impact.

Lemma 8.1. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$, $\underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\mathbf{b}) = o_{\mathrm{P}}(1)$ *as* $n \to \infty$, *for any* $\mathbf{b} \in \mathbb{R}^p$.

Proof. It follows from classical asymptotic linearity results (see, e.g., Jurečková, 1969) that, for any bounded sequence $\boldsymbol{\tau}^{(n)}$,

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b} + n^{-1/2}\mathbf{K}_1^{(n)}\boldsymbol{\tau}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b}) + \begin{pmatrix} \underset{\sim}{\boldsymbol{\Gamma}}_{f,g;22} \\ \underset{\sim}{\boldsymbol{\Gamma}}'_{\mathbf{P};f,g;24} \end{pmatrix} \boldsymbol{\tau}^{(n)} = o_{\mathrm{P}}(1)$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$ and contiguous alternatives. This and Lemma 4.2 by Kreiss (1987) then entail, for $\hat{\mathbf{b}}^{(n)}$ satisfying Assumptions (D),

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b}) = -\begin{pmatrix} \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f;22} \\ \underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f;24} \end{pmatrix} n^{1/2}\left(\mathbf{K}_1^{(n)}\right)^{-1}(\hat{\mathbf{b}}^{(n)} - \mathbf{b}) + o_{\mathrm{P}}(1).$$

Now,

$$\begin{aligned} \underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\mathbf{b}) &= \left(-\underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22})^{-1}, \mathbf{I}_{p\times p}\right) \begin{pmatrix} \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f;2}(\mathbf{b}) \\ \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f;4}(\mathbf{b}) \end{pmatrix} \\ &= \left(-\underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22})^{-1}, \mathbf{I}_{p\times p}\right) \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b}). \end{aligned}$$

Hence,

$$\begin{aligned} \underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)*}_{\mathbf{P};f}(\mathbf{b}) &= \left(-\underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22})^{-1}, \mathbf{I}_{p\times p}\right)\left(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b})\right) \\ &= -\left(-\underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}(\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22})^{-1}, \mathbf{I}_{p\times p}\right) \begin{pmatrix} \underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22} \\ \underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24} \end{pmatrix} \\ &\qquad \times \left(\mathbf{K}_1^{(n)}\right)^{-1} n^{1/2}(\hat{\mathbf{b}}^{(n)} - \mathbf{b}) + o_{\mathrm{P}}(1) \\ &= 0 + o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1). \end{aligned}$$

The result follows—under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$, not just under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};f}$. □

8.2. Cross-information quantities

The following lemma provides the consistent estimators of $\mathcal{I}_\varphi(f,g)$ and $\mathcal{I}_{\psi\varphi}(f,g)$ required in Section 5.2.

Lemma 8.2. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta},\mathbf{0};g}$,

$$\begin{aligned} \mathcal{I}^{(n)}_\varphi(f) := \frac{1}{p}\mathbf{1}'_p\Bigg(\mathbf{K}_1^{(n)}\frac{1}{n}\sum_{j=1}^n \left[\mathbf{x}_j - \bar{\mathbf{x}}^{(n)}\right]\left[\mathbf{x}_j - \bar{\mathbf{x}}^{(n)}\right]'\mathbf{K}_1^{(n)}\Bigg)^{-1} \\ \times \left(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f;2}(\hat{\mathbf{b}}^{(n)}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{f;2}(\hat{\mathbf{b}}^{(n)} + n^{-1/2}\mathbf{K}_1^{(n)}\mathbf{1}_p)\right) \end{aligned} \tag{8.1}$$

and

$$\mathcal{I}^{(n)}_{\psi\varphi}(f) := \frac{2}{p}\mathbf{1}_p'\Big(\mathbf{K}^{(n)}_{\mathbf{I};2}\frac{1}{n}\sum_{j=1}^{n}\Big[\mathbf{x}_j^{\odot 2}-\overline{\mathbf{x}^{\odot 2}}\Big]\Big[\mathbf{x}_j-\bar{\mathbf{x}}^{(n)}\Big]'\mathbf{K}^{(n)}_1\Big)^{-1} \times\Big(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{I};f;4}(\hat{\mathbf{b}}^{(n)})-\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{I};f;4}(\hat{\mathbf{b}}^{(n)}+n^{-1/2}\mathbf{K}^{(n)}_1\mathbf{1}_p)\Big) \tag{8.2}$$

are such that $\mathcal{I}^{(n)}_{\varphi}(f)=\mathcal{I}_{\varphi}(f,g)+o_{\mathrm{P}}(1)$ *and* $\mathcal{I}^{(n)}_{\psi\varphi}(f)=\mathcal{I}_{\psi\varphi}(f,g)+o_{\mathrm{P}}(1)$.

Proof. For any $\boldsymbol{d}\in\mathbb{R}^p$ and any orthogonal $\mathbf{P}$, asymptotic linearity implies

$$\begin{aligned}
&\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})-\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}+n^{-1/2}\mathbf{K}^{(n)}_1\boldsymbol{d})\\
&\quad=\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)})-\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b})+\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\mathbf{b})-\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\mathbf{P};f}(\hat{\mathbf{b}}^{(n)}+n^{-1/2}\mathbf{K}^{(n)}_1\boldsymbol{d})\\
&\quad=-\begin{pmatrix}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22}\\ \underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}\end{pmatrix}n^{1/2}(\mathbf{K}^{(n)}_1)^{-1}(\hat{\mathbf{b}}^{(n)}-\mathbf{b})\\
&\qquad+\begin{pmatrix}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22}\\ \underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}\end{pmatrix}n^{1/2}(\mathbf{K}^{(n)}_1)^{-1}(\hat{\mathbf{b}}^{(n)}+n^{-1/2}\mathbf{K}^{(n)}_1\boldsymbol{d}-\mathbf{b})+o_{\mathrm{P}}(1)\\
&\quad=\begin{pmatrix}\underset{\sim}{\boldsymbol{\Gamma}}^{(n)}_{f,g;22}\\ \underset{\sim}{\boldsymbol{\Gamma}}^{(n)\prime}_{\mathbf{P};f,g;24}\end{pmatrix}\boldsymbol{d}+o_{\mathrm{P}}(1)\\
&\quad=\begin{pmatrix}\mathcal{I}_{\varphi}(f,g)\big[\mathbf{I}_p-\mathbf{K}^{(n)}_1\bar{\mathbf{x}}^{(n)}\bar{\mathbf{x}}^{(n)\prime}\mathbf{K}^{(n)}_1\big]\boldsymbol{d}\\ \frac{\mathcal{I}_{\psi\varphi}(f,g)}{2}\mathbf{K}^{(n)}_1\big[\mathbf{C}^{(n)}_{\mathbf{P};3}-\bar{\mathbf{x}}^{(n)}\overline{(\mathbf{P}\mathbf{x})^{\odot 2}}^{(n)\prime}\big]\mathbf{K}^{(n)}_{\mathbf{P};2}\boldsymbol{d}\end{pmatrix}+o_{\mathrm{P}}(1).
\end{aligned}$$

Particularizing $\boldsymbol{d}=\mathbf{1}_p$ and $\mathbf{P}=\mathbf{I}$ yields (8.1) and (8.2). □

8.3. Proof of Proposition 5.1

Proposition 5.1 follows by piecing together the Hájek asymptotic representation theorem for linear rank statistics (Chapter V of Hájek and Šidák (1967)), asymptotic linearity, and Le Cam's third lemma, then proceeding as in the proof of Proposition 3.1. Substituting $\hat{\mathbf{b}}^{(n)}$ for $\mathbf{b}$ has no asymptotic impact in view of local asymptotic linearity and Lemma 8.1. The estimation of information and cross-information coefficients is taken care of by Lemma 8.2. Then, the asymptotic null distribution in *(i)* follows from the classical asymptotic normality results for linear rank statistics (same reference to Hájek and Šidák), the asymptotic distribution under the alternative from a standard application of Le Cam's third lemma. The directional maximin property in *(ii)* is obtained along the same argument as in the proof of of Proposition 3.1. □

8.4. Data-driven scores

Thanks to the independence, under the null (hence under contiguous alternatives), between the residual ranks and the residual order statistic, data-driven scores, based on the residual order statistic, safely can be used. Computing asymptotic α-level critical values for rank statistics based on such data-driven scores as if they were deterministic yields tests with asymptotic *conditional* level α, hence asymptotic *un*conditional level α as well (conditional here means conditional on the order statistic of the residuals). Starting from that idea (developed, e.g., in Dodge and Jurečková, 2000), Hallin and Mehta (2015), in the totally different context of R-estimation for independent component analysis, propose selecting the reference density f by fitting a skew-t distribution (see Azzalini and Capitanio, 2003) with location zero, scale one, and density

$$f_{\delta,\nu}(z) = 2t_\nu(z)T_{\nu+1}\left(\delta z\left(\frac{\nu+1}{\nu+z^2}\right)^{1/2}\right) \tag{8.3}$$

to the residuals $Z_j^{(n)}$; $\delta \in \mathbb{R}$ here is a skewness parameter, and $\nu > 0$ the degrees of freedom governing the tails, while t_ν and $T_{\nu+1}$ stand for the density distribution and cumulative distribution functions of the classical Student-t distributions with ν and $\nu+1$ degrees of freedom, respectively. Estimators $\hat{\delta}$ and $\hat{\nu}$ are obtained from the (order statistic of the) residuals $Z_j^{(n)}$ using a routine maximum likelihood method (namely, maximizing a skew-t likelihood with respect to (δ, ν)). The f-score functions to be used in the testing procedure then are those associated with the skew-t density $f_{\hat{\delta},\hat{\nu}}$. Although the actual density g, in general, does not belong to the skew- t family, those $f_{\hat{\delta},\hat{\nu}}$ scores profitably adapt to its skewness and tailweight features: see Tables 9.3 and 9.4 below for empirical evidence.

9. Finite-sample performance

Sections 9.1 and 9.2 report simulations results establishing *(i)* the poor relative performance of the pseudo-Gaussian methods described in Proposition 4.1 and *(ii)* the good finite-sample performance of our rank tests. Section 9.3 develops a short application to real data.

A small Monte-Carlo experiment was conducted based on 2500 replications of a sample of size $n = 100$ generated from

$$Y_j = \mu + [\mathbf{b} + \mathbf{P}'\mathbf{\Lambda}\mathbf{P}\,\boldsymbol{\xi}_j]'\,\mathbf{x}_j + \varepsilon_j, \quad j = 1, \ldots, n = 100, \tag{9.1}$$

where $\varepsilon_1, \ldots, \varepsilon_{100}$ are i.i.d. with Gaussian density ϕ_1, logistic ℓ_1, Student t_5, skew-normal $s\mathcal{N}(5)$, and skew-t $st_5(5)$, respectively, with

$$\mu = 1,\ \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},\ \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),\ \mathbf{P} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix},$$
$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},\ \text{and } \mathbf{x} := \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 10 & 1 \\ 1 & 20 \end{pmatrix}\right).$$

9.1. The poor performance of pseudo-Gaussian tests

The simulation results in Tables 9.1 (spherical $\mathbf{\Lambda}$) and 9.2 (eccentric $\mathbf{\Lambda}$) below reveal the poor performance of the pseudo-Gaussian methods described in Proposition 4.1 under two- and three-dimensional skew normal densities.

TABLE 9.1

Rejection frequencies, under skew-normal f ($f_1 = s\mathcal{N}(2)$ and $f_1 = s\mathcal{N}(3)$), of the parametric optimal test based on $T^{(n)*}_{f_1}$ and the pseudo-Gaussian test based on $T^{(n)\bullet}_{\phi_1}$, for various values of $\mathbf{\Lambda} = \lambda\mathbf{I}$, $n = 100$, and $\alpha = 5\%$; 2500 replications.

f_1	test statistic	$\lambda_1 = \lambda_2 = \lambda$: 0	0.05	0.075	0.1	0.125
$s\mathcal{N}(2)$	$T^{(n)*}_{f_1}$	0.0480	0.1712	0.3648	0.5532	0.7420
	$T^{(n)\bullet}_{\phi_1}$	0.0500	0.1080	0.2220	0.3740	0.5360
$s\mathcal{N}(3)$	$T^{(n)*}_{f_1}$	0.0596	0.2492	0.4644	0.6928	0.8424
	$T^{(n)\bullet}_{\phi_1}$	0.0516	0.1128	0.2072	0.3612	0.4960

TABLE 9.2

Rejection frequencies, under skew-normal f ($f_1 = s\mathcal{N}(2)$ and $f_1 = s\mathcal{N}(3)$), of the parametric optimal test based on $T^{(n)*}_{f_1}$ and the pseudo-Gaussian test based on $T^{(n)\bullet}_{\phi_1}$, for various values of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, $n = 100$, and $\alpha = 5\%$; 2500 replications.

f_1	test statistic	(λ_1, λ_2): $(0.075, 0.05)$	$(0.1, 0.05)$	$(0.125, 0.05)$	$(0.125, 0.075)$	$(0.125, 0.1)$
$s\mathcal{N}(2)$	$T^{(n)*}_{f_1}$	0.2596	0.3964	0.5400	0.5984	0.6708
	$T^{(n)\bullet}_{\phi_1}$	0.1624	0.2592	0.3648	0.3968	0.4484
$s\mathcal{N}(3)$	$T^{(n)*}_{f_1}$	0.3796	0.5252	0.6856	0.7336	0.8028
	$T^{(n)\bullet}_{\phi_1}$	0.1760	0.2388	0.3396	0.4056	0.4500

This weakness of the pseudo-Gaussian method is particularly marked under high-eccentricity covariance matrices $\mathbf{P}\mathbf{\Lambda}^2\mathbf{P}'$: in Table 9.2, rejection frequencies of the optimal parametric test, under skew-normal densities $s\mathcal{N}(3)$, are about twice those of the pseudo-Gaussian test.

9.2. A comparison of the finite-sample performance of pseudo-Gaussian and rank-based tests

Tables 9.3 and 9.4 report rejection frequencies ($\alpha = 5\%$), for various values of $\mathbf{\Lambda}$ (spherical $\mathbf{\Lambda}$ in Table 9.3, eccentric $\mathbf{\Lambda}$ in Table 9.4), of the following tests: pseudo-Gaussian(based on $T^{(n)\bullet}_{\phi_1}$), van der Waerden (based on $\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$), Wilcoxon (based on $\underset{\sim}{T}^{(n)}_{\mathrm{W}}$), Student-$t_5$ (based on $\underset{\sim}{T}^{(n)}_{t_5}$), and the test based on $\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$ involving data-driven skew Student $st_{\hat{\nu}}(\hat{\delta})$ scores (see Section 8.4).

Table 9.3

Rejection frequencies (2500 replications), for $\lambda_1 = \lambda_2 = 0$ (null hypothesis), $0.05,\ 0.1,\ 0.15,\ 0.2$ and $(\lambda_1, \lambda_2) = (0.05, 0),\ (0.1, 0)$, and $(0.2, 0)$ (local alternatives), with normal (ϕ_1), logistic (ℓ_1), Student (t_5), skew-normal ($s\mathcal{N}(5)$) and skew Student ($st_5(5)$) error distributions, of the pseudo-Gaussian test ($T^{(n)\bullet}_{\phi_1}$), the van der Waerden test ($\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$), the Wilcoxon test ($\underset{\sim}{T}^{(n)}_{\mathrm{W}}$), the Student ($t_5$-score) test ($\underset{\sim}{T}^{(n)}_{t_5}$), and the test $\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$ based on data-driven skew Student $st_{\hat{\nu}}(\hat{\delta})$ scores, for $n = 100$ and $\alpha = 5\%$.

		$\lambda_1 = \lambda_2$					$(\lambda_1, 0)$		
g_1	Test	0	0.05	0.1	0.15	0.2	0.05	0.1	0.2
ϕ_1	$T^{(n)\bullet}_{\phi_1}$	0.0484	0.1104	0.3836	0.6928	0.8480	0.0872	0.2236	0.6512
	$\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$	0.0468	0.1040	0.3524	0.6600	0.8288	0.0800	0.1980	0.6136
	$\underset{\sim}{T}^{(n)}_{\mathrm{W}}$	0.0428	0.0888	0.2860	0.5904	0.7736	0.0644	0.1664	0.5220
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.0432	0.0872	0.2644	0.5400	0.7388	0.0668	0.1536	0.4748
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.0480	0.1040	0.3300	0.6040	0.8040	0.0784	0.1844	0.5448
ℓ_1	$T^{(n)\bullet}_{\phi_1}$	0.0516	0.0684	0.1208	0.2404	0.3908	0.0628	0.0916	0.2392
	$\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$	0.0448	0.0728	0.1204	0.2500	0.4052	0.0652	0.0876	0.2408
	$\underset{\sim}{T}^{(n)}_{\mathrm{W}}$	0.0460	0.0688	0.1160	0.2440	0.3896	0.0604	0.0836	0.2284
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.0468	0.0692	0.1100	0.2324	0.3756	0.0628	0.0820	0.2220
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.0480	0.0560	0.1040	0.2100	0.3560	0.0508	0.0796	0.1804
t_5	$T^{(n)\bullet}_{\phi_1}$	0.0616	0.0824	0.1952	0.4192	0.6208	0.0748	0.1320	0.4000
	$\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$	0.0572	0.0932	0.2280	0.4760	0.6828	0.0768	0.1460	0.4500
	$\underset{\sim}{T}^{(n)}_{\mathrm{W}}$	0.0512	0.0940	0.2416	0.4672	0.6620	0.0736	0.1540	0.4212
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.0504	0.0924	0.2380	0.4504	0.6392	0.0744	0.1516	0.4000
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.0440	0.0800	0.2120	0.42400	0.6160	0.0548	0.1124	0.3840
$s\mathcal{N}(5)$	$T^{(n)\bullet}_{\phi_1}$	0.0592	0.1204	0.3464	0.6564	0.8084	0.1468	0.4660	0.8744
	$\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$	0.0596	0.2388	0.5560	0.8056	0.8964	0.1828	0.5052	0.8748
	$\underset{\sim}{T}^{(n)}_{\mathrm{W}}$	0.0448	0.1840	0.5192	0.7836	0.8844	0.1324	0.4040	0.8128
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.0440	0.1600	0.4856	0.7504	0.8652	0.1092	0.3648	0.7704
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.0560	0.4440	0.7560	0.8160	0.9080	0.2224	0.4808	0.8224
$st_5(5)$	$T^{(n)\bullet}_{\phi_1}$	0.0516	0.2344	0.6988	0.9088	0.9640	0.0880	0.1972	0.6176
	$\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$	0.0508	0.2856	0.7216	0.9060	0.9660	0.1508	0.3620	0.7436
	$\underset{\sim}{T}^{(n)}_{\mathrm{W}}$	0.0436	0.2208	0.6452	0.8744	0.9468	0.1108	0.3116	0.7076
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.0400	0.1996	0.6056	0.8496	0.9280	0.0976	0.2836	0.6804
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.0480	0.3560	0.7240	0.8960	0.9360	0.2008	0.3928	0.6844

Table 9.4

Rejection frequencies (2500 replications), for various values of (λ_1, λ_2), with normal (ϕ_1), logistic (ℓ_1), Student (t_5), skew-normal $(s\mathcal{N}(5))$ and skew Student $(st_5(5))$ error distributions, of the pseudo-Gaussian test $(T^{(n)\bullet}_{\phi_1})$, the van der Waerden test $(\underset{\sim}{T}^{(n)}_{\text{vdW}})$, the Wilcoxon test $(\underset{\sim}{T}^{(n)}_{\text{W}})$, the Student ($t_5$-score) test $(\underset{\sim}{T}^{(n)}_{t_5})$, and the test $\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$ based on data-driven skew Student $st_{\hat{\nu}}(\hat{\delta})$ scores, for $n = 100$ and $\alpha = 5\%$.

g_1	Test	(λ_1, λ_2) $(0.1, 0.05)$	$(0.15, 0.05)$	$(0.2, 0.05)$	$(0.3, 0.05)$	$(0.2, 0.1)$
ϕ_1	$T^{(n)\bullet}_{\phi_1}$	0.2640	0.4576	0.6748	0.8508	0.7260
	$\underset{\sim}{T}^{(n)}_{\text{vdW}}$	0.2444	0.4184	0.6372	0.8344	0.6916
	$\underset{\sim}{T}^{(n)}_{\text{W}}$	0.2008	0.3548	0.5548	0.7720	0.6184
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.1836	0.3172	0.5050	0.7332	0.5732
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.1560	0.2840	0.4240	0.6840	0.5040
ℓ_1	$T^{(n)\bullet}_{\phi_1}$	0.0956	0.1544	0.2724	0.4888	0.2660
	$\underset{\sim}{T}^{(n)}_{\text{vdW}}$	0.1032	0.1588	0.2696	0.4804	0.2964
	$\underset{\sim}{T}^{(n)}_{\text{W}}$	0.1008	0.1600	0.2468	0.4456	0.2824
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.0976	0.1568	0.2368	0.4136	0.2724
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.0760	0.1440	0.2240	0.4040	0.2520
t_5	$T^{(n)\bullet}_{\phi_1}$	0.1436	0.2708	0.4384	0.6852	0.4512
	$\underset{\sim}{T}^{(n)}_{\text{vdW}}$	0.1664	0.3052	0.4748	0.7260	0.5000
	$\underset{\sim}{T}^{(n)}_{\text{W}}$	0.1696	0.3124	0.4608	0.6912	0.5088
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.1668	0.3004	0.4368	0.6628	0.4892
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.1360	0.2640	0.4240	0.6440	0.4700
$s\mathcal{N}(5)$	$T^{(n)\bullet}_{\phi_1}$	0.5484	0.7740	0.8864	0.9416	0.9144
	$\underset{\sim}{T}^{(n)}_{\text{vdW}}$	0.5792	0.7752	0.8844	0.9496	0.9204
	$\underset{\sim}{T}^{(n)}_{\text{W}}$	0.4868	0.7132	0.8308	0.9272	0.8808
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.4436	0.6624	0.7940	0.9100	0.8580
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.5640	0.7760	0.8800	0.9240	0.9160
$st_5(5)$	$T^{(n)\bullet}_{\phi_1}$	0.2608	0.4484	0.6396	0.8308	0.6800
	$\underset{\sim}{T}^{(n)}_{\text{vdW}}$	0.4504	0.6308	0.7784	0.8996	0.8264
	$\underset{\sim}{T}^{(n)}_{\text{W}}$	0.4044	0.5692	0.7524	0.8796	0.7991
	$\underset{\sim}{T}^{(n)}_{t_5}$	0.3776	0.5384	0.7244	0.8564	0.7724
	$\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$	0.5120	0.5600	0.7760	0.8640	0.8440

Inspection of those two tables reveals that under the null all rank-based tests have rejection frequencies quite close to the nominal size of 5%, while the pseudo-Gaussian test significantly over-rejects under Student and skew-normal densities; van der Waerden slightly outperforms the pseudo-Gaussian test under logistic and t_5 densities, quite significantly so under a skew-normal one; the same van der Waerden has similar performances as the data-driven-score test based on $\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$ under Gaussian and logistic densities (under t_5, it is not valid), but does much worse under skew-normal and skew-t. The test based on $\underset{\sim}{T}^{(n)}_{st_{\hat{\nu}}(\hat{\delta})}$ has excellent overall performance, and constitutes the best choice in the presence of skewness and heavier-than-normal tails.

9.3. An empirical illustration

As an empirical illustration, we consider the housing price dataset `HPRICE1.RAW` studied by Wooldridge (2012) (available at `https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041`), consisting of $n = 88$ observations of the variables `price`=*house price in thousands of dollars*, `assess`= *assessed value in thousands of dollars*, and `sqrft`= *size of house in square feet*. The prices `price` play the role of dependent variable.

On this dataset, we implemented the classical Breusch-Pagan test T_{BP} and some of the tests proposed in this paper: the pseudo-Gaussian test ($T^{(n)\bullet}_{\phi_1}$), the van der Waerden test ($\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$), the Wilcoxon test ($\underset{\sim}{T}^{(n)}_{\mathrm{W}}$), and the Student ($t_5$-score) test.

Table 9.5 provides the various p-values: unlike T_{BP} and $T^{(n)\bullet}_{\phi_1}$, $\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$, $\underset{\sim}{T}^{(n)}_{\mathrm{W}}$, and the Student ($t_5$-score) test all quite significantly reject the hypothesis of constant regression coefficients.

TABLE 9.5
p-values of the Breusch-Pagan (T_{BP}), pseudo-Gaussian ($T^{(n)\bullet}_{\phi_1}$), van der Waerden ($\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$), Wilcoxon ($\underset{\sim}{T}^{(n)}_{\mathrm{W}}$), and Student ($t_5$-score) tests for the housing price dataset `HPRICE1.RAW` in the regression of `price` on `assess` and `sqrft`.

Test	T_{BP}	$T^{(n)\bullet}_{\phi_1}$	$\underset{\sim}{T}^{(n)}_{\mathrm{vdW}}$	$\underset{\sim}{T}^{(n)}_{\mathrm{W}}$	$\underset{\sim}{T}^{(n)}_{t_5}$
p-value	0.2174	0.1395	0.0022	0.0030	0.0038

10. Conclusions

The apparently simple problem of detecting random coefficients in multiple regression proves to be surprisingly complex, with a nonstandard ULAN structure, non-diagonal information matrices, cone-shaped alternatives, and nuisance parameters that are not identified under the null. Moreover, the pseudo-Gaussian test appears to have quite poor performances under skewed and heavy-tailed densities. We therefore construct rank-based tests, which exhibit remarkably high ARE values with respect to their pseudo-Gaussian counterpart. Their excellent performances are confirmed by simulations.

References

R. P. Abelson and J. W. Tukey. Efficient utilization of non-numerical information in quantitative analysis general theory and the case of simple order. *The Annals of Mathematical Statistics*, 34:1347–1369, 1963. MR0156411

A. Akharif and M. Hallin. Efficient detection of random coefficients in autoregessive models. *Annals of Statistics*, 31:675–704, 2003. MR1983546

A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B*, 65:367–389, 2003. MR1983753

R. Beran and P. G. Hall. Estimating coefficient distributions in random coefficient regressions. *Annals of Statistics*, 20:1970–1984, 1992. MR1193321

R. Beran and P. W. Millar. Minimum distance estimation in random coefficient regression. *Annals of Statistics*, 22:1976–1992, 1994. MR1329178

R. Beran, A. Feuerverger, and P. G. Hall. On nonparametric estimation of intercept and slope distributions in random coefficient regression. *Annals of Statistics*, 24:2569–2592, 1996. MR1425969

P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore, 1993. MR1245941

T. S. Breusch and A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47:1287–1294, 1979. MR0545960

H. Chernoff and I. R. Savage. Asymptotic normality and efficiency of certain nonparametric test statistics. *The Annals of Mathematical Statistics*, 29:972–994, 1958. MR0100322

P. Delicado and J. Romo. Goodness-of-fit tests in random coefficient regression models. *Annals of the Institute of Statistisitical Mathematics*, 51:125–148, 1999. MR1704650

P. Delicado and J. Romo. Random coefficient regressions: parametric goodness of fit. *Journal of Statistical Planning and Inference*, 119:377–400, 2004. MR2019647

Y. Dodge and J. Jurečková. *Adaptive Regression*. New York: Springer, 2000. MR1932533

M. Fihri, A. Akharif, A. Mellouk, and M. Hallin. Efficient pseudo-Gaussian and rank-based detection of random regression coefficients. *Journal of Nonparametric Statistics*, 32:367–402, 2020. MR4101384

B. Garel and M. Hallin. Local asymptotic normality of multivariate arma processes with a linear trend. *Annals of the Institute of Statistical Mathematics*, 47:551–579, 1995. MR1364260

J. Gu. Neyman's $C(\alpha)$ test for unobserved heterogeneity. *Econometric Theory*, 32:1483–1522, 2016. MR3569281

J. Gu, R. Koenker, and S. Volgushev. Testing for homogeneity in mixture models. *Econometric Theory*, 34:850–895, 2018. MR3815469

J. Hájek. Local asymptotic minimax and admissibility in estimation. *In Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1:175–194, 1972. MR0400513

J. Hájek and Z. Šidák. *Theory of Rank Tests*. Academic Press, New York-London, 1967. MR0229351

M. Hallin. On the Pitman-nonadmissibility of correlogram-based methods. *Journal of Time Series Analysis*, 15:607–612, 1994. MR1312324

M. Hallin and D. La Vecchia. R-estimation in semiparametric dynamic location-scale models. *Journal of Econometrics*, 196:233–247, 2017. MR3589479

M. Hallin and C. Mehta. R-estimation for asymmetric independent component analysis. *Journal of the American Statistical Association*, 110:218–232, 2015. MR3338498

M. Hallin and D. Paindaveine. Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Annals of Statistics*, 30:1103–1133, 2002a. MR1926170

M. Hallin and D. Paindaveine. Optimal procedures based oninterdirections and pseudo-Mahalanobis ranks for testing multivariate elliptic white noise against VARMA dependence. *Bernoulli*, 8:787–816, 2002b. MR1963662

M. Hallin and D. Paindaveine. Chernoff-Savage and Hodges-Lehmann results for Wilks' test of independence. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, pages 184–196. I.M.S. Lecture Notes-Monograph Series, 2008. MR2462206

M. Hallin and O. Tribel. The efficiency of some nonparametric rank-based competitors to correlogram methods. In F. T. Bruss and L. LeCam, editors, *Game Theory, Optimal Stopping, Probability, and Statistics: Papers in Honor of T.S. Ferguson on the Occasion of his 70th Birthday*, pages 249–262. I.M.S. Lecture Notes-Monograph Series, 2000. MR1833863

M. Hallin and B. J. M. Werker. Semiparametric efficiency, distribution-freeness and invariance. *Bernoulli*, 9:137–165, 2003. MR1963675

C. Hildreth and J. P. Houck. Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, 63:584–595, 1968. MR0232502

J. Jurečková. Asymptotic linearity of a rank statistic in regression parameter. *The Annals of Mathematical Statistics*, 40:1889–1900, 1969. MR0248931

J. P. Kreiss. On adaptive estimation in stationary ARMA processes. *Annals of Statistics*, 15:112–133, 1987. MR0885727

A. Kudo. A multivariate analogue of the one-sided test. *Biometrika*, 50:403–418, 1963. MR0163386

L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer series in Statistics. Springer-Verlag, New York, 1986. MR0856411

L. Le Cam and G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag, New York, 2nd edition, 2000. MR1784901

B. Lind and G. Roussas. A remark on quadratic mean differentiability. *The Annals of Mathematical Statistics*, 43:1030–1034, 1972. MR0307329

F. Lombard. An elementary proof of asymptotic normality for linear rank statistics. *South African Statistical Journal*, 20:29–35, 1986. MR0860548

A. Mallet. A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, 73:645–656, 1986. MR0897856

P. Newbold and T. Bos. *Stochastic Parameter Regression Models*. Beverly Hills California Sage Publications, 1985. MR0807151

D. F. Nicholls and A. R. Pagan. Varying coefficient regression. *E. J. Hannan, P. R. Krishnaiah, M. M. Rao, eds., Handbook of Statistics*, 5:413–449, 1985. MR0831758

P. A. Novikov. A locally directionally maximin test for a multidimensional parameter with order-restricted alternatives. *Russian Mathematics*, 55:33–41, 2011. MR2810192

D. Paindaveine. A Chernoff-Savage result for shape. on the non-admissibility of pseudo-Gaussian methods. *Journal of Multivariate Analysis*, 97:2206–2220, 2006. MR2301635

B. Raj. Linear regression with random coefficients: The finite sample and convergence properties. *Journal of the American Statistical Association*, 70:127–137, 1975.

B. Raj and A. Ullah. *Econometrics, A Varying Coefficient Approach*. Croom-Helm, London, 1981.

T. Ramanathan and M. Rajarshi. Rank tests for testing randomness of a regression coefficient in a linear regression model. *Metrika*, 39:113–124, 1992. MR1162691

W. Schaafsma and L. Smid. Most stringent somewhere most powerful tests against alternatives restricted by a number of linear inequalities. *The Annals of Mathematical Statistics*, 37:1161–1172, 1966. MR0199920

P. A. V. B. Swamy. *Statistical Inference in Random Coefficient Regression Models*. Springer-Verlag, New York, 1971. MR0329146

A. R. Swensen. The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis*, 16:54–70, 1985. MR0778489

A. Wald. A note on regression analysis. *The Annals of Mathematical Statistics*, 18:586–589, 1947. MR0023498

J. Wooldridge. *Introductory Econometrics: a Modern Approach, Fifth Edition*. South-Western Cengage Learning, 2012.