Document de travail
Discussion paper

N°2024-04

https://lem.univ-lille.fr

Semiparametrically highly efficient estimation of spatial autoregressive models

Nicolas Debarsy

Univ. Lille, CNRS, IESEG School of Management, UMR 9221 – LEM, Lille Economie Management, F-59000, France | nicolas.debarsy@cnrs.fr

Vincenzo Verardi

Université Catholique de Louvain, LouRIM, Place des doyens 1 /L2.01.01. B-1348 Louvain-la-Neuve

Catherine Vermandele

Université Libre de Bruxelles, LMTD, 50 Av. F. D. Roosevelt, B-1050 Brussels

Semiparametrically highly efficient estimation of spatial autoregressive models

Nicolas Debarsy[a,*], Vincenzo Verardi[b,1], Catherine Vermandele[c]

[a] *Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, Univ. Lille, site cité scientifique, Building SH2 , 59000, Lille, France*
[b] *Université catholique de Louvain. LouRIM, Place des Doyens 1/L2.01.01. , Louvain-la-Neuve, 1348, Belgium*
[c] *Université libre de Bruxelles, LMTD, Av. F. D. Roosevelt, 50, Brussels, 1050, Belgium*

Abstract

Spatial autoregressive (SAR) models cannot generally be estimated using ordinary least squares given the simultaneity that results from interactions among individuals. Instead, two-stage least squares (Kelejian and Prucha, 1998; Bramoullé et al., 2009), generalized method of moments (Liu et al., 2010), or (quasi-)maximum likelihood (Lee, 2004) approaches are used.

In this article, we propose a semiparametrically highly efficient estimator, based on the Local Asymptotic Normality theory of Le Cam (1960) and the rank-and-sign semiparametric approach developed by Hallin et al. (2006, 2008). Monte Carlo simulations show that the suggested estimator outperforms existing estimators as soon as one deviates from a normal distribution of the error term. A trade regression from Behrens et al. (2012) (used differently from the original paper) is mobilized to illustrate how empirical findings might be affected when the Gaussian distribution is not imposed.

Keywords: Spillovers, Efficiency, Local Asymptotic Normality, Semiparametric estimation, Ranks and Signs

---

*Corresponding author

[1] At the time the work was carried out, Vincenzo Verardi was Associated Researcher of the FNRS (Fonds National de la Recherche Scientifique) and gratefully acknowledges their financial support. Present affiliation: Université Catholique de Louvain. LouRIM. Place des Doyens 1/L2.01.01. B-1348 Louvain-la-Neuve.

1. Introduction

It is well known that spatial autoregressive (SAR) models cannot generally be estimated by Ordinary Least Squares. This has led to the development of various estimators based on Two Stages Least Squares (TSLS), the Generalized Method of Moments (GMM), or Maximum Likelihood (ML) (see Kelejian and Prucha, 1998, 1999; Lee, 2004, 2007; Bramoullé et al., 2009; Liu et al., 2010).

ML estimation yields the most efficient estimator if the assumed distribution (typically the Gaussian) of the error term coincides with the actual one. However, when the distribution of the error term is unknown, the maximum likelihood estimator cannot be estimated. If a distribution is imposed, convergence is guaranteed only when the assumed distribution is either the true one or belongs to the Linear Exponential Family (LEF) (Gourieroux et al., 1984); otherwise, convergence is not guaranteed.

Ord (1975) was the first to develop the MLE for the SAR model, assuming a normal distribution of the error term. Lee (2004) generalizes the estimator by proposing a quasi-maximum likelihood (QML) approach. The estimator is obtained by maximizing a Gaussian likelihood, but the associated inference is based on a Fisher information matrix that is adjusted to remain valid under departures from normality. In this sense, the Gaussian QML estimator is consistent under standard regularity conditions (since the Normal distribution is LEF) even when the true errors are not Gaussian, although it is generally less efficient than the maximum likelihood estimator that is correctly specified for the true distribution.

Using both linear and quadratic moments, Liu et al. (2010) introduce a GMM estimator that does not require any assumption about the distribution of error terms. When the error terms are normally distributed, the authors show that this estimator is as efficient as the ML estimator. Furthermore, it generally performs better than the Gaussian QML estimator when the normality assumption is not satisfied.

Robinson (2010) proposes a semiparametric estimator in which the density of the error term is jointly estimated with the parameters of interest through the series estimation of the score function. However, the proposed estimator is based on a specific type of interaction scheme in which each observation, in the limit, has an infinite number of neighbors. This rules out the large majority of the literature, where the network of each observation consists of a limited number of peers/neighbors. Lee (2002) showed that this specific type of interaction scheme was the key assumption to derive the consistency of the ordinary least squares estimator.

In this paper, we propose a *semiparametric* approach in which the innovation density is viewed as an infinite-dimensional nuisance parameter in the regression model.

More precisely, relying on the framework of Local Asymptotic Normality (LAN) introduced by Le Cam (1960) and subsequent developments by Hallin and Werker (2003); Hallin et al. (2006, 2008), we propose a rank-and-sign (R&S) semiparametric estimator for SAR models, which can be argued, on heuristic grounds, to be asymptotically semiparametrically efficient.

We then perform Monte Carlo experiments to evaluate the behavior of the proposed estimator in finite samples and observe that it outperforms existing alternatives as soon as one deviates from the Gaussian distribution of the error term.

Finally, using a trade model developed by Behrens et al. (2012), we illustrate the usefulness of the developed estimator in applied research, considering it in a different context than the authors of the original paper.

The rest of the paper is organized as follows. Section 2 presents the *semiparametric* model considered here. Section 3 is devoted to the Uniform Local Asymptotic Normality (ULAN) property of the parametric submodel associated with a fixed density f for the error term and presents how a one-step efficient estimator of the parameters of this submodel can be defined. Section 4 details how the semiparametric rank and sign-based estimator procedure is obtained. Section 5 presents the practical implementation of the proposed estimator. Section 6 is dedicated to Monte Carlo experiments that compare the R&S estimator with QMLE and GMME. The efficiency of the R&S estimator is comparable to that of the ML estimator when errors are normally distributed. However, when the error component is distributed according to another distribution function, the R&S estimator exhibits substantially higher efficiency compared to competing estimators. Section 7 applies the R&S estimator to a trade model developed by Behrens et al. (2012) and compares the point estimates and their standard errors to those obtained in the original paper, which relies on the Gaussian QML of Lee (2004) and the GMM estimator. Finally, Section 8 concludes. All notations used in the article are collected in Appendix Appendix A.1.

2. The semiparametric SAR model $\boldsymbol{\mathcal{E}}^{(n)}$

Consider the following SAR model.[2] For $i = 1, \ldots, n$,

$$y_i^{(n)} = (\mathbf{x}_i^{(n)})^{\mathrm{T}}\boldsymbol{\beta} + \lambda \sum_{\substack{j=1 \\ j\neq i}}^{n} w_{ij}^{(n)} y_j^{(n)} + \varepsilon_i^{(n)}, \tag{1}$$

where

- n is the considered sample size;
- $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. error terms with an unknown distribution function F and density f;
- $\mathbf{x}_i^{(n)} = \left(1, x_{i1}^{(n)}, \ldots, x_{iK}^{(n)}\right)^{\mathrm{T}}$ is the vector of explanatory variables values for individual i and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)^{\mathrm{T}} \in \mathbb{R}^{K+1}$ is the associated vector of regression parameters;
- $\sum_{j=1,j\neq i}^{n} w_{ij}^{(n)} y_j^{(n)}$ represents endogenous spillover effects and λ is the associated regression coefficient. The definition of the relevant interaction scheme, modeled by the elements $w_{ij}^{(n)}$ of the general connectivity matrix $\mathbf{W}^{(n)}$, depends on the question under study. In the social-network literature, peers are individuals who influence the behavior of a specific individual i, such as friends, geographic neighbors, housemates, or coworkers. In the context of international trade, Behrens et al. (2012) show that links between regions should be modeled by their relative share of the population. As normalization, $w_{ii}^{(n)} = 0$ for all i.

Remark 1.. Let $\mathbf{I}^{(n)}$ be the $(n \times n)$-identity matrix. By writing the model (1) for the whole sample and assuming that the matrix $\mathbf{I}^{(n)} - \lambda \mathbf{W}^{(n)}$ is nonsingular, we compute its reduced form as:

$$\mathbf{y}^{(n)} = \left(\mathbf{I}^{(n)} - \lambda \mathbf{W}^{(n)}\right)^{-1} \left(\mathbf{X}^{(n)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{(n)}\right),$$

[2] As soon as we abstract from a group interaction scheme (with groups of equal size) and assume a deterministic interaction scheme, the model can include contextual effects (neighbors' characteristics) without additional difficulties.

where $\mathbf{y}^{(n)} = (y_1^{(n)}, \ldots, y_n^{(n)})^{\mathrm{T}}$, $\mathbf{X}^{(n)} = (\mathbf{x}_1^{(n)}, \ldots, \mathbf{x}_n^{(n)})^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}^{(n)} = (\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)})^{\mathrm{T}}$. Further, we have

$$\mathbf{W}^{(n)}\mathbf{y}^{(n)} = \mathbf{G}^{(n)}(\lambda)\left(\mathbf{X}^{(n)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^{(n)}\right), \tag{2}$$

with $\mathbf{G}^{(n)}(\lambda) = \mathbf{W}^{(n)}\left(\mathbf{I}^{(n)} - \lambda\mathbf{W}^{(n)}\right)^{-1}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}}$, $\mathbf{W}_{i\cdot}^{(n)}$ be the i^{th} row of matrix $\mathbf{W}^{(n)}$, $\mathbf{G}_{i\cdot}^{(n)}(\lambda)$ be the i^{th} row of matrix $\mathbf{G}^{(n)}(\lambda)$, and $e_i^{(n)}(\boldsymbol{\theta})$ $(i = 1, \ldots, n)$ be the residuals associated with the value $\boldsymbol{\theta}$ of the parameters vector, with $\mathbf{e}^{(n)}(\boldsymbol{\theta}) = (e_1^{(n)}(\boldsymbol{\theta}), \ldots, e_n^{(n)}(\boldsymbol{\theta}))^{\mathrm{T}}$. Then, we may write:

$$\begin{aligned} e_i^{(n)}(\boldsymbol{\theta}) &= y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}}\boldsymbol{\beta} - \lambda\mathbf{W}_{i\cdot}^{(n)}\mathbf{y}^{(n)} & (3)\\ &= y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}}\boldsymbol{\beta} - \lambda\mathbf{G}_{i\cdot}^{(n)}(\lambda)\left(\mathbf{X}^{(n)}\boldsymbol{\beta} + \mathbf{e}^{(n)}(\boldsymbol{\theta})\right) & \\ &= y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}}\boldsymbol{\beta} - \lambda\mathbf{G}_{i\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} - \lambda\sum_{j=1}^{n} G_{ij}^{(n)}(\lambda)e_j^{(n)}(\boldsymbol{\theta}). & (4) \end{aligned}$$

2.1. Existing estimators

The literature has developed several estimators for the model (1), to account for the simultaneity caused by the endogenous spillover effect. Assuming the normality of the error term, Ord (1975) developed the ML procedure, which consists of maximizing the log-likelihood function shown in (5).

$$\ln L_f\left(\boldsymbol{\theta}, \sigma^2 \,\middle|\, \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln\left|\det\left(\mathbf{I}^{(n)} - \lambda\mathbf{W}^{(n)}\right)\right| - \frac{1}{2\sigma^2}(\mathbf{e}^{(n)}(\boldsymbol{\theta}))^{\mathrm{T}}\mathbf{e}^{(n)}(\boldsymbol{\theta}). \tag{5}$$

Lee (2004) went one step further and developed a quasi-MLE, assuming normal errors, which takes into consideration the third and fourth moments of the distribution in the Fisher Information matrix. This Gaussian QML estimator remains consistent even if the error distribution is non-normal, but its efficiency is lower than that of the ML estimator under the true distribution.

Alternatively, Kelejian and Prucha (1998) developed a Two-Stage Least Squares Estimator (TSLSE) where the endogenous spatial lag, $\mathbf{W}^{(n)}\mathbf{y}^{(n)}$, is instrumented by higher-order "spatial lags" of the explanatory variables, typically relying on linearly independent columns of $\mathbf{W}^{(n)}\mathbf{X}^{(n)}$ and $(\mathbf{W}^{(n)})^2\mathbf{X}^{(n)}$. In the context of social networks, Bramoullé et al. (2009) show that these instruments will identify endogenous effects as soon as the connectivity scheme includes intransitive triads, i.e. triads such that "peers of my peers are not my peers". Besides, Lee (2003) discusses the best set of instruments and show that this approach will always be less efficient than ML approach (under normality). Besides, he notes that the TSLSE would not be consistent if all regressors are irrelevant.

As a solution, Lee (2007) and Liu et al. (2010) have developed efficient GMM estimators of model (1) which rely, in addition to the linear moments used in TSLS, on nonlinear (quadratic) moment conditions. These additional moments exploit the information contained in the error terms of equation (2) and $\mathbf{W}^{(n)}\mathbf{y}^{(n)}$. Lee (2007) develops the best GMME assuming normality of the error terms and show that it is asymptotically as efficient as the MLE (under normality). Liu et al. (2010) go one step further and propose distribution-free best GMME, which are shown to be asymptotically more efficient than the QMLE.

Finally, Robinson (2010) has proposed an adaptive semiparametric estimator with unknown distribution of the error terms. His idea is to nonparametrically estimate the score function using series estimates. His proposed estimator is as efficient as the MLE under the correct distribution. However, according to Liu et al. (2010), the adaptivity property requires orthogonality conditions, which only hold in specific cases of the SAR models. One of these cases occurs when each observation is assumed to have many neighbors (the number of neighbors has to increase with the sample size). Lee (2002) showed that this particular setup leads to consistent OLS estimation of the model (1). Robinson (2010) focuses on this type of interaction matrix and considers $\mathbf{W}^{(n)}$ that has nonnegative elements that are uniformly of order $O(1/h_n)$. He further assumes that h_n increases with the sample size n with either $h_n/n^{1/2} \to \infty$ as $n \to \infty$, or $h_n \to \infty$ as $n \to \infty$ with either a symmetric $\mathbf{W}^{(n)}$ or a symmetric distribution of the error term $\varepsilon_i^{(n)}$. These convergence rates do not align with those used in this paper, which come from Lee (2004) and are presented in Assumption B. Consequently, the interaction framework adopted in Robinson (2010) is only suitable for modeling interactions in specific cases that are not considered in this paper. Instead, we focus on the arguably more general case where each observation has a limited number of peers, which constitute the network definition used in the majority of the existing literature.

2.2. Regularity conditions

The definition and properties of the rank-and-sign estimator proposed here require some regularity conditions detailed below. Assumption A concerns the covariate vectors $\mathbf{x}_i^{(n)}$, while Assumption B relates to the interaction terms $w_{ij}^{(n)}$ and the endogenous effects parameter λ; these first two assumptions come from Lee (2004). Assumption C, required to apply the results of Hallin et al. (2006) for rank-and-sign statistics, completes Assumptions A and B. Assumption D specifies regularity conditions for the unknown distribution of the error terms.

Assumption A. *The elements of* $\mathbf{x}_i^{(n)}$ *are uniformly bounded constants for all* n*. Besides, the* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^{(n)}(\mathbf{x}_i^{(n)})^{\mathrm{T}}$ *exists and is non-singular.*

Assumption B.

(B1) *The elements* $w_{ij}^{(n)}$ *of the matrix* $\mathbf{W}^{(n)}$ *are at most of order* $1/h^{(n)}$ *— they are* $O(1/h^{(n)})$ *— uniformly in all* i, j*, where the rate sequence* $\{h^{(n)}\}$ *is such that the ratio* $h^{(n)}/n \to 0$ *as* $n \to \infty$.[3]

(B2) *In model (1), the matrix* $\mathbf{I}^{(n)} - \lambda\mathbf{W}^{(n)}$ *is nonsingular. Moreover, the sequences* $\{\mathbf{W}^{(n)}\}$ *and* $\left\{\left(\mathbf{I}^{(n)} - \lambda\mathbf{W}^{(n)}\right)^{-1}\right\}$ *are uniformly bounded in both row and column sums (Horn and Johnson, 1985).*

(B3) *In the sequence* $\left\{\left(\mathbf{I}^{(n)} - \ell\mathbf{W}^{(n)}\right)^{-1}\right\}$*, matrices* $\left(\mathbf{I}^{(n)} - \ell\mathbf{W}^{(n)}\right)^{-1}$ *are bounded in either row or column sums, uniformly in* ℓ *in an open set parameter space* Λ*. In consequence, the true value of parameter* λ *in model (1) is assumed to belong to the interior of* Λ.

The definition of the parameter space Λ in the above assumption depends on $\mathbf{W}^{(n)}$. For a connectivity matrix with real eigenvalues, Λ may be defined as the open subset $\left(1/\omega_{\max}^{(n)}, 1/\omega_{\min}^{(n)}\right)$, where $\omega_{\min}^{(n)}$ and $\omega_{\max}^{(n)}$ are respectively the minimal and maximal eigenvalues of $\mathbf{W}^{(n)}$. To ensure the same parameter space for λ across different connectivity matrices, it is normalized. Kelejian and

[3]That is, for some real constant c, there exists a finite integer N such that, for all $n \geq N$, $\left|h^{(n)}w_{ij}^{(n)}\right| < c$ for all i, j (see, e.g. White, 1984, p.14).

Prucha (2010) proposes two matrix norms, namely the spectral radius and the minimum between the absolute row and column sum norms, which allow restricting Λ to be the open subset $(-1, 1)$.[4]

Assumption C. *The following sets of constants satisfy the Noether condition*[5]*:*

(i) $\left\{x_{ik}^{(n)}; i = 1, \ldots, n\right\}$ *for each* $k = 1, \ldots, K$*;*

(ii) $\left\{G_{ii}^{(n)}(\lambda); i = 1, \ldots, n\right\}$*;*

(iii) $\left\{\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}; i = 1, \ldots, n\right\}$*;*

(iv) $\left\{G_{ij}^{(n)}(\lambda); i \neq j \in \{1, \ldots, n\}\right\}$.

Assumption D. *The distribution function* F *and density function* f *of the i.i.d. error terms* $\varepsilon_i^{(n)}$ $(i = 1, \ldots, n)$ *should satisfy the following regularity conditions:*

(D1) $F(0) = \int_{-\infty}^{0} f(e)de = 1/2$*, that is, the distribution of the error terms has a zero median.*

(D2) $\mu_f = \int_{-\infty}^{\infty} ef(e)de < \infty$ *and* $0 < \nu_f = \int_{-\infty}^{\infty} e^2 f(e)de < \infty$*;*

(D3) f *is continuous, strictly positive for all points in* $\mathbb{R}$*, and has a continuous first-order derivative* f'*. The second-order derivative of* f *also exists. Moreover,* f *gives a finite Fisher information for location* $\mathcal{I}_f = \int_{\infty}^{\infty} \phi_f^2(e)f(e)de$*, where* $\phi_f(\cdot) = -\frac{f'(\cdot)}{f(\cdot)}$*.*

(D4) f *is strongly unimodal, i.e. the function* ϕ_f *is non-decreasing*[6]*;*

(D5) $\mathcal{K}_f = \int_{-\infty}^{\infty} \phi_f^2(e)ef(e)de < \infty$ *and* $0 < \mathcal{Q}_f = \int_{-\infty}^{\infty} \phi_f^2(e)e^2 f(e)de < \infty$*.*

Let $\mathcal{F} = \{f : \mathbb{R} \to [0, \infty)$ such that f satisfies Assumption D$\}$. Since the density of errors in model (1) is unknown but assumed to belong to $\mathcal{F}$, it plays the role of a nonparametric (infinite dimensional) nuisance. Hence, specification (1) defines a *semiparametric* model

$$\mathcal{E}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)} : f \in \mathcal{F}, \boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}} \in \Theta = \mathbb{R}^{K+1} \times \Lambda\right\}\right).$$

Under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, the residuals $e_i^{(n)}(\boldsymbol{\theta})$ $(i = 1, \ldots, n)$ defined by (3) are i.i.d. with (marginal) density $f \in \mathcal{F}$.

---

[4]The row-normalization is also widely used in applied work. However, unless it is theoretically grounded (see, for instance, Patacchini and Zenou, 2012), or for special cases, such as assigning the same number of neighbors to each observation, this normalization should not be used as it introduces misspecification in the model (see Neumayer and Plümper, 2016).

[5]The set of constants $\{c_\ell; \ell = 1, \ldots, L\}$ is said to satisfy the *Noether condition* if (with the convention $\frac{0}{0} = 0$)

$$\lim_{L\to\infty} \left[\max_{1\leq\ell\leq L} \left(c_\ell - \overline{c}^{(L)}\right)^2 \Big/ \sum_{\ell=1}^{L} \left(c_\ell - \overline{c}^{(L)}\right)^2\right] = 0,$$

where $\overline{c}^{(L)} = \frac{1}{L}\sum_{\ell=1}^{L} c_\ell$.

[6]This is a classical assumption for semiparametric estimation involving ranks.

The rationale behind selecting the zero-median over the traditional zero-mean assumption for the error terms distribution is motivated by the fact that the former allows identification of a simple group of transformations of $\mathbb{R}^n$ that "generates" the semiparametric model and, consequently, to define a semiparametrically efficient estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}}$ using the so-called maximal invariant associated with this group of transformations. This will be further detailed in Section 4. Note that assuming zero-median error has an effect on the interpretation of the marginal effects. Indeed, all interpretations refer to the effect on the conditional median of the dependent variable. While the spillover parameter λ admits a straightforward structural interpretation, particular care is required when assessing how the conditional median responds to changes in covariates: the median is a nonlinear operator, and this nonlinearity must be accounted for both when quantifying the direct, indirect, and total effects of variations in explanatory variables.

3. ULAN property and one-step efficient estimation of the parametric submodel $\mathcal{E}_f^{(n)}$

Let us first consider that the error term density is known and is equal to a specific $f \in \mathcal{F}$. Estimation of the parameters vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}}$ occurs then in the context of the *parametric* submodel

$$\mathcal{E}_f^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}} \in \Theta\right\}\right)$$

of $\mathcal{E}^{(n)}$. In this purely parametric context, maximum likelihood estimation of $\boldsymbol{\theta}$ is straightforward. The log-likelihood function associated to $\mathcal{E}_f^{(n)}$ is

$$\ln L_f\left(\boldsymbol{\theta}\,\middle|\,\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) = \ln\left|\det\left(\mathbf{I}^{(n)} - \lambda \mathbf{W}^{(n)}\right)\right| + \sum_{i=1}^{n} \ln f\left(e_i^{(n)}(\boldsymbol{\theta})\right). \tag{6}$$

Proposition 1. *The sequence of parametric submodels $\mathcal{E}_f^{(n)}$, $n \in \mathbb{N}$, is uniformly locally asymptotically normal (ULAN), with central sequence*

$$\begin{aligned}
\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) &= \begin{pmatrix} \boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) \\ \Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta}) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} \left\{\ln f\left(e_i^{(n)}(\boldsymbol{\theta})\right)\right\} \\ \frac{1}{\sqrt{n}} \frac{\partial}{\partial \lambda} \left\{\ln\left|\det\left(\mathbf{I}^{(n)} - \lambda \mathbf{W}^{(n)}\right)\right|\right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial \lambda} \left\{\ln f\left(e_i^{(n)}(\boldsymbol{\theta})\right)\right\} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right) \mathbf{x}_i^{(n)} \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right) \mathbf{W}_{i\cdot}^{(n)} \mathbf{y}^{(n)} - \frac{1}{\sqrt{n}} \mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) \end{pmatrix},
\end{aligned} \tag{7}$$

where $\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)$ *is the trace of matrix* $\mathbf{G}^{(n)}(\lambda)$*. Under* $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$*, as* $n \to \infty$*,*

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_f(\boldsymbol{\theta})\right),$$

where $\mathbf{I}_f(\boldsymbol{\theta})$ *is the (parametric) Fisher information matrix for* $\boldsymbol{\theta}$ *given by*

$$\mathbf{I}_f(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathbf{I}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) \\ (\mathbf{I}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}))^{\mathrm{T}} & I_{f;\lambda}(\boldsymbol{\theta}) \end{pmatrix},$$

where, using the notations defined in Appendix Appendix A.1,

$$\mathbf{I}_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) = \lim_{n\to\infty}\left\{\mathcal{I}_f \mathbf{C}^{(n)}_{\mathbf{x}}\right\},$$

$$\mathbf{I}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) = \lim_{n\to\infty}\left\{\mathcal{I}_f\, \mathbf{C}^{(n)}_{\mathbf{x},\mathbf{Gx}}(\boldsymbol{\theta}) + \mathcal{K}_f\, \mathbf{C}^{(n)}_{\mathbf{x},\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_f\mu_f\, \mathbf{C}^{(n)}_{\mathbf{x},\mathbf{G},2}(\boldsymbol{\theta})\right\},$$

and

$$\begin{aligned} I_{f;\lambda}(\boldsymbol{\theta}) = \lim_{n\to\infty}\Big\{&\mathcal{I}_f\, C^{(n)}_{\mathbf{Gx}}(\boldsymbol{\theta}) + (\mathcal{Q}_f - 1)\, C^{(n)}_{\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_f\nu_f\, C^{(n)}_{\mathbf{G},2}(\boldsymbol{\theta}) \\ &+ 2\,\mathcal{K}_f\mu_f\, C^{(n)}_{\mathbf{G},3}(\boldsymbol{\theta}) + C^{(n)}_{\mathbf{G},4}(\boldsymbol{\theta}) + \mathcal{I}_f\mu_f^2\, C^{(n)}_{\mathbf{G},5}(\boldsymbol{\theta}) \\ &+ 2\,\mathcal{K}_f\, C^{(n)}_{\mathbf{Gx},\mathbf{G},1}(\boldsymbol{\theta}) + 2\,\mathcal{I}_f\mu_f\, C^{(n)}_{\mathbf{Gx},\mathbf{G},2}(\boldsymbol{\theta})\Big\}. \end{aligned}$$

Appendix Appendix A.2 contains some guidelines for the derivation of these results. The theory developed by Le Cam tells us how, in ULAN models, *parametrically efficient* inference procedures can be based on the central sequence $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$, essentially formalizing the statement that likelihood-based inference is efficient in "regular" models. In particular, if $\widetilde{\boldsymbol{\theta}}^{(n)}$ denotes a $\sqrt{n}$-consistent and *locally discrete* (see Remark 2) estimator of $\boldsymbol{\theta}$, then the one-step estimator

$$\widehat{\boldsymbol{\theta}}^{(n)}_f = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}}\left(\mathbf{I}_f(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1}\boldsymbol{\Delta}^{(n)}_f(\widetilde{\boldsymbol{\theta}}^{(n)})$$

is an *asymptotically efficient* estimator of $\boldsymbol{\theta}$. In other words, under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\theta}}^{(n)}_f$ is asymptotically equivalent to the ML estimator of $\boldsymbol{\theta}$ as $n \to \infty$:

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_f - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, (\mathbf{I}_f(\boldsymbol{\theta}))^{-1}\right).$$

Remark 2.. The estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ is *locally discrete* if, for each $M > 0$, n and $\boldsymbol{\theta}$, the random variable $\sqrt{n}(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta})1_{\left[\sqrt{n}\left\|\widetilde{\boldsymbol{\theta}}^{(n)}-\boldsymbol{\theta}\right\|<M\right]}$ takes only a finite number of values. The assumption that $\widetilde{\boldsymbol{\theta}}^{(n)}$ is $\sqrt{n}$-consistent and locally discrete implies that we may treat it as if it were of the form $\boldsymbol{\theta}^{(n)} + \boldsymbol{\tau}^{(n)}/\sqrt{n}$ for some sequence $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + O(1/\sqrt{n})$ and some bounded and deterministic sequence $\boldsymbol{\tau}^{(n)}$. As indicated in Hallin et al. (2008, Remark 4, p. 398), any $\sqrt{n}$-consistent estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ can quite easily be turned into a locally discrete estimator $\widetilde{\boldsymbol{\theta}}^{(n)}_{\#}$ by mapping each component $\widetilde{\theta}^{(n)}_k$ of $\widetilde{\boldsymbol{\theta}}^{(n)}$ onto

$$\widetilde{\theta}^{(n)}_{\#k} = \mathrm{sgn}(\widetilde{\theta}^{(n)}_k)\frac{\left\lceil c\sqrt{n}\left|\widetilde{\theta}^{(n)}_k\right|\right\rceil}{c\sqrt{n}},$$

where c is an arbitrary positive constant and $\lceil x \rceil$ denotes the smallest integer that is larger than or equal to x. If such discretizations are quite standard in Le Cam's one-step construction of estimators (see for instance Le Cam and Yang, 2000, pp. 125,188), they actually have no effect on practical implementation (where n is fixed), as c can be arbitrarily large, whereas the numerical precision is bounded.

4. Rank-and-sign estimation of $\boldsymbol{\theta}$ in the semiparametric model $\mathcal{E}^{(n)}$

The parametrically efficient procedures based on $\boldsymbol{\Delta}_f^{(n)}(\cdot)$ are generally valid under density f only. The problem lies in the fact that $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$, defined on the basis of the score function $\phi_f(\cdot)$, is no longer appropriately centered under innovation densities $h \neq f$, because $\int_{-\infty}^{\infty} \phi_f(e)h(e)de$ generally differs from zero.

Semiparametric theory usually palliates this, in an optimal way, by projecting the central sequence $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ along the so-called *tangent space* that is associated with variations of the error term density (see Bickel et al., 1993). This projection yields a semiparametrically efficient score function, defining a *semiparametric central sequence* $\boldsymbol{\Delta}_f^{(n)*}(\boldsymbol{\theta})$. However, general results in Hallin and Werker (2003) indicate that in the presence of a suitable group of transformations that "generates" any fixed-$\boldsymbol{\theta}$ submodel of the semiparametric model, a semiparametric central sequence can be obtained more easily and more intuitively — with the additional advantage of distribution freeness — by conditioning $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ on the *maximal invariant* for this group of transformations. In this paper, we rely on this second approach to determine $\boldsymbol{\Delta}_f^{(n)*}(\boldsymbol{\theta})$.

Consider the fixed-$\boldsymbol{\theta}$ submodel $\mathcal{E}_{\boldsymbol{\theta}}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)} : f \in \mathcal{F}\right\}\right)$ of $\mathcal{E}^{(n)}$. It is characterized by (i) the *residual function* $r_{\boldsymbol{\theta}}^{(n)}(\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}) = \left(e_1^{(n)}(\boldsymbol{\theta}), \ldots, e_n^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}} = \mathbf{e}^{(n)}(\boldsymbol{\theta})$ defined by (3) and (ii) a concept of *white noise* with (marginal) unknown density f such that $\mathbf{y}^{(n)}$ has a distribution $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$ if and only if $r_{\boldsymbol{\theta}}^{(n)}(\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)})$ is white noise with (marginal) density f. Denote by $\mathbf{R}^{(n)}(\boldsymbol{\theta}) = \left(R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}}$ and by $\mathbf{s}^{(n)}(\boldsymbol{\theta}) = \left(s_1^{(n)}(\boldsymbol{\theta}), \ldots, s_n^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}}$ the rank vector and the sign vector associated with the residuals $e_1^{(n)}(\boldsymbol{\theta}), \ldots, e_n^{(n)}(\boldsymbol{\theta})$. Define $N_+^{(n)}(\boldsymbol{\theta}) = \sum_{i=1}^n 1_{\left[s_i^{(n)}(\boldsymbol{\theta})=+1\right]}$ and $N_-^{(n)}(\boldsymbol{\theta}) = \sum_{i=1}^n 1_{\left[s_i^{(n)}(\boldsymbol{\theta})=-1\right]}$ as the numbers of positive and negative residuals, respectively. Clearly, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ is uniformly distributed in the set of the $n!$ permutations of $\{1, \ldots, n\}$, $N_-^{(n)}(\boldsymbol{\theta}) + N_+^{(n)}(\boldsymbol{\theta}) = n$ almost surely, and $N_+^{(n)}(\boldsymbol{\theta})$ and $N_-^{(n)}(\boldsymbol{\theta})$ are binomial random variables with 0.5 probability. Moreover, it is well known that the vector of ranks $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ is stochastically independent of the order statistics, and thus of $\mathbf{N}^{(n)}(\boldsymbol{\theta}) = \left(N_-^{(n)}(\boldsymbol{\theta}), N_+^{(n)}(\boldsymbol{\theta})\right)$.

Let T_0 be the set of all continuous, strictly monotonic increasing transformations $t : \mathbb{R} \to \mathbb{R}$ such that $\lim_{e \to \pm\infty} t(e) = \pm\infty$ and $t(0) = 0$. Then, defining, for $t \in T_0$, the transformation $t^{(n)} : \mathbb{R}^n \to \mathbb{R}^n$ by $t^{(n)}(e_1, \ldots, e_n) = \big(t(e_1), \ldots, t(e_n)\big)$, we have that the *group of order preserving transformations* (acting on $\mathbb{R}^n$)

$$T_{0;\boldsymbol{\theta}}^{(n)} = \left\{ \left(r_{\boldsymbol{\theta}}^{(n)}\right)^{-1} \circ t^{(n)} \circ r_{\boldsymbol{\theta}}^{(n)}; t \in T_0 \right\}$$

is a generating group for $\mathcal{E}_{\boldsymbol{\theta}}^{(n)}$. This generating group has for *maximal invariant* the vectors $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{s}^{(n)}(\boldsymbol{\theta})$ of residual ranks and signs, or, equivalently, the vectors $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{N}^{(n)}(\boldsymbol{\theta})$ (see Hallin et al., 2006).

In this context, following the conditioning argument of Hallin and Werker (2003), we should get a semiparametric central sequence by taking the expectation of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ conditionally to $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{N}^{(n)}(\boldsymbol{\theta})$. Proposition 2 characterizes the asymptotic behavior of this central sequence based on ranks and signs, under density $h \in \mathcal{F}$ (with distribution function H), as well as under local alternatives for $\boldsymbol{\theta}$.

4.1. Rank-and-sign central sequence for $\boldsymbol{\theta}$, with reference density f

Let $\varphi_f(u) = \phi_f\left(F^{-1}(u)\right)$ for $u \in (0,1)$ and define, for $i = 1, \ldots, n$:

$$\widetilde{R}_i^{(n)}(\boldsymbol{\theta}) = 1_{\left[s_i^{(n)}(\boldsymbol{\theta})=-1\right]} \left\{ \frac{1}{2} \frac{R_i^{(n)}(\boldsymbol{\theta})}{N_-^{(n)}(\boldsymbol{\theta})+1} \right\} + 1_{\left[s_i^{(n)}(\boldsymbol{\theta})=+1\right]} \left\{ \frac{1}{2} + \frac{1}{2} \frac{R_i^{(n)}(\boldsymbol{\theta}) - (n - N_+^{(n)}(\boldsymbol{\theta}))}{N_+^{(n)}(\boldsymbol{\theta})+1} \right\}.$$

Proposition 2. *Let us define*

$$\boldsymbol{\Delta}_{f;\text{exact}}^{(n)*}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Delta}_{f;\boldsymbol{\beta};\text{exact}}^{(n)*}(\boldsymbol{\theta}) \\ \Delta_{f;\lambda;\text{exact}}^{(n)*}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{f;\boldsymbol{\theta}}^{(n)}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) \,\middle|\, \mathbf{N}^{(n)}(\boldsymbol{\theta}), \mathbf{R}^{(n)}(\boldsymbol{\theta})\right] \\ \mathbb{E}_{f;\boldsymbol{\theta}}^{(n)}\left[\Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta}) \,\middle|\, \mathbf{N}^{(n)}(\boldsymbol{\theta}), \mathbf{R}^{(n)}(\boldsymbol{\theta})\right] \end{pmatrix},$$

where $\mathbb{E}_{f;\boldsymbol{\theta}}^{(n)}$ denotes expectation under the assumption that $e_1^{(n)}(\boldsymbol{\theta}), \ldots, e_n^{(n)}(\boldsymbol{\theta})$ are i.i.d. with density $f \in \mathcal{F}$.

(i) *Under $\mathrm{P}_{h;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$, for any $h \in \mathcal{F}$,*

$$\begin{aligned} \boldsymbol{\Delta}_{f;\text{exact}}^{(n)*}(\boldsymbol{\theta}) &= \boldsymbol{\Delta}_{fh}^{(n)*}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1) \\ &= \boldsymbol{\Delta}_{f;\text{appr}}^{(n)*}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1), \end{aligned} \tag{8}$$

where, for H being the cumulative distribution function associated with h,

$$\boldsymbol{\Delta}_{fh;\boldsymbol{\beta}}^{(n)*}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) + \overline{\mathbf{x}}^{(n)} \frac{2f(0)}{\sqrt{n}} \sum_{i=1}^n s_i^{(n)}(\boldsymbol{\theta}), \tag{9}$$

$$\boldsymbol{\Delta}_{f;\boldsymbol{\beta};\text{appr}}^{(n)*}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(\widetilde{R}_i^{(n)}(\boldsymbol{\theta})\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) + \overline{\mathbf{x}}^{(n)} \frac{2f(0)}{\sqrt{n}} \left(N_+^{(n)}(\boldsymbol{\theta}) - N_-^{(n)}(\boldsymbol{\theta})\right), \tag{10}$$

and

$$\begin{aligned} \Delta_{fh;\lambda}^{(n)*}(\boldsymbol{\theta}) = {} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\left(\mathbf{G}_{i\cdot}^{(n)}(\lambda) - \overline{\mathbf{G}}_{\cdot}^{(n)}(\lambda)\right)\mathbf{X}^{(n)}\boldsymbol{\beta} \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right) F^{-1}\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\left(G_{ii}^{(n)}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right) \\ & + \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \varphi_f\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right) F^{-1}\left(H(e_j^{(n)}(\boldsymbol{\theta}))\right)\left(G_{ij}^{(n)}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right) \\ & + \frac{2f(0)}{\sqrt{n}} \sum_{i=1}^n s_i^{(n)}(\boldsymbol{\theta})\left(\overline{\mathbf{G}}_{\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n}\right), \end{aligned} \tag{11}$$

$$\Delta^{(n)*}_{f;\lambda;\mathrm{appr}}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f \left(\widetilde{R}^{(n)}_i(\boldsymbol{\theta})\right) \left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda) - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\right) \mathbf{X}^{(n)}\boldsymbol{\beta}$$
$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f \left(\widetilde{R}^{(n)}_i(\boldsymbol{\theta})\right) F^{-1}\left(\widetilde{R}^{(n)}_i(\boldsymbol{\theta})\right) \left(G^{(n)}_{ii}(\lambda) - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right)$$
$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varphi_f \left(\widetilde{R}^{(n)}_i(\boldsymbol{\theta})\right) F^{-1}\left(\widetilde{R}^{(n)}_j(\boldsymbol{\theta})\right) \left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right) \tag{12}$$
$$+ \frac{2f(0)}{\sqrt{n}} \left(N^{(n)}_{+}(\boldsymbol{\theta}) - N^{(n)}_{-}(\boldsymbol{\theta})\right) \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n}\right).$$

Moreover,

$$\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{exact/appr}}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}^*_f(\boldsymbol{\theta})\right), \tag{13}$$

with

$$\mathbf{I}^*_f(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}^*_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathbf{I}^*_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) \\ \left(\mathbf{I}^*_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta})\right)^{\mathrm{T}} & I^*_{f;\lambda}(\boldsymbol{\theta}) \end{pmatrix},$$

where

$$\mathbf{I}^*_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) = \lim_{n\to\infty} \left\{\mathcal{I}_f\, \mathbf{C}^{(n)*}_{\mathbf{x}} + (2f(0))^2\, \overline{\mathbf{x}}^{(n)} \left(\overline{\mathbf{x}}^{(n)}\right)^{\mathrm{T}}\right\},$$

$$\mathbf{I}^*_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) = \lim_{n\to\infty} \left\{\mathcal{I}_f\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{Gx}}(\boldsymbol{\theta}) + \mathcal{K}_f\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_f \mu_f\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{G},2}(\boldsymbol{\theta}) + (2f(0))^2\, \overline{\mathbf{x}}^{(n)} \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n}\right)\right\},$$

and

$$I^*_{f;\lambda}(\boldsymbol{\theta}) = \lim_{n\to\infty} \Big\{\mathcal{I}_f\, C^{(n)*}_{\mathbf{Gx}}(\boldsymbol{\theta}) + (\mathcal{Q}_f - 1)\, C^{(n)*}_{\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_f \nu_f\, C^{(n)*}_{\mathbf{G},2}(\boldsymbol{\theta}) + 2\,\mathcal{K}_f \mu_f\, C^{(n)*}_{\mathbf{G},3}(\boldsymbol{\theta}) + C^{(n)*}_{\mathbf{G},4}(\boldsymbol{\theta}) + \mathcal{I}_f \mu_f^2\, C^{(n)*}_{\mathbf{G},5}(\boldsymbol{\theta}) + 2\,\mathcal{K}_f\, C^{(n)*}_{\mathbf{Gx},\mathbf{G},1}(\boldsymbol{\theta}) + 2\,\mathcal{I}_f \mu_f\, C^{(n)*}_{\mathbf{Gx},\mathbf{G},2}(\boldsymbol{\theta}) + (2f(0))^2 \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n}\right)^2\Big\}.$$

(ii) *Moreover, under* $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}}$*, as* $n\to\infty$*, for any* $h\in\mathcal{F}$*,*

$$\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{exact/appr}}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\boldsymbol{\tau}, \mathbf{I}^*_f(\boldsymbol{\theta})\right) \tag{14}$$

with

$$\mathbf{I}^*_{fh}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}^*_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) & \mathbf{I}^*_{fh;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) \\ \mathbf{I}^*_{fh;\lambda,\boldsymbol{\beta}}(\boldsymbol{\theta}) & I^*_{fh;\lambda}(\boldsymbol{\theta}) \end{pmatrix}, \tag{15}$$

where

$$\mathbf{I}^*_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) = \lim_{n\to\infty} \left\{ \mathcal{I}_{fh}\, \mathbf{C}^{(n)*}_{\mathbf{x}} + (2f(0))\,(2h(0))\, \overline{\mathbf{x}}^{(n)} \left(\overline{\mathbf{x}}^{(n)}\right)^{\mathrm{T}} \right\},$$

$$\mathbf{I}^*_{fh;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) = \lim_{n\to\infty} \Big\{ \mathcal{I}_{fh}\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{Gx}}(\boldsymbol{\theta}) + \mathcal{K}_{fh}\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\mu_h\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{G},2}(\boldsymbol{\theta})$$
$$(2f(0))\,(2h(0))\, \overline{\mathbf{x}}^{(n)} \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_h \frac{g^{(n)}(\lambda)}{n} \right) \Big\},$$

$$\mathbf{I}^*_{fh;\lambda,\boldsymbol{\beta}}(\boldsymbol{\theta}) = \lim_{n\to\infty} \Big\{ \mathcal{I}_{fh}\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{Gx}}(\boldsymbol{\theta}) + \mathcal{K}_{hf}\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\mu_f\, \mathbf{C}^{(n)*}_{\mathbf{x},\mathbf{G},2}(\boldsymbol{\theta})$$
$$(2f(0))\,(2h(0)) \left(\overline{\mathbf{x}}^{(n)}\right)^{T} \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n} \right) \Big\},$$

and

$$\begin{aligned} I^*_{fh;\lambda}(\boldsymbol{\theta}) = \lim_{n\to\infty} \Big\{ & \mathcal{I}_{fh}\, C^{(n)*}_{\mathbf{Gx}}(\boldsymbol{\theta}) + (\mathcal{Q}_{fh} - 1)\, C^{(n)*}_{\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\nu_{fh}\, C^{(n)*}_{\mathbf{G},2}(\boldsymbol{\theta}) \\ & + (\mathcal{K}_{fh}\mu_f + \mathcal{K}_{hf}\mu_h)\, C^{(n)*}_{\mathbf{G},3}(\boldsymbol{\theta}) + \mathcal{J}_{fh}\mathcal{J}_{hf}\, C^{(n)*}_{\mathbf{G},4}(\boldsymbol{\theta}) \\ & + \mathcal{I}_{fh}\mu_f\mu_h\, C^{(n)*}_{\mathbf{G},5}(\boldsymbol{\theta}) \\ & + (\mathcal{K}_{fh} + \mathcal{K}_{hf})\, C^{(n)*}_{\mathbf{Gx},\mathbf{G},1}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\,(\mu_f + \mu_h)\, C^{(n)*}_{\mathbf{Gx},\mathbf{G},2}(\boldsymbol{\theta}) \\ & + (2f(0))\,(2h(0)) \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n} \right) \left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_h \frac{g^{(n)}(\lambda)}{n} \right) \Big\}. \end{aligned}$$

Note that $\mathbf{I}^*_{ff}(\boldsymbol{\theta}) = \mathbf{I}^*_{f}(\boldsymbol{\theta})$. The central sequence $\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{exact}}(\boldsymbol{\theta})$ is based on so-called *exact* rank-and-sign scores, whereas $\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\boldsymbol{\theta})$ is defined on the basis of *approximate* rank-and-sign scores (see Appendix Appendix A.3). The proof of Proposition 2 is given in Appendix Appendix A.4 for Part (i) and Appendix Appendix A.5 for Part (ii).

According to the *invariance* properties of signs and ranks, the limiting distribution (13) depends only on the reference density f, and *not* on the actual density h. However, since the invariance property only holds under $\boldsymbol{\theta}$, the behavior of $\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{exact/appr}}(\boldsymbol{\theta})$ depends on h under local alternatives, implying that some choices of reference density f lead to more efficient inference about $\boldsymbol{\theta}$ than others.

4.2. Rank-and-sign estimator of $\boldsymbol{\theta}$, with reference density f

We first consider an estimation procedure that is based on a reference density f, but its properties are investigated under a general "true" density h.[7]

In accordance with the standard Le Cam theory, we formally introduce our rank-and-sign based estimator as a one-step update of an initial $\sqrt{n}$-consistent estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ (in principle, after due

[7]Considering the normal distribution as the reference density f leads to a variant of the Van der Waerden scores, while a logistic reference density yields a variant of the Wilcoxon scores. We refer to these as variants because the original scores were developed in the context of rank-based inference, while our maximal invariant also account for the signs of error terms.

discretization) involving the rank-and-sign based central sequence $\boldsymbol{\Delta}^{(n)*}_{f;\text{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)})$ for which the rank-and-sign approximate scores are associated with a reference density $f \in \mathcal{F}$:

$$\widehat{\boldsymbol{\theta}}^{(n)}_{f;\text{R\&S}} = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}} \left(\widehat{\mathbf{I}}^{(n)*}_{f\widehat{h}_n}(\widetilde{\boldsymbol{\theta}}^{(n)}) \right)^{-1} \boldsymbol{\Delta}^{(n)*}_{f;\text{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}), \tag{16}$$

where $\widehat{\mathbf{I}}^{(n)*}_{f\widehat{h}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})$ is a consistent estimate of $\mathbf{I}^*_{fh}(\widetilde{\boldsymbol{\theta}}^{(n)})$. Actually, $\widehat{\mathbf{I}}^{(n)*}_{f\widehat{h}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})$ is defined from the expression of $\mathbf{I}^*_{fh}(\widetilde{\boldsymbol{\theta}}^{(n)})$ in which we consider the various sums for the fixed size n of the sample (as an approximation of the limit, for $n \to \infty$, of these sums) and in which, using an estimate $\widehat{h}_n$ of the density h,we replace $h(0)$, μ_h, ν_{fh}, $\mathcal{I}_{fh}$, $\mathcal{J}_{fh}$, $\mathcal{J}_{hf}$, $\mathcal{K}_{fh}$, $\mathcal{K}_{hf}$, and $\mathcal{Q}_{fh}$ by consistent estimates $\widehat{h}_n(0)$, $\widehat{\mu}_{\widehat{h}_n}$, $\widehat{\nu}_{f\widehat{h}_n}$, $\widehat{\mathcal{I}}_{f\widehat{h}_n}$, $\widehat{\mathcal{J}}_{f\widehat{h}_n}$, $\widehat{\mathcal{J}}_{\widehat{h}_n f}$, $\widehat{\mathcal{K}}_{f\widehat{h}_n}$, $\widehat{\mathcal{K}}_{\widehat{h}_n f}$, and $\widehat{\mathcal{Q}}_{f\widehat{h}_n}$ (the actual error term density function h and distribution function H are estimated from the residuals $e_i^{(n)}(\widetilde{\boldsymbol{\theta}}^{(n)})$, $i = 1, \ldots, n$).

Proposition 3. *Under* $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$ *, as* $n \to \infty$,

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}^{(n)}_{f;\text{R\&S}} - \boldsymbol{\theta} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\mathbf{0}, \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta}) \right)^{-1} \mathbf{I}^*_{f}(\boldsymbol{\theta}) \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta}) \right)^{-1} \right),$$

where $\mathbf{I}^*_{fh}(\boldsymbol{\theta})$ *is given in Proposition 2 and can be estimated consistently by* $\widehat{\mathbf{I}}^{(n)*}_{f\widehat{h}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})$.

The proof of Proposition 3 is given in Appendix Appendix A.6. Proposition 3 indicates that our rank-and-sign based estimator $\widehat{\boldsymbol{\theta}}^{(n)}_{f;\text{R\&S}}$ remains consistent under a *misspecified* reference density (i.e. when the considered reference density f does not coincide with the true density h). However, its efficiency depends on both the reference and the actual densities (f and h).

By Proposition 3, if the reference density function f coincides with the actual error term density function h, the asymptotic covariance matrix of $\sqrt{n}\,\widehat{\boldsymbol{\theta}}^{(n)}_{f;\text{R\&S}}$ reduces to $(\mathbf{I}^*_f(\boldsymbol{\theta}))^{-1}$. Following Hallin and Werker (2003) and in view of (13), a heuristic argument suggests that $(\mathbf{I}^*_f(\boldsymbol{\theta}))^{-1}$ should, under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, be the semiparametric efficiency bound for the estimation of $\boldsymbol{\theta}$ (see Remark 3). As a consequence, our R&S-based estimator $\widehat{\boldsymbol{\theta}}^{(n)}_{f;\text{R\&S}}$ would be asymptotically semiparametrically efficient under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$.

Remark 3.. The rigorous proof that $\boldsymbol{\Delta}^{(n)*}_{f;\text{exact/appr}}(\boldsymbol{\theta})$ is a semi-parametrically efficient (under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$) central sequence is challenging as we have to show that condition (LF1) of Hallin and Werker (2003, p. 144) holds. More precisely, by defining[8]

$$\mathbf{H}^{(n)}_{\text{lf};f}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{H}^{(n)}_{\text{lf};f;\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ H^{(n)}_{\text{lf};f;\lambda}(\boldsymbol{\theta}) \end{pmatrix} = \boldsymbol{\Delta}^{(n)}_{f}(\boldsymbol{\theta}) - \boldsymbol{\Delta}^{(n)*}_{f;\text{exact}}(\boldsymbol{\theta}), \tag{17}$$

[8]Note that, under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$,

$$\mathbf{H}^{(n)}_{\text{lf};f;\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \phi_f \left(e_i^{(n)}(\boldsymbol{\theta}) \right) - 2f(0) s_i^{(n)}(\boldsymbol{\theta}) \right\} \overline{\mathbf{x}}^{(n)}$$

we have to show that expression (17) can be obtained as the $\boldsymbol{\eta}$-part of the central sequence $\left((\mathbf{H}^{(n)}_{\mathrm{lf};f}(\boldsymbol{\theta}))^{\mathrm{T}}, (\boldsymbol{\Delta}^{(n)}_{f}(\boldsymbol{\theta}))^{\mathrm{T}}\right)^{\mathrm{T}}$ of some LAN parametric submodel $\mathcal{E}^{(n)}_{\mathrm{sub}|f}$ of the form

$$\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}^{(n)}_{\boldsymbol{\eta};\boldsymbol{\theta}} \overset{\mathrm{def}}{=} \mathrm{P}^{(n)}_{f_{\boldsymbol{\eta}};\boldsymbol{\theta}} : \boldsymbol{\eta} \in (-1,1)^{K+2} \text{ such that } f_{\mathbf{0}} = f, \boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}} \in \Theta\right\}\right).$$

This question remains open and could benefit from further exploration in future research.

4.3. Fully semiparametric rank-and-sign estimator of $\boldsymbol{\theta}$

Our objective is to define a *fully* semiparametrically highly efficient estimator of $\boldsymbol{\theta}$ in $\mathcal{E}^{(n)}$, i.e. an estimator that is highly efficient regardless of the actual error density.

For this, we should be able to neutralize the dependence on f in the definition of the R&S estimator.

Proposition 4. *Let $f \in \mathcal{F}$ be the (unknown) actual error density and $\widehat{f}_n$ be a consistent estimator of f defined from the residuals $e^{(n)}_i(\widetilde{\boldsymbol{\theta}}^{(n)})$, $i = 1, \ldots, n$. Let $\widehat{\mathbf{I}}^{(n)*}_{\widehat{f}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})$ be a consistent estimate of $\mathbf{I}^*_f(\widetilde{\boldsymbol{\theta}}^{(n)})$ defined by considering the size n as fixed and by replacing the unknown quantities $f(0)$, μ_f, $\mathcal{I}_f$, $\mathcal{K}_f$, and $\mathcal{Q}_f$ by consistent estimates (based on $\widehat{f}_n$) in the expression of $\mathbf{I}^*_f(\widetilde{\boldsymbol{\theta}}^{(n)})$. Denote by $\widehat{\Delta}^{(n)*}_{\widehat{f}_n;\mathrm{appr}}(\boldsymbol{\theta})$ the central sequence obtained by replacing, in $\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\boldsymbol{\theta})$, $f(0)$ and μ_f by consistent estimates (based on $\widehat{f}_n$), and the score functions $\varphi_f(u) = \phi_f\left(F^{-1}(u)\right)$ and $\psi_f(u) = F^{-1}(u)$ by some estimates $\widehat{\varphi}_{f,n}(u)$ and $\widehat{\psi}_{f,n}(u)$. If $\widehat{\varphi}_{f,n}$ and $\widehat{\psi}_{f,n}$ are estimators of φ_f and ψ_f such that, for $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + O(1/\sqrt{n})$,*

$$\widehat{\Delta}^{(n)*}_{\widehat{f}_n;\mathrm{appr}}(\boldsymbol{\theta}^{(n)}) = \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\boldsymbol{\theta}^{(n)}) + o_{\mathrm{P}}(1) \tag{18}$$

under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$, then the R&S-based estimator

$$\widehat{\boldsymbol{\theta}}^{(n)}_{\mathrm{R\&S}} = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}}\left(\widehat{\mathbf{I}}^{(n)*}_{\widehat{f}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \widehat{\Delta}^{(n)*}_{\widehat{f}_n;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) \tag{19}$$

is fully semiparametrically efficient in the sense that, for all $f \in \mathcal{F}$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\mathrm{R\&S}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{I}^*_f(\boldsymbol{\theta})\right)^{-1}\right)$$

under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$.

---

and

$$H^{(n)}_{\mathrm{lf};f;\lambda}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left\{\phi_f\left(e^{(n)}_i(\boldsymbol{\theta})\right) - 2f(0)s^{(n)}_i(\boldsymbol{\theta})\right\}\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right.$$
$$\left. + \left\{\phi_f\left(e^{(n)}_i(\boldsymbol{\theta})\right)\sum_{\substack{j=1\\ j\neq i}}^{n} e^{(n)}_j(\boldsymbol{\theta}) - 2f(0)s^{(n)}_i(\boldsymbol{\theta})\mu_f\right\}\frac{g^{(n)}(\lambda)}{n(n-1)}\right] + o_{\mathrm{P}}(1).$$

In Appendix Appendix A.7, we provide a *sufficient* condition that characterizes the consistency requirements for the estimators $\widehat{\varphi}_{f,n}$ and $\widehat{\psi}_{f,n}$ under which (18) holds. We also briefly discuss the existence of such estimators. Establishing that this sufficient condition is satisfied in our setting is left for future work. Appendix Appendix A.8 contains the proof of Proposition 4.

5. Practical implementation

Implementation of our proposed R&S estimator requires first to select a preliminary estimator. Next, we have to estimate the unknown error density function relying on the preliminary residuals and finally compute the central sequence and information matrix to obtain the one-step estimator. Despite the complexity of the formulas presented in the previous sections, they are all explicitly defined and the integrals can be numerically approximated. In this section, we thus focus on the first two requirements and then present a refinement procedure that can be used in finite samples to potentially improve efficiency. We note, as illustrated by the simulation study, that the efficiency gains from refining the one-step estimator depend primarily on the choice of preliminary estimator.

5.1. The preliminary estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ *of* $\boldsymbol{\theta}$

As mentioned in Section 4.2, we only require that the preliminary estimator of $\boldsymbol{\theta}$ be $\sqrt{n}$-consistent. As such, we may choose between the TSLSE of Kelejian and Prucha (1998); Bramoullé et al. (2009), an efficient GMME of Lee (2007); Liu et al. (2010), or the QMLE of Lee (2004).[9] Even though the efficiency of these estimators differs, this initial difference can be compensated by the proposed refinement procedure. Finally, as the preliminary estimated intercept represents the mean of the residuals, we correct it to ensure that residuals have zero-median.

5.2. Data dependent variable-bandwidth kernel estimation of error's density

The second requirement to implement our proposed estimator is to consistently estimate the density f of the error term. Theoretically, as discussed in section 4.3 and Appendix Appendix A.7, one should rely on an estimator that considers only the negative residuals for the left-hand side estimation of the density and only the positive residuals for the right-hand side. However, to the best of our knowledge, such an estimator does not exist. To address this, we rely on a variable-bandwidth (Gaussian) kernel estimation. When point concentration varies significantly across locations, as is the case for skewed and/or heavy-tailed distributions, a fixed bandwidth estimator may be problematic, as it could result in excessive smoothing and loss of detail in highly populated areas and under-smoothing and excess variability in regions with low point density. Additionally, our data dependent variable bandwidth estimator mimics, to some extent, the theoretically required property of considering only negative or positive residuals for estimating each part of the density. The only problem arises for the estimation of the density close to the median of the residuals. However, as this area is quite dense, the bandwidth will be small, reducing the number of residuals with the "wrong" signs to be used.

The formula used for the variable-bandwidth is $bw_i = bw \times \left\{ M_{\text{geom}} / \widehat{f}_{n,\text{prel}}(e_i) \right\}^{0.5}$, where M_{geom} is the geometric mean of a preliminary fixed bandwidth (bw) density estimate $\widehat{f}_{n,\text{prel}}$ evaluated at

[9] Naturally, when no exogenous regressors are present or if all are irrelevant, the TSLSE cannot be used since there is no internal instrument available, and QMLE or GMME should be considered as a starting point.

each point (see Abramson, 1982; Van Kerm, 2003). The bandwidth of the preliminary density estimator is chosen according to a Silverman rule of thumb, namely $bw = 0.9 \min\left(\widehat{\sigma}, \frac{\text{IQR}}{1.349}\right) n^{-\frac{1}{5}}$, where $\widehat{\sigma}$ corresponds to the estimated standard deviation and IQR is the fitted interquartile range.

5.3. Refinement procedure

Since the one-step estimator is based on asymptotic results, we propose a refinement procedure for the estimator $\widehat{\boldsymbol{\theta}}^{(n)}_{\text{R\&S}}$ presented in (19) to improve its performance in small samples. The residuals $e_i^{(n)}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\text{R\&S}}\right)$ $(i = 1, \ldots, n)$ are computed and used to estimate once again the underlying density function f and to evaluate the log-likelihood function in (6) at the parameter value $\widehat{\boldsymbol{\theta}}^{(n)}_{\text{R\&S}}$. Then we update $\widehat{\boldsymbol{\theta}}^{(n)}_{\text{R\&S}}$ applying (19) in which $\widehat{\boldsymbol{\theta}}^{(n)}_{\text{R\&S}}$ acts as a preliminary estimator, and we evaluate (6) for this new estimate of $\boldsymbol{\theta}$. This iterative process stops (usually very fast) when the log-likelihood value stops increasing. Our simulation study indicates that the efficiency gains obtained from this refinement procedure depend on the initial estimator considered.

6. Simulations

The experimental design considered is

$$y_i^{(n)} = \beta_0 + \beta_1 x_i^{(n)} + \lambda \mathbf{W}^{(n)}_{i\cdot} \mathbf{y}^{(n)} + \varepsilon_i^{(n)}, \quad i = 1, \ldots, n \tag{20}$$

where the $x_i^{(n)}$'s are generated once (and kept constant over all the simulations) from a standard normal and $\beta_0 = \beta_1 = 1$.

The interaction scheme we consider is binary 10 nearest neighbors constructed from randomly set coordinates.[10] This matrix has been normalized using the spectral radius norm of Kelejian and Prucha (2010). As such, the parameter λ takes values from -2.7 to 0.7, increasing in steps of 0.2, and also includes the value 0. These values span the parameter space Λ, defined in section 2.2.

We evaluate the finite-sample performance of five estimators: the QMLE developed by Lee (2004), the efficient GMME of Liu et al. (2010), our proposed agnostic rank-and-sign (R&S) estimator (agnostic in the sense that we make no assumptions in advance about the distribution of the error term, which is therefore entirely estimated from the data), and two constrained, parametric versions of this rank-and-sign approach, where a specific reference density f is considered (see section 4.2). The first, denoted VdW, assumes a normal distribution as the reference density f in the expression (16), leading to the rank-and-sign version of the Van der Waerden scores, optimal under normality. The second parametric rank-and-sign estimator, denoted Wil, relies on a logistic reference density, leading to the rank-and-sign version of the Wilcoxon scores, optimal under logistic errors.[11] These parametric rank-and-sign estimators are standard benchmarks in rank-based semiparametric inference. Comparing against them highlights the gains from using a kernel-based estimate of the error density rather than imposing a parametric score.

The simulations are conducted across six distributions for the error term, sketched in Figure 1: the standard Normal distribution, the Student distribution with 2 degrees of freedom, a shifted

[10] As for the $x_i^{(n)}$'s, the coordinates have been generated once and kept constant across designs.
[11] Appendix Appendix A.9 develops the associated formulas.

LogNormal(0,1)-distribution, a Logistic(0,1)-distribution, a Beta-Student mixture, and an asymmetric bimodal Normal mixture.[12] The last two distributions have been considered to test the robustness of our estimator with respect to violations of its fundamental assumptions. As such, the Beta-Student mixture is discontinuous, while the strong unimodality assumption underlying the R&S procedure is not met for the bimodal Normal mixture.

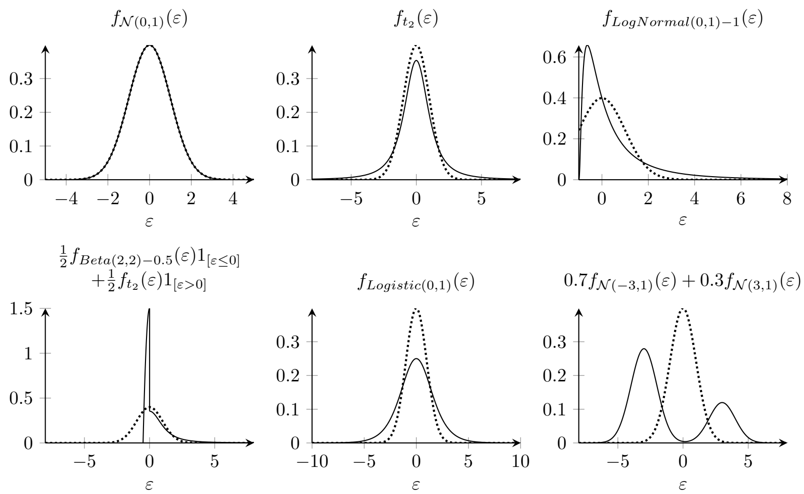


Figure 1: Solid lines: Density functions of the distributions considered in simulations – Dotted line: The standard normal density function

In total, 432 alternative scenarios are considered, and each of them has been replicated 1000 times. The simulation setup is run for 2 sample sizes: $n = 300$ and $n = 900$.[13]

We compare the bias of the competing estimators using the median difference of the estimated coefficients to the true values, and we rely on the interquartile range (divided by 1.349 to ensure Gaussian consistency towards the standard deviation) of the point estimates as a measure of dispersion. The results related to the constant term are not reported, as its definition is not comparable across estimators.

The TSLSE is computed using the 2 first-order neighborhood's characteristics as instruments for the endogenous effects (Kelejian and Prucha, 1998; Bramoullé et al., 2009). The efficient GMM estimator of Liu et al. (2010) is obtained by an iterated procedure used to refine the estimation of the covariance matrix of moment conditions. QMLE serves as preliminary estimator for the R&S estimators (agnostic and parametric), and we report results allowing for refinements of the one-step

[12]For the standard Normal distribution, we consider the MLE rather than QMLE and the moment conditions that lead to the efficient GMM under normality. Additionally, the mixture of Normal distributions is not zero-median but this is handled by the residuals recentering (around the median) and the presence of a constant term in the model.

[13]All simulations have been run with Matlab R2019a on the calculation center of the Université de Lille (Mésocentre de Calcul Scientifique Intensif de l'Université de Lille). Moreover, the proposed R&S estimator has been programmed in Stata, Matlab, and R softwares.

estimator. Finally, in the core of the paper, we only present the results associated with the largest sample size ($n = 900$).[14]

Analysis of bias

Figure 2 presents the bias of all estimators of λ across the range of values considered for this parameter. The bias of the R&S estimator is minimal in all configurations and remains strikingly consistent over the different values assumed for λ. The other estimators also perform satisfactorily for all distributions (within ± 0.02), but their bias depends more strongly on λ. A key advantage of the R&S estimator is therefore that its bias is largely unaffected by the true value of λ.

The bias in the estimation of β_1, shown in Figure 3, is minimal for all estimators in all setups (within ± 0.01), regardless of the value of λ. The R&S estimator performs consistently well for all distributions.

Relative Dispersion

Figure 4 summarizes the dispersion of $\widehat{\lambda}$ for all the estimators considered. The figure exhibits an inverted U-shape pattern across the values of λ for all estimators, with the maximum dispersion occurring around $\lambda \approx -1$ and decreasing towards the extremes of the parameter space. This pattern is not surprising and is analogous to the well-known behavior of the correlation coefficient estimator, whose variance is highest for intermediate values and decreases as the true correlation approaches the boundaries of the parameter space.

Under Normal errors, all estimators perform similarly, with ML and GMM enjoying a slight advantage, as expected, since QML achieves parametric efficiency under correct specification, while GMM uses the first order conditions of ML as moment conditions, leading to an asymptotic equivalent estimator. However, under non-normal distributions, the R&S estimator clearly dominates with the lowest dispersion. The efficiency gains are substantial, particularly under heavy-tailed or asymmetric distributions (Student t_2 and Shifted LogNormal(0,1)).

Under Logistic errors, all estimators perform comparably, with the logistic-based rank-and-sign estimator showing strong performance as expected given its optimality under this distribution.

A striking finding is that the R&S estimator achieves the lowest dispersion even under the asymmetric Bimodal Normal and Beta-Student mixtures. This suggests a considerable robustness of the proposed procedure with respect to violations of its core assumptions.

The dispersion of the estimators of the slope parameter β_1 are sketched in Figure 5. The observed patterns are qualitatively different from those observed for $\widehat{\lambda}$. The dispersion is relatively flat across λ values for all estimators, without the pronounced inverted U-shape.

Under Normal and Logistic errors, all estimators exhibit nearly identical dispersion, with all the curves overlapping almost perfectly. Under all other distributions, however, the R&S estimator consistently achieves the lowest dispersion.

Role of the refinement procedure

The results presented above are based on the refined R&S estimator described in Section 5.3, which updates the one-step estimator until the log-likelihood ceases to increase. The supplementary file reports additional simulations using only a single one-step estimator. While the unrefined

[14] In a supplementary file, we also report the results for $n = 300$, when the TSLSE is used as the preliminary estimator and the results without refinement of the one-step estimator.

estimator remains consistent and displays similar qualitative behavior, the iterative procedure yields modest but systematic reductions in both bias and dispersion, particularly for the small sample size ($n = 300$). In our experiments, convergence typically occurred within two to four iterations, indicating that the computational overhead of refinement is negligible. We therefore recommend the refined version for applied work.

The efficiency gains obtained from this refinement procedure also depend on the initial estimator considered. Relying on QMLE, a single application of the one-step estimator yields an estimate close to the fully refined estimator. By contrast, starting from TSLSE, multiple iterations are typically needed to attain comparable precision, even though convergence occurs rapidly. Upon full convergence, both initializations lead to estimators with equivalent properties. Interestingly, in the simulation study, the number of iterations required for convergence when starting from TSLSE increases with λ.

Robustness to the preliminary estimator

The supplementary file also examines the sensitivity of the R&S estimator to the choice of preliminary estimator. Relying on the TSLSE rather than on the QMLE estimator yields virtually identical results for both bias and dispersion (after refining the one-step based on TSLSE), confirming that the one-step correction effectively eliminates the influence of the initial estimator, as predicted by the asymptotic theory.

Reliability of Standard Errors

Figures 6–8 assess the reliability of the asymptotic standard errors by comparing the median of estimated standard errors (dashed line) with the actual sampling dispersion measured by the interquartile range (solid line). For valid inference, these two quantities should coincide.

Figure 6 shows the excellent reliability of asymptotic standard errors of the R&S estimator across all error distributions. These results indicate that the asymptotic variance formula provides reliable inference for the R&S estimator.

By contrast, the estimator based on the Van der Waerden R&S scores exhibits systematic overestimation, as indicated in Figure 7. The median standard error consistently lies above the IQR dispersion across all distributions. The gap is smallest under Normal errors, as expected since VdW is optimal in this case, and for Logistic errors.

The Wilcoxon estimator displays a pattern similar to Van der Waerden scores, with systematic overestimation of dispersion (see Figure 8). The smallest gap between estimated and actual dispersion occurs under Logistic errors, consistent with the optimality of Wilcoxon scores for this distribution. As such, the asymptotic standard errors seem to be conservative as soon as the distribution of the error term deviates from the Logistic.

As such, the simulation results lead to the following conclusions:

(i) **Bias:** The R&S estimator exhibits negligible bias for both λ and β_1 for all error distributions and is notably less sensitive to the true value of λ than competing estimators.

(ii) **Efficiency:** Under non-normal errors, the R&S estimator achieves the lowest dispersion for both $\widehat{\lambda}$ and $\widehat{\beta}_1$, with substantial efficiency gains under heavy-tailed and mixture distributions. Under normality, R&S performs comparably to existing parametric estimators, sacrificing only a small amount of efficiency.

(iii) **Robustness:** The R&S estimator maintains excellent performance even under the Bimodal Normal and Beta-Student mixtures, which violate its core assumptions, suggesting the robustness of the estimator.

(iv) **Inference:** Asymptotic standard errors for the R&S estimator provide reliable inference. However, we observe discrepancies between the real and theoretical dispersion for rank-and-sign estimators based on a specific (symmetric) reference density of the errors (Van der Waerden and Wilcoxon).

(v) **Sensitivity to λ:** The dispersion of the R&S estimator is less sensitive to the true value of λ than competing estimators, a desirable property given that the true spillover parameter is unknown in applications.

Table 1 summarizes the main findings of the simulation study.

Table 1: Summary of estimator performance

Criterion	Best under Normal	Best under Non-Normal	Most Robust
Bias of $\widehat{\lambda}$	None	R&S	R&S
Bias of $\widehat{\beta}_1$	None	None	R&S
Dispersion of $\widehat{\lambda}$	QML/GMM	R&S	R&S
Dispersion of $\widehat{\beta}_1$	None	R&S	R&S
S.E. reliability	All adequate	R&S	R&S

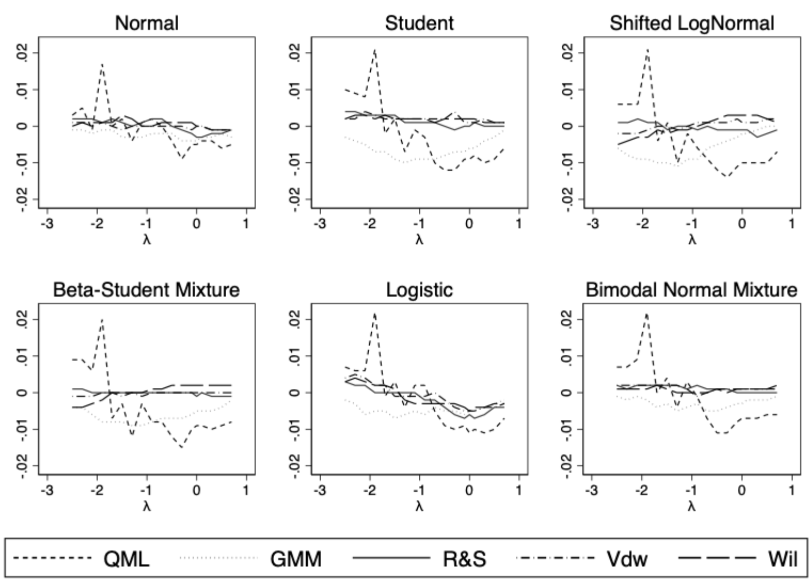


Figure 2: Bias of $\widehat{\lambda}$, $n = 900$

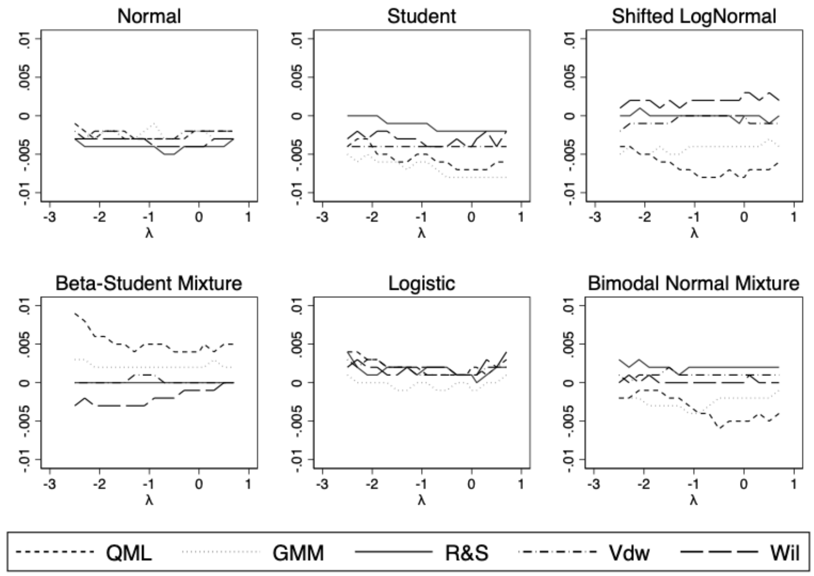


Figure 3: Bias of $\widehat{\beta}_1$, $n = 900$

7. Empirical illustration

To illustrate the practical utility of the R&S estimation framework, we rerun a trade regression initially developed by Behrens et al. (2012) (BEK hereafter). These authors theoretically derive a trade model using spatial econometric techniques to assess the effect of the Canada-U.S. border on trade flows. Their sample includes 30 US states and 10 Canadian regions, leading to a sample of size $n = 1600$. In one of their intermediary results (Table III, p.788), they report the estimation of a SAR specification, shown in (21). Our aim is to use their dataset as an empirical setting to assess how sensitive the parameter estimates are to the choice of estimation method, not to validate or challenge their theoretical claims, especially since their final empirical specification differs from the one we adopt here.

$$\ln(Z_{ij}) = \beta_0 + \beta_1 d_{ij} + \beta_2 \ln(w_i) + \beta_3 b_{ij} + \lambda \sum_{\substack{k=1 \\ k \neq i}}^{n} \frac{L_k}{L} \ln(Z_{kj}) + \varepsilon_{ij}, \tag{21}$$

where Z_{ij} are the GDP-standardized manufacturing exports from the region i to j, d_{ij} is the great circle distance (in kilometers) between regional and provincial capitals. The internal distance is measured as $\frac{2}{3}\sqrt{\text{surface}_i/\pi}$, with surface$_i$ denotes the region's surface in square kilometers (see Redding and Venables, 2004).[15] The regression also includes w_i, which measures the average hourly manufacturing wage in region i and the dummy variable b_{ij}, which takes a value of 1 if region i

[15]Behrens et al. (2012) consider also alternative measures of internal distances as robustness analysis. However, in this illustration, we focus on the first definition but all the results hold for the 2 other definitions.

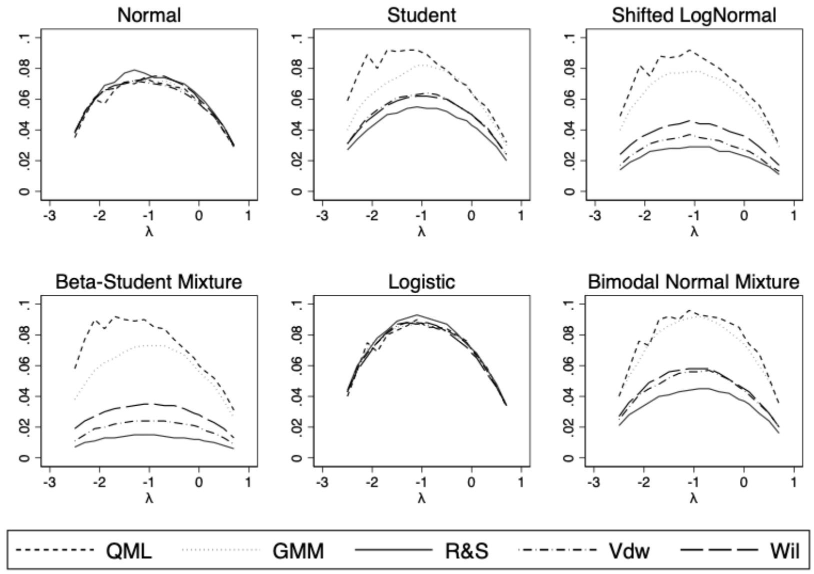


Figure 4: Dispersion of $\widehat{\lambda}$, $n = 900$

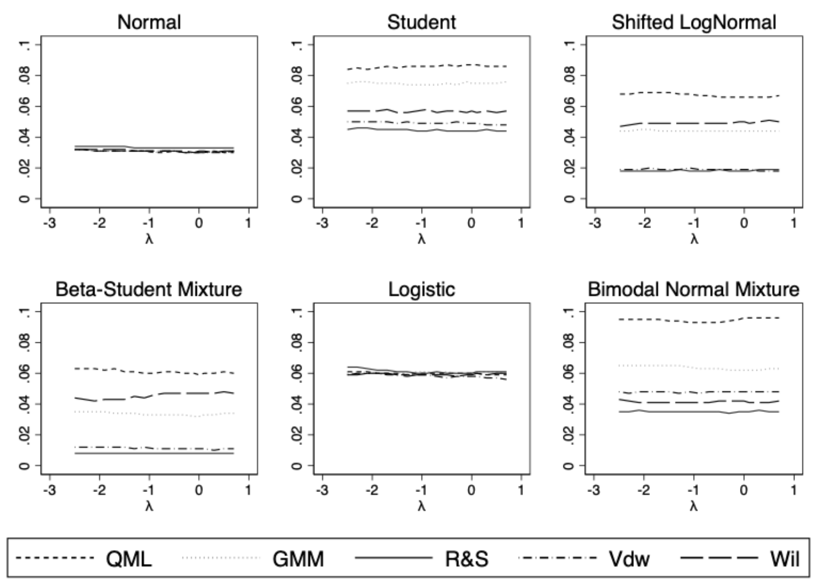


Figure 5: Dispersion of $\widehat{\beta}_1$, $n = 900$

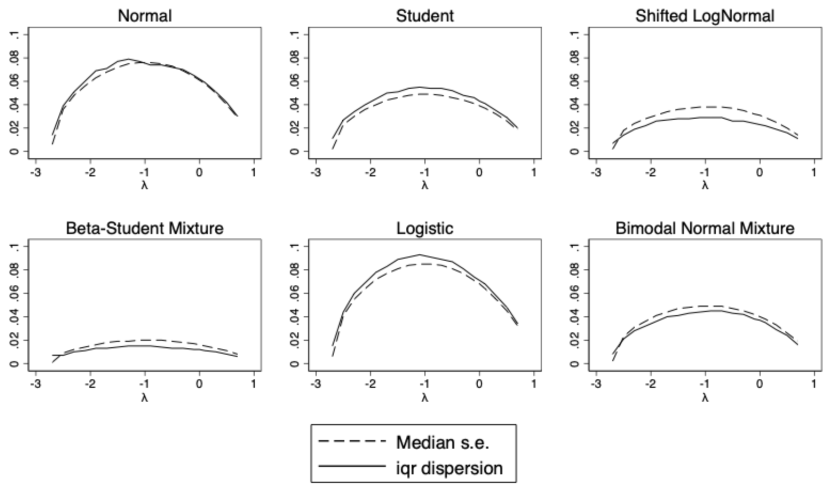


Figure 6: R&S estimator, $n = 900$

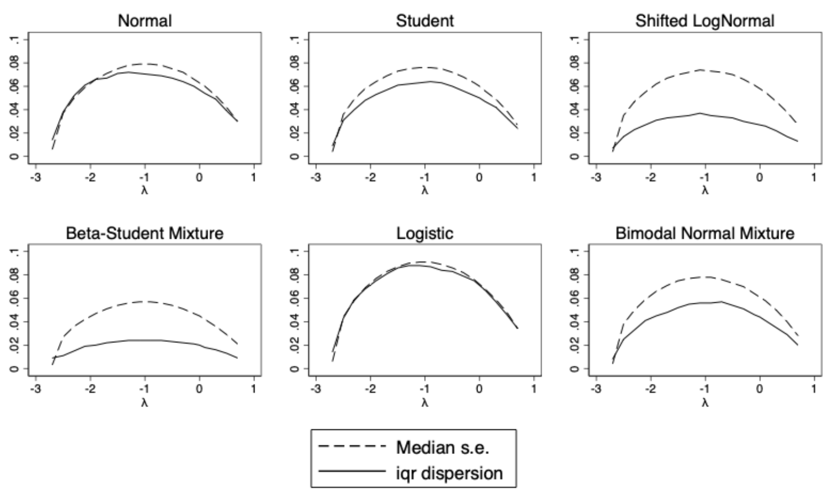


Figure 7: Van der Waerden scores, $n = 900$

belongs to Canada while j is part of the U.S. or conversely and 0 otherwise.[16] Finally, the exports from i to j depend on the exports of other regions k to region j, where the connectivity between k and j is constructed from the share of population in region k over the total sample (L_k/L),

[16]To account for zero flow observations in their logarithmic bilateral export model, BEK add a unit to these flows and introduce a dummy variable to identify the original zero flows among the regressors (not reported in the model specification), yielding a total of five regression parameters and a constant to estimate.

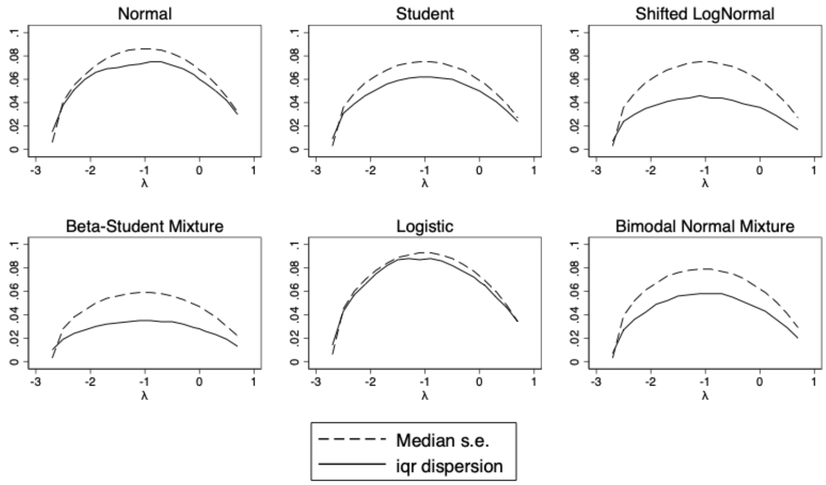


Figure 8: Wilcoxon scores, $n = 900$

and further normalized by its spectral radius. Their characterization of the interaction scheme is directly derived from the trade model and thus avoids the complex question of selecting the relevant neighborhood for each unit.

Our objective here is to compare the estimation results of the SAR model obtained by QML under the normality assumption, as estimated by BEK, to the R&S estimator. We also provide a comparison with the GMM estimator of Liu et al. (2010) and the two rank-and-sign estimators constructed, respectively, from the Normal and Logistic reference densities, which are potentially more efficient than QML.

The first column in Table 2 presents the estimation results reported in BEK, based on QML under normality. We observe statistically insignificant spillover effects (parameter λ).[17] In the left panel of Figure 9, we show the qqplot with respect to the Normal distribution of the residuals based on the QML estimation. The tails of the empirical distribution differ greatly from those of a Gaussian distribution, indicating that a substantial gain in efficiency can be achieved. This gain can be visualized by looking at the right panel of Figure 9, which shows the histogram of the residuals computed from the R&S estimator, the associated estimated density (in red), and the normal density as the reference (in black).[18] We clearly observe that the estimated density substantially deviates from the Gaussian density.

Column 2 of Table 2 presents the estimation results obtained by efficient GMM, which are slightly more precise than QML, although qualitatively similar. The last three columns of Table 2 report the results of three R&S estimators. In column 3 we consider the scores of Van der Waerden (normal distribution), in column 4 the Wilcoxon scores (logistic distribution), while in column 5

[17]Behrens et al. (2012) show that this coefficient should be negative. This insignificance may come from the fact that cross-section dependence in the error terms is omitted in this specification.

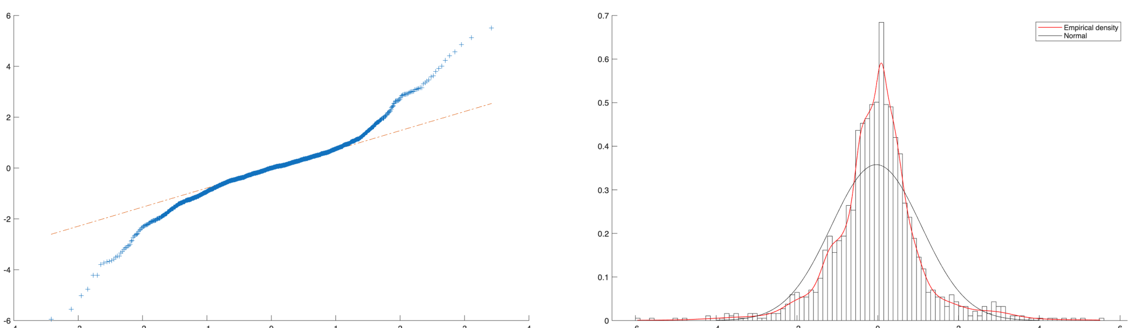
[18]The density has been estimated relying on a Gaussian kernel and a data dependent variable-bandwidth whose details are presented in section 5.2.

we estimate the model relying on the agnostic R&S estimator (i.e. without assuming any reference density for the errors). For the three proposed R&S estimators, we observe an increase in the point estimate of λ, as well as an increase in statistical significance. For the estimator based on Wilcoxon's scores, this result seems driven by the increase in the point estimate, as the standard error is similar to that of QML. Nevertheless, for the agnostic R&S estimator, we observe a markedly decrease in all standard errors.

Table 2: Comparison of estimation results

Dependent variable: $\ln(Z_{ij})$					
	QML	GMM	VdW scores	Wilcoxon Scores	R&S
Constant	-13.740	-12.248	-13.005	-12.226	-12.119
	(0.720)	(0.706)	(0.664)	(0.789)	(0.464)
	[-19.083]	[-17.352]	[-19.602]	[-15.501]	[-26.098]
d_{ij}	-1.331	-1.379	-1.318	-1.279	-1.302
	(0.038)	(0.037)	(0.035)	(0.044)	(0.026)
	[-35.258]	[-37.681]	[-37.344]	[-28.884]	[-50.638]
$\ln(w_i)$	-1.238	-1.785	-1.276	-1.217	-1.284
	(0.179)	(0.173)	(0.167)	(0.209)	(0.122)
	[-6.925]	[-10.294]	[-7.638]	[-5.814]	[-10.563]
b_{ij}	-1.067	-0.831	-1.091	-1.183	-1.261
	(0.067)	(0.065)	(0.062)	(0.078)	(0.045)
	[-16.058]	[-12.897]	[-17.545]	[-15.156]	[-27.847]
λ	0.030	0.043	0.063	0.116	0.113
	(0.030)	(0.023)	(0.027)	(0.032)	(0.019)
	[1.003]	[1.473]	[2.304]	[3.669]	[6.041]
Rel. eff	1	1.059	1.153	0.757	2.226

Notes: standard errors between parentheses and t-stats between square brackets. Rel. eff. computes the relative efficiency of each estimator compared to QML.



(a) Normal QQ plot

(b) Residuals density

Figure 9: Residuals diagnostics

Finally, the bottom panel of Table 2 presents the relative efficiency of each of the four estimation methods with respect to QML. This measure of relative efficiency is computed as (see Serfling, 2011):

$$\text{Rel. eff}_q = \left(\frac{\det\left(\mathbf{I}(\boldsymbol{\theta})^{-1}_{\text{QML}}\right)}{\det\left(\mathbf{I}(\boldsymbol{\theta})^{-1}_{q}\right)} \right)^{1/K_{comp}}, \quad q = \text{QML}, \text{GMM}, \text{VdW}, \text{Wil}, \text{R\&S}, \quad K_{comp} = 5.$$

with $\mathbf{I}(\boldsymbol{\theta})_q^{-1}$ being the asymptotic covariance matrix of the q^{th} estimator and K_{comp} the number of estimates to be compared.[19].

In this empirical application, GMM is slightly more efficient than QML (around 6%), while the agnostic R&S estimator is approximately two times more efficient than QML. Intermediate efficiency gains are observed when one uses the R&S estimator constructed from the Van der Waerden scores, while results indicate relatively lower efficiency for the R&S estimator based on the Wilcoxon scores. These last two results can be explained by the fact that the density of the residuals indicates some asymmetry, while symmetric reference densities (Normal for Van der Waerden and Logistic for Wilcoxon) are imposed, leading to a loss of efficiency, as predicted by Section 4.1.

8. Conclusions

Due to its inherent simultaneity, the SAR model cannot generally be estimated by ordinary least squares and calls for more advanced procedures such as two-stage least squares, generalized method of moments, or quasi-maximum likelihood.

When the error distribution is known (and possesses the appropriate differentiability properties), the maximum likelihood framework provides the most efficient estimator. However, if the distribution of the errors is unknown, maximum likelihood estimation becomes infeasible. In such cases, the quasi-maximum likelihood method under normal errors still produces consistent estimators, although not efficient.

In this paper, we develop a new estimator based on the concept of Local Asymptotic Normality and previous research by Hallin and Werker (2003) and Hallin et al. (2006, 2008). This estimator, constructed from the ranks and signs of the residuals of a preliminary $\sqrt{n}$-consistent estimator, should be asymptotically *semiparametrically* efficient. Monte Carlo experiments show that it generally performs better than the other methods considered, once the assumption of a normal error distribution is relaxed.

In future research, we plan to relax the i.i.d. errors assumption and propose R&S estimators that address heteroskedasticity and/or clustering.

References

Abramson, I. S. (1982), 'On bandwidth variation in kernel estimates–a square root law', *Annals of Statistics* **10**, 1217–1223.

Behrens, K., Ertur, C. and Koch, W. (2012), 'Dual gravity: using spatial econometrics to control for multilateral resistance', *Journal of Applied Econometrics* **27**, 773–794.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993), *Efficient and Adaptive Statistical Inference for Semiparametric Models*, John Hopkins University Press, Baltimore.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**, 41–55.

[19]We do not include the constant here while the dummy that identifies the original zero flows is accounted for (see footnote 16).

Gourieroux, C., Monfort, A. and Trognon, A. (1984), 'Pseudo maximum likelihood methods: Theory', *Econometrica* **52**, 681–700.

Hájek, J. and Sidák, Z. (1967), *Theory of Rank Tests*, Academic Press, New York.

Hallin, M., Vermandele, C. and Werker, B. J. M. (2006), 'Serial and nonserial sign-and-rank statistics: Asymptotic representation and asymptotic normality', *The Annals of Statistics* **34**(1), 254–289.

Hallin, M., Vermandele, C. and Werker, B. J. M. (2008), 'Semiparametrically efficient inference based on signs and ranks for median-restricted models', *Journal of the Royal Statistical Society, Serie B* **70**(2), 389–412.

Hallin, M. and Werker, B. J. M. (2003), 'Semi-parametric efficiency, distribution-freeness and invariance', *Bernouilli* **9**(1), 137–165.

Horn, R. A. and Johnson, C. R. (1985), *Matrix Analysis*, Cambridge University Press, London.

Kelejian, H. H. and Prucha, I. R. (1998), 'A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics* **17**, 99–121.

Kelejian, H. H. and Prucha, I. R. (1999), 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review* **40**, 509–533.

Kelejian, H. H. and Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.

Le Cam, L. (1960), 'Locally asymptotically normal families of distributions', *University of California Publications in Statistics* **3**, 37–98.

Le Cam, L. and Yang, G. L. (2000), *Asymptotics in Statistics; Some Basic Concepts, 2nd Edition*, Springer-Verlag, New York.

Lee, L.-f. (2002), 'Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models', *Econometric Theory* **18**, 252–277.

Lee, L.-F. (2003), 'Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances', *Econometric Reviews* **22**, 307–335.

Lee, L.-f. (2004), 'Asymptotic distributions of Quasi-Maximum Likelihood estimators for spatial autoregressive models', *Econometrica* **72**, 1899–1925.

Lee, L.-f. (2007), 'GMM and 2SLS estimation of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **137**, 489–514.

Liu, X., Lee, L.-f. and Bollinger, C. R. (2010), 'An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* **159**, 302–319.

Neumayer, E. and Plümper, T. (2016), 'W', *Political Science Research and Methods* **4**, 175–193.

Ord, K. (1975), 'Estimation methods for models of spatial interaction', *Journal of the American Statistical Association* **70**, 120–126.

Patacchini, E. and Zenou, Y. (2012), 'Juvenile delinquency and conformism', *The Journal of Law, Economics and Organization* **28**, 1–31.

Redding, S. and Venables, A. J. (2004), 'Economic geography and international inequality', **62**, 53–82.

Robinson, P. M. (2010), 'Efficient estimation of the semiparametric spatial autoregressive model', *Journal of Econometrics* **157**, 6–17.

Serfling, R. (2011), Asymptotic relative efficiency in estimation, *in* M. Lovric, ed., 'International Encyclopedia of Statistical Science', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 68–72.

Van Kerm, P. (2003), 'Adaptive kernel density estimation', *Stata Journal* **3**, 148–156.

Vermandele, C. (2000), Semiparametrically Efficient Sign-and-Rank Methods for Median Regression and AR Models, Phd thesis, Université Libre de Bruxelles.

White, H. (1984), *Asymptotic theory for econometricians*, Academic Press, Orlando.

Acknowledgments

We thank Serguei Dachian, Sophie Dabo, Arnaud Dufays, Abhimanyu Gupta, Xiaodong Liu, Davy Paindaveine, Florian Pelgrin and Francesca Rossi, as well as seminar participants at Ohio State University, Venice University, Université de Caen, Université de Lille, Université de Grenoble, participants of London and Essex workshops, the Journées de la Société Française de Statistique and those of the Stata workshop organized in Aix-Marseille University and in the Katholieke Universiteit Leuven. We also would like to thank Arthur Hemdat for having implemented the Matlab code in R. All remaining errors are our responsibility.

Appendix A. Appendix

Appendix A.1. Notations used in the paper

This section defines all notations used frequently in the paper.

For a square $n \times n$ matrix $\mathbf{A}^{(n)}$:

- $\mathrm{tr}\left(\mathbf{A}^{(n)}\right) = \sum_{i=1}^{n} A_{ii}^{(n)}$ denotes the trace of $\mathbf{A}^{(n)}$;
- $\mathbf{A}_{i\boldsymbol{\cdot}}^{(n)}$ represents the ith row of $\mathbf{A}^{(n)}$;
- $\overline{\mathbf{A}}_{\boldsymbol{\cdot}}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_{i\boldsymbol{\cdot}}^{(n)}$ is the average $(1 \times n)$-vector of the n rows of $\mathbf{A}^{(n)}$;
- $\overline{A}_{\boldsymbol{\cdot\cdot}}^{(n)} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}^{(n)}$ is the arithmetic mean of the n^2 components of $\mathbf{A}^{(n)}$.

Let $\mathbf{G}^{(n)}(\lambda) = \mathbf{W}^{(n)} \left(\mathbf{I}^{(n)} - \lambda \mathbf{W}^{(n)}\right)^{-1}$. We define

$$g^{(n)}(\lambda) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} G_{ij}^{(n)}(\lambda) = n^2 \overline{G}_{\boldsymbol{\cdot\cdot}}^{(n)}(\lambda) - \operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right).$$

We also define the following sums:

$$\overline{\mathbf{x}}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)},$$

$$\begin{aligned}
\mathbf{C}_{\mathbf{x}}^{(n)} =& \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} (\mathbf{x}_i^{(n)})^{\mathrm{T}}, \\
\mathbf{C}_{\mathbf{x}}^{(n)*} =& \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right)^{\mathrm{T}},
\end{aligned}$$

$$\begin{aligned}
\mathbf{C}_{\mathbf{x},\mathbf{Gx}}^{(n)}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} \left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta}\right), \\
\mathbf{C}_{\mathbf{x},\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) \left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) - \overline{\mathbf{G}}_{\boldsymbol{\cdot}}^{(n)}(\lambda)\right) \mathbf{X}^{(n)} \boldsymbol{\beta},
\end{aligned}$$

$$\begin{aligned}
\mathbf{C}_{\mathbf{x},\mathbf{G},1}^{(n)}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} G_{ii}^{(n)}(\lambda), \\
\mathbf{C}_{\mathbf{x},\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) \left(G_{ii}^{(n)}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right),
\end{aligned}$$

$$\begin{aligned}
\mathbf{C}_{\mathbf{x},\mathbf{G},2}^{(n)}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbf{x}_i^{(n)} G_{ij}^{(n)}(\lambda), \\
\mathbf{C}_{\mathbf{x},\mathbf{G},2}^{(n)*}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) \left(G_{ij}^{(n)}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right),
\end{aligned}$$

$$\begin{aligned}
C_{\mathbf{Gx}}^{(n)}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta}\right)^2, \\
C_{\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \left[\left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) - \overline{\mathbf{G}}_{\boldsymbol{\cdot}}^{(n)}(\lambda)\right) \mathbf{X}^{(n)} \boldsymbol{\beta}\right]^2,
\end{aligned}$$

$$C^{(n)}_{\mathbf{G},1}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(G^{(n)}_{ii}(\lambda)\right)^2,$$
$$C^{(n)*}_{\mathbf{G},1}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(G^{(n)}_{ii}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right)^2,$$

$$C^{(n)}_{\mathbf{G},2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\left(G^{(n)}_{ij}(\lambda)\right)^2,$$
$$C^{(n)*}_{\mathbf{G},2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right)^2,$$

$$C^{(n)}_{\mathbf{G},3}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}G^{(n)}_{ii}(\lambda)G^{(n)}_{ij}(\lambda),$$
$$C^{(n)*}_{\mathbf{G},3}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\left(G^{(n)}_{ii}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right)\left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right),$$

$$C^{(n)}_{\mathbf{G},4}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}G^{(n)}_{ij}(\lambda)G^{(n)}_{ji}(\lambda),$$
$$C^{(n)*}_{\mathbf{G},4}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right)\left(G^{(n)}_{ji}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right),$$

$$C^{(n)}_{\mathbf{G},5}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n}G^{(n)}_{ij}(\lambda)G^{(n)}_{ik}(\lambda),$$
$$C^{(n)*}_{\mathbf{G},5}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n}\left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right)\left(G^{(n)}_{ik}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right),$$

$$C^{(n)}_{\mathbf{Gx},\mathbf{G},1}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G^{(n)}_{ii}(\lambda),$$
$$C^{(n)*}_{\mathbf{Gx},\mathbf{G},1}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda) - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\right)\mathbf{X}^{(n)}\boldsymbol{\beta}\left(G^{(n)}_{ii}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right),$$

$$C^{(n)}_{\mathbf{Gx},\mathbf{G},2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G^{(n)}_{ij}(\lambda),$$

$$C^{(n)*}_{\mathbf{Gx},\mathbf{G},2}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda) - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\right)\mathbf{X}^{(n)}\boldsymbol{\beta}\left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right).$$

Let $f \in \mathcal{F}$ and F be the associated distribution function. Let $\phi_f(e) = -\frac{f'(e)}{f(e)}$ for $e \in \mathbb{R}$ and $\varphi_f(u) = \phi_f\left(F^{-1}(u)\right)$ for $u \in (0,1)$. We will use throughout the paper the following integrals:

$$\int_{-\infty}^{\infty}\phi_f(e)f(e)de = \int_0^1 \varphi_f(u)du = 0, \quad \int_0^{1/2}\varphi_f(u)du = -f(0), \quad \int_{1/2}^{1}\varphi_f(u)du = f(0),$$

$$\int_0^{1/2}\varphi_f(u)F^{-1}(u)du = \int_{1/2}^{1}\varphi_f(u)F^{-1}(u)du = \frac{1}{2},$$

$$\mu_f = \int_{-\infty}^{\infty} ef(e)de = \int_0^1 F^{-1}(u)du,$$

$$\nu_f = \int_{-\infty}^{\infty} e^2 f(e)de = \int_0^1 \left(F^{-1}(u)\right)^2 du,$$

$$\mathcal{I}_f = \int_{-\infty}^{\infty}\phi_f^2(e)f(e)de = \int_0^1 \varphi_f^2(u)du,$$

$$\mathcal{J}_f = \int_{-\infty}^{\infty}\phi_f(e)ef(e)de = \int_0^1 \varphi_f(u)F^{-1}(u)du = 1,$$

$$\mathcal{K}_f = \int_{-\infty}^{\infty}\phi_f^2(e)ef(e)de = \int_0^1 \varphi_f^2(u)F^{-1}(u)du,$$

$$\mathcal{Q}_f = \int_{-\infty}^{\infty}\phi_f^2(e)e^2f(e)de = \int_0^1 \left(\varphi_f(u)F^{-1}(u)\right)^2 du.$$

Moreover, for $h \in \mathcal{F}$ and its associated distribution function H:

$$\nu_{fh} = \int_0^1 F^{-1}(u)H^{-1}(u)du, \qquad \mathcal{I}_{fh} = \int_0^1 \varphi_f(u)\varphi_h(u)du,$$

$$\mathcal{J}_{fh} = \int_0^1 \varphi_f(u)H^{-1}(u)du, \qquad \mathcal{J}_{hf} = \int_0^1 \varphi_h(u)F^{-1}(u)du,$$

$$\mathcal{K}_{fh} = \int_0^1 \varphi_f(u)\varphi_h(u)H^{-1}(u)du, \qquad \mathcal{K}_{hf} = \int_0^1 \varphi_h(u)\varphi_f(u)F^{-1}(u)du,$$

$$\mathcal{Q}_{fh} = \int_0^1 \left(\varphi_f(u)F^{-1}(u)\right)\left(\varphi_h(u)H^{-1}(u)\right)du.$$

Note that $\nu_{ff} = \nu_f$, $\mathcal{I}_{ff} = \mathcal{I}_f$, $\mathcal{J}_{ff} = \mathcal{J}_f = 1$, $\mathcal{K}_{ff} = \mathcal{K}_f$, and $\mathcal{Q}_{ff} = \mathcal{Q}_f$.

Appendix A.2. Elements of proof of Proposition 1 and parametric Fisher information matrix for $\boldsymbol{\theta}$ under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$

The sequence of parametric submodels $\mathcal{E}_f^{(n)}$, $n \in \mathbb{N}$, satisfies the ULAN property if there exists a sequence $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ of $(K+2)$-dimensional and $\left(\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}, \boldsymbol{\theta}\right)$-measurable random vectors, and a $(K+2) \times (K+2)$ symmetric positive semi-definite matrix $\mathbf{I}_f(\boldsymbol{\theta})$ continuous in $\boldsymbol{\theta}$ such that, for any $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^{\mathrm{T}}, \lambda_0)^{\mathrm{T}} \in \Theta$, any sequence $\left(\boldsymbol{\theta}^{(n)}\right)$ in Θ defined as $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}_0 + O(1/\sqrt{n})$ (where the sequence $\left(\boldsymbol{\theta}^{(n)}\right)$ converges to $\boldsymbol{\theta}_0$ for $n \to \infty$), and any bounded sequence $\left(\boldsymbol{\tau}^{(n)}\right) \in \mathbb{R}^{K+2}$, we have, under $P_{f;\boldsymbol{\theta}^{(n)}}^{(n)}$, as $n \to \infty$:

(i)

$$\ln\left(\frac{L_f\left(\boldsymbol{\theta}^{(n)} + \boldsymbol{\tau}^{(n)}/\sqrt{n}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{L_f\left(\boldsymbol{\theta}^{(n)}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}\right) = (\boldsymbol{\tau}^{(n)})^{\mathrm{T}}\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}^{(n)}) - \frac{1}{2}(\boldsymbol{\tau}^{(n)})^{\mathrm{T}}\mathbf{I}_f(\boldsymbol{\theta}_0)\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1)\,;$$

(ii)

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}^{(n)}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_f(\boldsymbol{\theta}_0)\right).$$

Under Assumption (D3), the decomposition (i) of the logarithm of the likelihood ratio is based on a second order Taylor expansion of $\ln L_f\left(\boldsymbol{\theta}^{(n)} + \boldsymbol{\tau}^{(n)}/\sqrt{n}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)$ around $\boldsymbol{\theta}^{(n)}$, with

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}^{(n)}) = \frac{1}{\sqrt{n}}\left(\left.\frac{\partial \ln L_f\left(\boldsymbol{\theta}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(n)}}\right)$$

and $\mathbf{I}_f(\boldsymbol{\theta}_0)$ being the limit in probability of

$$-\frac{1}{n}\left(\left.\frac{\partial^2 \ln L_f\left(\boldsymbol{\theta}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(n)}}\right).$$

As shown in Proposition 1, the (so-called) central sequence for $\boldsymbol{\theta}$ in the parametric submodel $\mathcal{E}_f^{(n)}$ takes the following form:

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\begin{pmatrix} \frac{\partial \ln L_f\left(\boldsymbol{\theta}\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ln L_f\left(\boldsymbol{\theta}\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) \\ \Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta}) \end{pmatrix},$$

with

$$\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)\mathbf{x}_i^{(n)}$$

and

$$\Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)\mathbf{W}_{i\boldsymbol{.}}^{(n)}\mathbf{y}^{(n)} - \frac{1}{\sqrt{n}}\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right).$$

Since, in view of (2),

$$\begin{aligned}\mathbf{W}_{i\bullet}^{(n)}\mathbf{y}^{(n)} &= \mathbf{G}_{i\bullet}^{(n)}(\lambda)\left(\mathbf{X}^{(n)}\boldsymbol{\beta}+\mathbf{e}^{(n)}(\boldsymbol{\theta})\right)\\ &= \mathbf{G}_{i\bullet}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}+\sum_{j=1}^{n}G_{ij}^{(n)}(\lambda)e_j^{(n)}(\boldsymbol{\theta}),\end{aligned}$$

we get

$$\Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta}) = L_{1;f}^{(n)}(\boldsymbol{\theta})+L_{2;f}^{(n)}(\boldsymbol{\theta})+L_{3;f}^{(n)}(\boldsymbol{\theta})+L_{4;f}^{(n)}(\boldsymbol{\theta}),$$

with

$$\begin{aligned}L_{1;f}^{(n)}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)\mathbf{G}_{i\bullet}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta},\\ L_{2;f}^{(n)}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)e_i^{(n)}(\boldsymbol{\theta})G_{ii}^{(n)}(\lambda),\\ L_{3;f}^{(n)}(\boldsymbol{\theta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)e_j^{(n)}(\boldsymbol{\theta})G_{ij}^{(n)}(\lambda),\\ L_{4;f}^{(n)}(\boldsymbol{\theta}) &= -\frac{1}{\sqrt{n}}\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right).\end{aligned}$$

Under $P_{f;\boldsymbol{\theta}}^{(n)}$, the error terms $e_1^{(n)}(\boldsymbol{\theta}),\ldots,e_n^{(n)}(\boldsymbol{\theta})$ are i.i.d. with density function f and we have, for all $i=1,\ldots,n$:

- $\mathbb{E}\left[e_i^{(n)}(\boldsymbol{\theta})\right]=\int_{-\infty}^{\infty}ef(e)de \overset{\text{def}}{=} \mu_f$;
- $\mathbb{E}\left[\left(e_i^{(n)}(\boldsymbol{\theta})\right)^2\right]=\int_{-\infty}^{\infty}e^2f(e)de \overset{\text{def}}{=} \nu_f$;
- $\mathbb{E}\left[\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)\right]=\int_{-\infty}^{\infty}\phi_f(e)f(e)de=-\int_{-\infty}^{\infty}f'(e)de=-\left[f(e)\right]_{-\infty}^{\infty}=0$;
- $\mathbb{E}\left[\phi_f^2\left(e_i^{(n)}(\boldsymbol{\theta})\right)\right]=\int_{-\infty}^{\infty}\phi_f^2(e)f(e)de \overset{\text{def}}{=} \mathcal{I}_f$;
- $\mathbb{E}\left[\phi_f\left(e_i^{(n)}(\boldsymbol{\theta})\right)e_i^{(n)}(\boldsymbol{\theta})\right] = \int_{-\infty}^{\infty}\phi_f(e)ef(e)de \overset{\text{def}}{=} \mathcal{J}_f = -\int_{-\infty}^{\infty}f'(e)ede = -\left[f(e)e\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}f(e)de=0+1=1$;
- $\mathbb{E}\left[\phi_f^2\left(e_i^{(n)}(\boldsymbol{\theta})\right)e_i^{(n)}(\boldsymbol{\theta})\right]=\int_{-\infty}^{\infty}\phi_f^2(e)ef(e)de \overset{\text{def}}{=} \mathcal{K}_f$;
- $\mathbb{E}\left[\phi_f^2\left(e_i^{(n)}(\boldsymbol{\theta})\right)\left(e_i^{(n)}(\boldsymbol{\theta})\right)^2\right]=\int_{-\infty}^{\infty}\phi_f^2(e)e^2f(e)de \overset{\text{def}}{=} \mathcal{Q}_f$.

It follows that, under $P_{f;\boldsymbol{\theta}}^{(n)}$,

$$\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\right]=\mathbf{0}$$

and

$$
\begin{aligned}
\mathbb{E}\left[\Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta})\right] &= \mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[L_{2;f}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[L_{3;f}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[L_{4;f}^{(n)}(\boldsymbol{\theta})\right] \\
&= 0 + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_{ii}^{(n)}(\lambda) + 0 - \frac{1}{\sqrt{n}} \mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) \\
&= \frac{1}{\sqrt{n}} \mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) - \frac{1}{\sqrt{n}} \mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) \\
&= 0.
\end{aligned}
$$

Moreover, under $P_{f;\boldsymbol{\theta}}^{(n)}$,

$$
\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) \left(\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right] = \mathcal{I}_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} (\mathbf{x}_i^{(n)})^{\mathrm{T}} \right\}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) \Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta})\right] &= \mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{1;f}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{2;f}^{(n)}(\boldsymbol{\theta})\right] \\
&\quad + \mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{3;f}^{(n)}(\boldsymbol{\theta})\right] + \mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{4;f}^{(n)}(\boldsymbol{\theta})\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{1;f}^{(n)}(\boldsymbol{\theta})\right] &= \mathcal{I}_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} \left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta}\right) \right\}, \\
\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{2;f}^{(n)}(\boldsymbol{\theta})\right] &= \mathcal{K}_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} G_{ii}^{(n)}(\lambda) \right\}, \\
\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{3;f}^{(n)}(\boldsymbol{\theta})\right] &= \mathcal{I}_f \mu_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbf{x}_i^{(n)} G_{ij}^{(n)}(\lambda) \right\}, \\
\mathbb{E}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}(\boldsymbol{\theta}) L_{4;f}^{(n)}(\boldsymbol{\theta})\right] &= 0.
\end{aligned}
$$

Finally, under $P_{f;\boldsymbol{\theta}}^{(n)}$,

$$
\begin{aligned}
\mathbb{E}\left[\left(\Delta_{f;\lambda}^{(n)}(\boldsymbol{\theta})\right)^2\right] &= \mathbb{E}\left[\left(L_{1;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] + \mathbb{E}\left[\left(L_{2;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] + \mathbb{E}\left[\left(L_{3;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] + \mathbb{E}\left[\left(L_{4;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] \\
&\quad + 2\,\mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta}) L_{2;f}^{(n)}(\boldsymbol{\theta})\right] + 2\,\mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta}) L_{3;f}^{(n)}(\boldsymbol{\theta})\right] + 2\,\mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta}) L_{4;f}^{(n)}(\boldsymbol{\theta})\right] \\
&\quad + 2\,\mathbb{E}\left[L_{2;f}^{(n)}(\boldsymbol{\theta}) L_{3;f}^{(n)}(\boldsymbol{\theta})\right] + 2\,\mathbb{E}\left[L_{2;f}^{(n)}(\boldsymbol{\theta}) L_{4;f}^{(n)}(\boldsymbol{\theta})\right] + 2\,\mathbb{E}\left[L_{3;f}^{(n)}(\boldsymbol{\theta}) L_{4;f}^{(n)}(\boldsymbol{\theta})\right],
\end{aligned}
$$

where

$$
\mathbb{E}\left[\left(L_{1;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] = \mathcal{I}_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta}\right)^2 \right\},
$$

$$\mathbb{E}\left[\left(L_{2;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] = (\mathcal{Q}_f - 1)\left\{\frac{1}{n}\sum_{i=1}^{n}\left(G_{ii}^{(n)}(\lambda)\right)^2\right\} + \frac{\left(\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)\right)^2}{n},$$

$$\mathbb{E}\left[\left(L_{3;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] = \mathcal{I}_f\nu_f\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\left(G_{ij}^{(n)}(\lambda)\right)^2\right\} + \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}G_{ij}^{(n)}(\lambda)G_{ji}^{(n)}(\lambda)$$
$$+\,\mathcal{I}_f\mu_f^2\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{k=1\\k\neq i,j}}^{n}G_{ij}^{(n)}(\lambda)G_{ik}^{(n)}(\lambda)\right\},$$

$$\mathbb{E}\left[\left(L_{4;f}^{(n)}(\boldsymbol{\theta})\right)^2\right] = \frac{\left(\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)\right)^2}{n},$$

$$\mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta})L_{2;f}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{K}_f\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G_{ii}^{(n)}(\lambda)\right\},$$

$$\mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta})L_{3;f}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{I}_f\mu_f\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta}\right)G_{ij}^{(n)}(\lambda)\right\},$$

$$\mathbb{E}\left[L_{2;f}^{(n)}(\boldsymbol{\theta})L_{3;f}^{(n)}(\boldsymbol{\theta})\right] = \mathcal{K}_f\mu_f\lim_{n\to\infty}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}G_{ii}^{(n)}(\lambda)G_{ij}^{(n)}(\lambda)\right\},$$

$$\mathbb{E}\left[L_{1;f}^{(n)}(\boldsymbol{\theta})L_{4;f}^{(n)}(\boldsymbol{\theta})\right] = \mathbb{E}\left[L_{3;f}^{(n)}(\boldsymbol{\theta})L_{4;f}^{(n)}(\boldsymbol{\theta})\right] = 0,$$

and

$$\mathbb{E}\left[L_{2;f}^{(n)}(\boldsymbol{\theta})L_{4;f}^{(n)}(\boldsymbol{\theta})\right] = -\frac{\left(\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)\right)^2}{n}.$$

The expression for the (parametric) Fisher information matrix $\mathbf{I}_f(\boldsymbol{\theta})$ follows directly from the above results.

Appendix A.3. Rank-and-sign statistics: preliminaries

Let $\mathcal{H}_h^{(n)}$ denote the hypothesis under which $e_1^{(n)}, \ldots, e_n^{(n)}$ are i.i.d. with density h defined (with respect to Lebesgue's measure) on $\mathbb{R}$ and cumulative distribution function H such that $H(0) = \int_{-\infty}^{0} h(e)de = 1/2$. Define the uniform (under $\mathcal{H}_h^{(n)}$) random variables $U_i^{(n)} = H(e_i^{(n)})$. The rank of $U_i^{(n)}$ and the sign of $U_i^{(n)} - 1/2$ clearly coincide with those of $e_i^{(n)}$. Finally, we write U for a generic random variable with uniform distribution over $(0, 1)$.

We define $\mathbf{N}^{(n)} = \left(N_-^{(n)}, N_+^{(n)}\right)$ with $N_-^{(n)} = \sum_{i=1}^{n} 1_{\left[U_i^{(n)}<1/2\right]}$, $N_+^{(n)} = \sum_{i=1}^{n} 1_{\left[U_i^{(n)}>1/2\right]}$, and $N_-^{(n)} + N_+^{(n)} = n$ almost surely.

Appendix A.3.1. Linear rank-and-sign statistics of order 1

a) Definition. A linear rank-and-sign statistic of order 1 — also called linear nonserial rank-and-sign statistic — is a statistic of the form

$$\mathbf{S}^{(n)}_{1;\text{exact/appr}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}^{(n)}_i a^{(n)}_{\text{exact/appr}}\left(\mathbf{N}^{(n)}; R^{(n)}_i\right),$$

where $\mathbf{c}^{(n)}_i$ $(i = 1, \ldots, n)$ are constant vectors in $\mathbb{R}^p$ and, for $n_-, n_+ \in \{0, 1, \ldots, n\}$ with $n_+ = n - n_-$, and $r \in \{1, \ldots, n\}$,

$$a^{(n)}_{\text{exact}}\left((n_-, n_+); r\right) = \mathbb{E}\left[J\left(U^{(n)}_i\right)\Big|\mathbf{N}^{(n)} = (n_-, n_+); R^{(n)}_i = r\right]$$

(*exact score*) and

$$a^{(n)}_{\text{appr}}\left((n_-, n_+); r\right) = \begin{cases} J\left(\frac{1}{2}\,\frac{r}{n_-+1}\right) & \text{if } r \le n_- \\ J\left(\frac{1}{2} + \frac{1}{2}\,\frac{r-n_-}{n_++1}\right) & \text{if } r > n_- \end{cases}$$

(*approximate score*) for some nonconstant and square-integrable *score-generating* function $J : (0, 1) \mapsto \mathbb{R}$ satisfying $\mu_J = \mathbb{E}\left[J(U)\right] < \infty$ and $0 < \sigma^2_J = \mathbb{E}\left[J^2(U)\right] < \infty$. In the approximate form case, J moreover is assumed to be the difference of two non-decreasing square-integrable functions. Note that, under $\mathcal{H}^{(n)}_h$,

$$\mathbf{S}^{(n)}_{1;\text{exact}} = \mathbb{E}\left[\mathbf{T}^{(n)}_1 \Big|\mathbf{N}^{(n)}; \mathbf{R}^{(n)}\right],$$

with

$$\mathbf{T}^{(n)}_1 = \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}^{(n)}_i J\left(H(e^{(n)}_i)\right).$$

Moreover,

$$a^{(n)}_{\text{appr}}\left(\mathbf{N}^{(n)}; R^{(n)}_i\right) = J\left(\widetilde{R}^{(n)}_i\right),$$

with

$$\widetilde{R}^{(n)}_i = 1_{\left[s^{(n)}_i=-1\right]}\left\{\frac{1}{2}\,\frac{R^{(n)}_i}{N^{(n)}_- + 1}\right\} + 1_{\left[s^{(n)}_i=+1\right]}\left\{\frac{1}{2} + \frac{1}{2}\,\frac{R^{(n)}_i - (n - N^{(n)}_+)}{N^{(n)}_+ + 1}\right\},$$

where $s^{(n)}_i$ and $R^{(n)}_i$ are the sign and the rank of $e^{(n)}_i$ among $e^{(n)}_1, \ldots, e^{(n)}_n$.

b) Assumptions on the constants $\mathbf{c}^{(n)}_i \in \mathbb{R}^p$ *$(i = 1, \ldots, n)$* . The elements of $\mathbf{c}^{(n)}_i$ are uniformly bounded constants for all n. Besides, the $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbf{c}^{(n)}_i (\mathbf{c}^{(n)}_i)^{\mathrm{T}}$ exists and is non-singular. Finally, for each $k = 1, \ldots, p$, the set of constants $\left\{c^{(n)}_{ik}; i = 1, \ldots, n\right\}$ satisfies the classical *Noether condition*: with the convention $\frac{0}{0} = 0$ and $\overline{c}^{(n)}_k = \frac{1}{n}\sum_{i=1}^{n} c^{(n)}_{ik}$,

$$\lim_{n\to\infty}\left[\frac{\max_{1\le i\le n}\left(c^{(n)}_{ik} - \overline{c}^{(n)}_k\right)^2}{\sum_{i=1}^{n}\left(c^{(n)}_{ik} - \overline{c}^{(n)}_k\right)^2}\right] = 0.$$

c) Results about asymptotic representation. By Proposition 3.2 of Hallin et al. (2006), we have that, under $\mathcal{H}_h^{(n)}$, as $n \to \infty$,

$$\mathbf{S}_{1;\text{exact}}^{(n)} - \mathbb{E}\left[\mathbf{S}_{1;\text{exact}}^{(n)}\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{c}_i^{(n)} - \overline{\mathbf{c}}^{(n)}\right) J\left(H(e_i^{(n)})\right) + \overline{\mathbf{c}}^{(n)}\left\{2\frac{N_-^{(n)}}{n}\mu_J^- + 2\frac{N_+^{(n)}}{n}\mu_J^+ - \mu_J\right\} + o_{\mathrm{P}}(1/\sqrt{n}),$$

where $\mu_J^- = \int_0^{1/2} J(u)du$, $\mu_J^+ = \int_{1/2}^1 J(u)du$, and $\mu_J = \int_0^1 J(u)du$. Since, under $\mathcal{H}_h^{(n)}$, $\mathbf{S}_{1;\text{exact}}^{(n)} = \mathbb{E}\left[\mathbf{T}_1^{(n)} \middle| \mathbf{N}^{(n)}; \mathbf{R}^{(n)}\right]$ and

$$\begin{aligned}\mathbb{E}\left[\mathbf{S}_{1;\text{exact}}^{(n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{T}_1^{(n)} \middle| \mathbf{N}^{(n)}; \mathbf{R}^{(n)}\right]\right] = \mathbb{E}\left[\mathbf{T}_1^{(n)}\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{c}_i^{(n)}\mathbb{E}\left[J\left(H(e_i^{(n)})\right)\right] = \overline{\mathbf{c}}^{(n)}\mu_J,\end{aligned}$$

it follows that, under $\mathcal{H}_h^{(n)}$, as $n \to \infty$,

$$\mathbb{E}\left[\mathbf{T}_1^{(n)} \middle| \mathbf{N}^{(n)}; \mathbf{R}^{(n)}\right] = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{c}_i^{(n)} - \overline{\mathbf{c}}^{(n)}\right) J\left(H(e_i^{(n)})\right) + \overline{\mathbf{c}}^{(n)}\left\{2\frac{N_-^{(n)}}{n}\mu_J^- + 2\frac{N_+^{(n)}}{n}\mu_J^+\right\} + o_{\mathrm{P}}(1/\sqrt{n}). \tag{A.1}$$

Moreover, Lemma 3.1 – more precisely, relation (3.7) – of Hallin et al. (2006) implies that, under $\mathcal{H}_h^{(n)}$, as $n \to \infty$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{c}_i^{(n)} - \overline{\mathbf{c}}^{(n)}\right) J\left(H(e_i^{(n)})\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{c}_i^{(n)} - \overline{\mathbf{c}}^{(n)}\right) a_{\text{appr}}^{(n)}\left(\mathbf{N}^{(n)}; R_i^{(n)}\right) + o_{\mathrm{P}}(1/\sqrt{n}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{c}_i^{(n)} - \overline{\mathbf{c}}^{(n)}\right) J\left(\widetilde{R}_i^{(n)}\right) + o_{\mathrm{P}}(1/\sqrt{n}).\end{aligned} \tag{A.2}$$

Appendix A.3.2. Linear rank-and-sign statistics of order 2
a) Definition. We will consider linear rank-and-sign statistics of the form

$$S_{2;\text{exact/appr}}^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}^{(n)} a_{\text{exact/appr}}^{(n)}\left(\mathbf{N}^{(n)}; R_i^{(n)}, R_j^{(n)}\right),$$

where $c_{ij}^{(n)} \in \mathbb{R}$ $(i, j = 1, \ldots, n)$ and, for $n_-, n_+ \in \{0, 1, \ldots, n\}$ with $n_+ = n - n_-$, and $r, s \in \{1, \ldots, n\}$, $r \neq s$,

$$a_{\text{exact}}^{(n)}\left((n_-, n_+); r, s\right) = \mathbb{E}\left[J\left(U_i^{(n)}, U_j^{(n)}\right) \middle| \mathbf{N}^{(n)} = (n_-, n_+); R_i^{(n)} = r, R_j^{(n)} = s\right]$$

(*exact score*) and

$$a^{(n)}_{\text{appr}}\left((n_-,n_+);r,s\right)=\begin{cases} J\left(\frac{1}{2}\,\frac{r}{n_-+1},\frac{1}{2}\,\frac{s}{n_-+1}\right) & \text{if } r\leq n_- \text{ and } s\leq n_-\\ J\left(\frac{1}{2}+\frac{1}{2}\,\frac{r-n_-}{n_++1},\frac{1}{2}\,\frac{s}{n_-+1}\right) & \text{if } r> n_- \text{ and } s\leq n_-\\ J\left(\frac{1}{2}\,\frac{r}{n_-+1},\frac{1}{2}+\frac{1}{2}\,\frac{s-n_-}{n_++1}\right) & \text{if } r\leq n_- \text{ and } s> n_-\\ J\left(\frac{1}{2}+\frac{1}{2}\,\frac{r-n_-}{n_++1},\frac{1}{2}+\frac{1}{2}\,\frac{s-n_-}{n_++1}\right) & \text{if } r> n_- \text{ and } s> n_-\end{cases}$$

(*approximate score*) for some *score-generating* function $J:(0,1)^2\mapsto\mathbb{R}$ satisfying $\mu_J=\mathbb{E}\left[J(U_1,U_2)\right]<\infty$ and $0<\sigma^2_J=\mathbb{E}\left[J^2(U_1,U_2)\right]<\infty$. In the approximate form case, J moreover is assumed to be the difference of two non-decreasing functions. Note that, under $\mathcal{H}^{(n)}_h$,

$$S^{(n)}_{2;\text{exact}}=\mathbb{E}\left[T^{(n)}_2\,\middle|\,\mathbf{N}^{(n)};\mathbf{R}^{(n)}\right],$$

with

$$T^{(n)}_2=\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}c^{(n)}_{ij}J\left(H(e^{(n)}_i),H(e^{(n)}_j)\right).$$

Moreover,

$$a^{(n)}_{\text{appr}}\left(\mathbf{N}^{(n)};R^{(n)}_i,R^{(n)}_j\right)=J\left(\widetilde{R}^{(n)}_i,\widetilde{R}^{(n)}_j\right).$$

b) Assumptions on the constants $c^{(n)}_{ij}\in\mathbb{R}$ *(*$i\neq j\in\{1,\ldots,n\}$*).* Let $\overline{c}^{(n)}=\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}c^{(n)}_{ij}$. The constants $c^{(n)}_{ij}$ have to satisfy the classical *Noether condition*: with the convention $\frac{0}{0}=0$,

$$\lim_{n\to\infty}\left[\frac{\max_{1\leq i\neq j\leq n}\left(c^{(n)}_{ij}-\overline{c}^{(n)}\right)^2}{\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(c^{(n)}_{ij}-\overline{c}^{(n)}\right)^2}\right]=0.$$

This also implies that

$$\lim_{n\to\infty}\left[\frac{\max_{1\leq i\neq j\leq n}\left(c^{(n)}_{ij}\right)^2}{\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(c^{(n)}_{ij}\right)^2}\right]=0$$

(uniform asymptotic negligibility of each $c^{(n)}_{ij}$).

c) Results about asymptotic representation. The following result extends to the present context the asymptotic representation result part of Lemma 4.1 in Hallin et al. (2006); the proof follows along the same steps and is omitted. Under $\mathcal{H}^{(n)}_h$, as $n\to\infty$,

$$S^{(n)}_{2;\text{exact/appr}}-\mathbb{E}\left[S^{(n)}_{2;\text{exact/appr}}\,\middle|\,\mathbf{N}^{(n)}\right]=T^{(n)}_2-\mathbb{E}\left[T^{(n)}_2\,\middle|\,\mathbf{e}^{(n)}_{(\cdot)}\right]+o_{\mathrm{P}}(1/\sqrt{n}),\tag{A.3}$$

where $\mathbf{e}^{(n)}_{(\cdot)}$ denotes the vector of order statistics associated with $e^{(n)}_1,\ldots,e^{(n)}_n$.

(i) Note first that, under $\mathcal{H}_h^{(n)}$,

$$\begin{aligned}
\mathbb{E}\left[T_2^{(n)}\,\middle|\,\mathbf{e}_{(.)}^{(n)}\right] &= \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}^{(n)}\,\mathbb{E}\left[J\left(U_i^{(n)},U_j^{(n)}\right)\,\middle|\,\mathbf{U}_{(.)}^{(n)}\right] \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}^{(n)}\right)\mathbb{E}\left[J\left(U_1^{(n)},U_2^{(n)}\right)\,\middle|\,\mathbf{U}_{(.)}^{(n)}\right] \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}^{(n)}\right)\frac{1}{n(n-1)}\sum_{k=1}^{n}\sum_{\substack{\ell=1\\ \ell\neq k}}^{n} J\left(U_{(k)}^{(n)},U_{(\ell)}^{(n)}\right) \\
&= \left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}^{(n)}\right)\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} J\left(U_i^{(n)},U_j^{(n)}\right) \\
&= \overline{c}^{(n)}\,\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} J\left(H(e_i^{(n)}),H(e_j^{(n)})\right).
\end{aligned}$$

Consequently, under $\mathcal{H}_h^{(n)}$,

$$T_2^{(n)}-\mathbb{E}\left[T_2^{(n)}\,\middle|\,\mathbf{e}_{(.)}^{(n)}\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(c_{ij}^{(n)}-\overline{c}^{(n)}\right)J\left(H(e_i^{(n)}),H(e_j^{(n)})\right). \tag{A.4}$$

(ii) Moreover, under $\mathcal{H}_h^{(n)}$,

$$\begin{aligned}
\mathbb{E}\left[S_{2;\text{exact}}^{(n)}\,\middle|\,\mathbf{N}^{(n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[T_2^{(n)}\,\middle|\,\mathbf{N}^{(n)};\mathbf{R}^{(n)}\right]\,\middle|\,\mathbf{N}^{(n)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[T_2^{(n)}\,\middle|\,\mathbf{s}^{(n)}\right]\,\middle|\,\mathbf{N}^{(n)}\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}^{(n)}\,\mathbb{E}\left[J(U_i^{(n)},U_j^{(n)})\,\middle|\,s_i^{(n)},s_j^{(n)}\right]\,\middle|\,\mathbf{N}^{(n)}\right],
\end{aligned}$$

with, for $i \neq j$,

$$\mathbb{E}\left[J(U_i^{(n)}, U_j^{(n)}) \middle| s_i^{(n)}, s_j^{(n)}\right] = 1_{\left[s_i^{(n)}=-1, s_j^{(n)}=-1\right]} \int_0^{1/2} \left(\int_0^{1/2} J(u_1, u_2) 2du_2 \right) 2du_1$$
$$+ 1_{\left[s_i^{(n)}=-1, s_j^{(n)}=+1\right]} \int_0^{1/2} \left(\int_{1/2}^{1} J(u_1, u_2) 2du_2 \right) 2du_1$$
$$+ 1_{\left[s_i^{(n)}=+1, s_j^{(n)}=-1\right]} \int_{1/2}^{1} \left(\int_0^{1/2} J(u_1, u_2) 2du_2 \right) 2du_1$$
$$+ 1_{\left[s_i^{(n)}=+1, s_j^{(n)}=+1\right]} \int_{1/2}^{1} \left(\int_{1/2}^{1} J(u_1, u_2) 2du_2 \right) 2du_1.$$

Defining

$$\mu_J^{--} = \int_0^{1/2} \left(\int_0^{1/2} J(u_1, u_2) du_2 \right) du_1,$$
$$\mu_J^{-+} = \int_0^{1/2} \left(\int_{1/2}^{1} J(u_1, u_2) du_2 \right) du_1,$$
$$\mu_J^{+-} = \int_{1/2}^{1} \left(\int_0^{1/2} J(u_1, u_2) du_2 \right) du_1,$$
$$\mu_J^{++} = \int_{1/2}^{1} \left(\int_{1/2}^{1} J(u_1, u_2) du_2 \right) du_1,$$

we have, for $i \neq j$,

$$\mathbb{E}\left[J(U_i^{(n)}, U_j^{(n)}) \middle| s_i^{(n)}, s_j^{(n)}\right] = 4 \times \left\{ 1_{\left[s_i^{(n)}=-1, s_j^{(n)}=-1\right]} \mu_J^{--} + 1_{\left[s_i^{(n)}=-1, s_j^{(n)}=+1\right]} \mu_J^{-+} + 1_{\left[s_i^{(n)}=+1, s_j^{(n)}=-1\right]} \mu_J^{+-} + 1_{\left[s_i^{(n)}=+1, s_j^{(n)}=+1\right]} \mu_J^{++} \right\},$$

such that

$$\mathbb{E}\left[\mathbb{E}\left[J(U_i^{(n)}, U_j^{(n)}) \middle| s_i^{(n)}, s_j^{(n)}\right] \middle| \mathbf{N}^{(n)}\right] = 4 \times \left\{ \mathbb{P}\left[s_i^{(n)} = -1, s_j^{(n)} = -1 \middle| \mathbf{N}^{(n)}\right] \mu_J^{--} + \mathbb{P}\left[s_i^{(n)} = -1, s_j^{(n)} = +1 \middle| \mathbf{N}^{(n)}\right] \mu_J^{-+} + \mathbb{P}\left[s_i^{(n)} = +1, s_j^{(n)} = -1 \middle| \mathbf{N}^{(n)}\right] \mu_J^{+-} + \mathbb{P}\left[s_i^{(n)} = +1, s_j^{(n)} = +1 \middle| \mathbf{N}^{(n)}\right] \mu_J^{++} \right\}$$
$$= 4 \times \left\{ \frac{N_-^{(n)}\left(N_-^{(n)} - 1\right)}{n(n-1)} \mu_J^{--} + \frac{N_-^{(n)} N_+^{(n)}}{n(n-1)} \left(\mu_J^{-+} + \mu_J^{+-}\right) + \frac{N_+^{(n)}\left(N_+^{(n)} - 1\right)}{n(n-1)} \mu_J^{++} \right\}.$$

Consequently, under $\mathcal{H}_h^{(n)}$,

$$\begin{aligned}
&\mathbb{E}\left[S_{2;\text{exact}}^{(n)} \,\middle|\, \mathbf{N}^{(n)}\right] \\
&= (n-1)\overline{c}^{(n)} 4 \left\{ \frac{N_-^{(n)}\left(N_-^{(n)}-1\right)}{n(n-1)} \mu_J^{--} + \frac{N_-^{(n)} N_+^{(n)}}{n(n-1)} \left(\mu_J^{-+} + \mu_J^{+-}\right) + \frac{N_+^{(n)}\left(N_+^{(n)}-1\right)}{n(n-1)} \mu_J^{++} \right\} \\
&= \frac{4\overline{c}^{(n)}}{n} \left\{ N_-^{(n)}\left(N_-^{(n)}-1\right)\mu_J^{--} + N_-^{(n)} N_+^{(n)}\left(\mu_J^{-+} + \mu_J^{+-}\right) + N_+^{(n)}\left(N_+^{(n)}-1\right)\mu_J^{++} \right\}. \qquad \text{(A.5)}
\end{aligned}$$

(iii) Finally, under $\mathcal{H}_h^{(n)}$,

$$\begin{aligned}
\mathbb{E}\left[S_{2;\text{appr}}^{(n)} \,\middle|\, \mathbf{N}^{(n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[S_{2;\text{appr}}^{(n)} \,\middle|\, \mathbf{s}^{(n)}\right] \,\middle|\, \mathbf{N}^{(n)}\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} c_{ij}^{(n)} \mathbb{E}\left[\mathbb{E}\left[J\left(\widetilde{R}_i^{(n)}, \widetilde{R}_j^{(n)}\right) \,\middle|\, \mathbf{s}^{(n)}\right] \,\middle|\, \mathbf{N}^{(n)}\right].
\end{aligned}$$

For $i \neq j$,

$$\begin{aligned}
&\mathbb{E}\left[J\left(\widetilde{R}_i^{(n)}, \widetilde{R}_j^{(n)}\right) \,\middle|\, \mathbf{s}^{(n)}\right] \\
&= \mathbb{E}\left[J\left(\widetilde{R}_i^{(n)}, \widetilde{R}_j^{(n)}\right) \,\middle|\, \mathbf{N}^{(n)}, s_i^{(n)}, s_j^{(n)}\right] \\
&= 1_{\left[s_i^{(n)}=-1, s_j^{(n)}=-1\right]} \frac{1}{N_-^{(n)}(N_-^{(n)}-1)} \sum_{k=1}^{N_-^{(n)}} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_-^{(n)}} J\left(\frac{1}{2}\frac{k}{N_-^{(n)}+1}, \frac{1}{2}\frac{\ell}{N_-^{(n)}+1}\right) \\
&\quad + 1_{\left[s_i^{(n)}=-1, s_j^{(n)}=+1\right]} \frac{1}{N_-^{(n)} N_+^{(n)}} \sum_{k=1}^{N_-^{(n)}} \sum_{\ell=1}^{N_+^{(n)}} J\left(\frac{1}{2}\frac{k}{N_-^{(n)}+1}, \frac{1}{2} + \frac{1}{2}\frac{\ell}{N_+^{(n)}+1}\right) \\
&\quad + 1_{\left[s_i^{(n)}=+1, s_j^{(n)}=-1\right]} \frac{1}{N_-^{(n)} N_+^{(n)}} \sum_{k=1}^{N_+^{(n)}} \sum_{\ell=1}^{N_-^{(n)}} J\left(\frac{1}{2} + \frac{1}{2}\frac{k}{N_+^{(n)}+1}, \frac{1}{2}\frac{\ell}{N_-^{(n)}+1}\right) \\
&\quad + 1_{\left[s_i^{(n)}=+1, s_j^{(n)}=+1\right]} \frac{1}{N_+^{(n)}(N_+^{(n)}-1)} \sum_{k=1}^{N_+^{(n)}} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_+^{(n)}} J\left(\frac{1}{2} + \frac{1}{2}\frac{k}{N_+^{(n)}+1}, \frac{1}{2} + \frac{1}{2}\frac{\ell}{N_+^{(n)}+1}\right).
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbb{P}\left[s_i^{(n)} = -1, s_j^{(n)} = -1 \,\middle|\, \mathbf{N}^{(n)}\right] = \frac{N_-^{(n)}(N_-^{(n)}-1)}{n(n-1)}, \\
&\mathbb{P}\left[s_i^{(n)} = -1, s_j^{(n)} = +1 \,\middle|\, \mathbf{N}^{(n)}\right] = \mathbb{P}\left[s_i^{(n)} = +1, s_j^{(n)} = -1 \,\middle|\, \mathbf{N}^{(n)}\right] = \frac{N_-^{(n)} N_+^{(n)}}{n(n-1)},
\end{aligned}$$

and

$$\mathbb{P}\left[s_i^{(n)} = +1, s_j^{(n)} = +1 \,\middle|\, \mathbf{N}^{(n)}\right] = \frac{N_+^{(n)}(N_+^{(n)}-1)}{n(n-1)},$$

we have, for $i \neq j$,

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[J\left(\widetilde{R}_i^{(n)}, \widetilde{R}_j^{(n)}\right)\Big| \mathbf{s}^{(n)}\right]\Big| \mathbf{N}^{(n)}\right] \\
&= \frac{1}{n(n-1)}\left[1_{\left[N_-^{(n)} \geq 2\right]} \sum_{k=1}^{N_-^{(n)}} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_-^{(n)}} J\left(\frac{1}{2}\frac{k}{N_-^{(n)}+1}, \frac{1}{2}\frac{\ell}{N_-^{(n)}+1}\right)\right. \\
&\qquad + 1_{\left[N_-^{(n)} \geq 1, N_+^{(n)} \geq 1\right]} \sum_{k=1}^{N_-^{(n)}} \sum_{\ell=1}^{N_+^{(n)}} J\left(\frac{1}{2}\frac{k}{N_-^{(n)}+1}, \frac{1}{2}+\frac{1}{2}\frac{\ell}{N_+^{(n)}+1}\right) \\
&\qquad + 1_{\left[N_-^{(n)} \geq 1, N_+^{(n)} \geq 1\right]} \sum_{k=1}^{N_+^{(n)}} \sum_{\ell=1}^{N_-^{(n)}} J\left(\frac{1}{2}+\frac{1}{2}\frac{k}{N_+^{(n)}+1}, \frac{1}{2}\frac{\ell}{N_-^{(n)}+1}\right) \\
&\qquad \left. +1_{\left[N_+^{(n)} \geq 2\right]} \sum_{k=1}^{N_+^{(n)}} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_+^{(n)}} J\left(\frac{1}{2}+\frac{1}{2}\frac{k}{N_+^{(n)}+1}, \frac{1}{2}+\frac{1}{2}\frac{\ell}{N_+^{(n)}+1}\right)\right] \\
&= \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} J\left(\widetilde{R}_k^{(n)}, \widetilde{R}_\ell^{(n)}\right).
\end{aligned}
$$

So,

$$
\mathbb{E}\left[S_{2;\mathrm{appr}}^{(n)} \Big| \mathbf{N}^{(n)}\right] = \frac{\overline{c}^{(n)}}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} J\left(\widetilde{R}_i^{(n)}, \widetilde{R}_j^{(n)}\right). \tag{A.6}
$$

(iv) From (A.3), (A.4) and (A.5), it follows that, under $\mathcal{H}_h^{(n)}$, as $n \to \infty$,

$$
\begin{aligned}
S_{2;\mathrm{exact}}^{(n)} &= \mathbb{E}\left[T_2^{(n)} \Big| \mathbf{N}^{(n)}; \mathbf{R}^{(n)}\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(c_{ij}^{(n)} - \overline{c}^{(n)}\right) J\left(H(e_i^{(n)}), H(e_j^{(n)})\right) \\
&\quad + \frac{4\overline{c}^{(n)}}{n}\left\{N_-^{(n)}\left(N_-^{(n)}-1\right)\mu_J^{--} + N_-^{(n)} N_+^{(n)}\left(\mu_J^{-+} + \mu_J^{+-}\right) + N_+^{(n)}\left(N_+^{(n)}-1\right)\mu_J^{++}\right\} \\
&\quad + o_P(1/\sqrt{n}).
\end{aligned} \tag{A.7}
$$

On the other hand, (A.3), (A.4) and (A.6) imply that, under $\mathcal{H}_h^{(n)}$, as $n \to \infty$,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(c_{ij}^{(n)}-\overline{c}^{(n)}\right)J\left(H(e_i^{(n)}),H(e_j^{(n)})\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(c_{ij}^{(n)}-\overline{c}^{(n)}\right)J\left(\widetilde{R}_i^{(n)},\widetilde{R}_j^{(n)}\right)+o_{\mathrm{P}}(1/\sqrt{n}). \qquad \text{(A.8)}
\end{aligned}
$$

Appendix A.4. Proof of Proposition 2 - Part (i)

Under $\mathrm{P}_{h;\boldsymbol{\theta}}^{(n)}$, the error terms $e_1^{(n)}(\boldsymbol{\theta}),\ldots,e_n^{(n)}(\boldsymbol{\theta})$ are i.i.d. with density h and the random variables $H\left(e_1^{(n)}(\boldsymbol{\theta})\right),\ldots,H\left(e_n^{(n)}(\boldsymbol{\theta})\right)$ are i.i.d. $\mathcal{U}(0,1)$.

Throughout this proof, to prevent the notations from becoming overly complex, we simply write $e_i^{(n)}$, $s_i^{(n)}$, $R_i^{(n)}$, $\mathbf{N}^{(n)}$, $\mathbf{R}^{(n)}$, $\mathbf{G}^{(n)}$, ... for $e_i^{(n)}(\boldsymbol{\theta})$, $s_i^{(n)}(\boldsymbol{\theta})$, $R_i^{(n)}(\boldsymbol{\theta})$, $\mathbf{N}^{(n)}(\boldsymbol{\theta})$, $\mathbf{R}^{(n)}(\boldsymbol{\theta})$, $\mathbf{G}^{(n)}(\lambda)$, ... We then have that

$$
\boldsymbol{\Delta}_f^{(n)}=\begin{pmatrix}\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}\\ \Delta_{f;\lambda}^{(n)}\end{pmatrix},
$$

with

$$
\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(F(e_i^{(n)})\right)\mathbf{x}_i^{(n)}
$$

and

$$
\Delta_{f;\lambda}^{(n)}=L_{1;f}^{(n)}+L_{2;f}^{(n)}+L_{3;f}^{(n)}-\frac{1}{\sqrt{n}}\mathrm{tr}\left(\mathbf{G}^{(n)}\right),
$$

where

$$
\begin{aligned}
L_{1;f}^{(n)}&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(F(e_i^{(n)})\right)\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta},\\
L_{2;f}^{(n)}&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(F(e_i^{(n)})\right)F^{-1}\left(F(e_i^{(n)})\right)G_{ii}^{(n)},\\
L_{3;f}^{(n)}&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varphi_f\left(F(e_i^{(n)})\right)F^{-1}\left(F(e_j^{(n)})\right)G_{ij}^{(n)}.
\end{aligned}
$$

(i). Let us first consider $\boldsymbol{\Delta}_{f;\boldsymbol{\beta};\mathrm{exact}}^{(n)*}$. We have that

$$
\begin{aligned}
\boldsymbol{\Delta}_{f;\boldsymbol{\beta};\mathrm{exact}}^{(n)*}&=\mathbb{E}_{f;\boldsymbol{\theta}}^{(n)}\left[\boldsymbol{\Delta}_{f;\boldsymbol{\beta}}^{(n)}\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]\\
&=\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(U_i^{(n)}\right)\mathbf{x}_i^{(n)}\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right],
\end{aligned}
$$

this latter expectation being taken under the assumption that $F(e_i^{(n)}) = U_i^{(n)}$, $i = 1, \ldots, n$, are i.i.d. $\mathcal{U}(0,1)$. Applying results (A.1) and (A.2) with $\mathbf{c}_i^{(n)} = \mathbf{x}_i^{(n)}$ for $i = 1, \ldots, n$, and $J(u) = \varphi_f(u)$ such that

$$\mu_J = \int_0^1 \varphi_f(u)du = \int_{-\infty}^{\infty} \phi_f(e)f(e)de = 0,\ \mu_J^- = \int_0^{1/2} \varphi_f(u)du = -f(0),\ \mu_J^+ = \int_{1/2}^1 \varphi_f(u)du = f(0),$$

we have that, under $\mathrm{P}_{h;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\begin{aligned}\boldsymbol{\Delta}_{f;\boldsymbol{\beta};\text{exact}}^{(n)*} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(H(e_i^{(n)})\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) + \overline{\mathbf{x}}^{(n)} \frac{2f(0)}{\sqrt{n}}\left(N_+^{(n)} - N_-^{(n)}\right) + o_{\mathrm{P}}(1)\\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(\widetilde{R}_i^{(n)}\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) + \overline{\mathbf{x}}^{(n)} \frac{2f(0)}{\sqrt{n}} \sum_{i=1}^n s_i^{(n)} + o_{\mathrm{P}}(1).\end{aligned}$$

(ii). Applying once again (A.1) and (A.2) with $J(u) = \varphi_f(u)$ and $c_i^{(n)} = \mathbf{G}_{i\bullet}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta}$, we get that, under $\mathrm{P}_{h;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\begin{aligned}&\mathbb{E}_{f;\boldsymbol{\theta}}^{(n)}\left[L_{1;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right]\\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(H(e_i^{(n)})\right)\left(\mathbf{G}_{i\bullet}^{(n)} - \overline{\mathbf{G}}_{\bullet}^{(n)}\right)\mathbf{X}^{(n)}\boldsymbol{\beta} + \overline{\mathbf{G}}_{\bullet}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta} \frac{2f(0)}{\sqrt{n}}\left(N_+^{(n)} - N_-^{(n)}\right) + o_{\mathrm{P}}(1)\\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi_f\left(\widetilde{R}_i^{(n)}\right)\left(\mathbf{G}_{i\bullet}^{(n)} - \overline{\mathbf{G}}_{\bullet}^{(n)}\right)\mathbf{X}^{(n)}\boldsymbol{\beta} + \overline{\mathbf{G}}_{\bullet}^{(n)}\mathbf{X}^{(n)}\boldsymbol{\beta} \frac{2f(0)}{\sqrt{n}} \sum_{i=1}^n s_i^{(n)} + o_{\mathrm{P}}(1).\end{aligned}$$

(iii). Applying (A.1) and (A.2) with $c_i^{(n)} = G_{ii}^{(n)}$ — it gives $\overline{c}^{(n)} = \mathrm{tr}\left(\mathbf{G}^{(n)}\right)/n$ — and $J(u) = \varphi_f(u)F^{-1}(u)$, we get

$$\mu_J = \int_0^1 \varphi_f(u)F^{-1}(u)du = 1 = \mathcal{J}_f$$

and

$$\mu_J^- = \int_0^{1/2} \varphi_f(u)F^{-1}(u)du = \frac{1}{2} = \int_{1/2}^1 \varphi_f(u)F^{-1}(u)du = \mu_J^+.$$

Then, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$
\begin{aligned}
&\mathbb{E}^{(n)}_{f;\boldsymbol{\theta}}\left[L^{(n)}_{2;f}\middle|\mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(H(e_i^{(n)})\right)F^{-1}\left(H(e_i^{(n)})\right)\left(G^{(n)}_{ii} - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right) \\
&\qquad + \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{\sqrt{n}}\left(2\frac{N_-^{(n)}}{n}\frac{1}{2} + 2\frac{N_+^{(n)}}{n}\frac{1}{2}\right) + o_{\mathrm{P}}(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(H(e_i^{(n)})\right)F^{-1}\left(H(e_i^{(n)})\right)\left(G^{(n)}_{ii} - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right) + \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{\sqrt{n}} + o_{\mathrm{P}}(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(\widetilde{R}_i^{(n)}\right)F^{-1}\left(\widetilde{R}_i^{(n)}\right)\left(G^{(n)}_{ii} - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right) + \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{\sqrt{n}} + o_{\mathrm{P}}(1).
\end{aligned}
$$

(iv). Finally, applying (A.7) with $c_{ij}^{(n)} = G_{ij}^{(n)}$ and $J(u_1, u_2) = \varphi_f(u_1)F^{-1}(u_2)$, such that

$$
\overline{c}^{(n)} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}^{(n)} = \frac{g^{(n)}}{n(n-1)},
$$

and, since $\int_0^{1/2}\varphi_f(u)du = -f(0)$, $\int_{1/2}^{1}\varphi_f(u)du = f(0)$, $\int_0^{1/2}F^{-1}(u)du = \int_{-\infty}^{0} ef(e)de$, $\int_{1/2}^{1}F^{-1}(u)du = \int_0^{\infty} ef(e)de$,

$$
\begin{aligned}
\mu_J^{--} &= -f(0)\int_{-\infty}^{0} ef(e)de, \quad \mu_J^{-+} = -f(0)\int_0^{\infty} ef(e)de, \\
\mu_J^{+-} &= f(0)\int_{-\infty}^{0} ef(e)de, \quad \mu_J^{++} = f(0)\int_0^{\infty} ef(e)de.
\end{aligned}
$$

We then have that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$
\begin{aligned}
&\mathbb{E}^{(n)}_{f;\boldsymbol{\theta}}\left[L^{(n)}_{3;f}\,\middle|\,\mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varphi_f\left(H(e_i^{(n)})\right) F^{-1}\left(H(e_j^{(n)})\right)\left(G_{ij}^{(n)} - \frac{g^{(n)}}{n(n-1)}\right) \\
&\quad + \frac{4}{\sqrt{n}}\,\frac{g^{(n)}}{n(n-1)} f(0)\left\{-N_-^{(n)}\left(N_-^{(n)}-1\right)\int_{-\infty}^{0} ef(e)de \right. \\
&\qquad\qquad + N_-^{(n)}N_+^{(n)}\left(\int_{-\infty}^{0} ef(e)de - \int_{0}^{\infty} ef(e)de\right) \\
&\qquad\qquad \left. + N_+^{(n)}\left(N_+^{(n)}-1\right)\int_{0}^{\infty} ef(e)de\right\} + o_{\mathrm{P}}(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varphi_f\left(H(e_i^{(n)})\right) F^{-1}\left(H(e_j^{(n)})\right)\left(G_{ij}^{(n)} - \frac{g^{(n)}}{n(n-1)}\right) \\
&\quad + 4f(0)\frac{g^{(n)}}{n}\sqrt{n}\left\{\frac{N_-^{(n)}N_+^{(n)} - N_-^{(n)}\left(N_-^{(n)}-1\right)}{n(n-1)}\int_{-\infty}^{0} ef(e)de \right. \\
&\qquad\qquad \left. + \frac{N_+^{(n)}\left(N_+^{(n)}-1\right) - N_-^{(n)}N_+^{(n)}}{n(n-1)}\int_{0}^{\infty} ef(e)de\right\} + o_{\mathrm{P}}(1).
\end{aligned}
$$

But

$$
\sqrt{n}\frac{N_-^{(n)}N_+^{(n)} - N_-^{(n)}\left(N_-^{(n)}-1\right)}{n(n-1)} = \frac{1}{2}\,\frac{N_+^{(n)} - N_-^{(n)}}{\sqrt{n}} + o_{\mathrm{P}}(1)
$$

and, similarly,

$$
\sqrt{n}\frac{N_+^{(n)}\left(N_+^{(n)}-1\right) - N_-^{(n)}N_+^{(n)}}{n(n-1)} = \frac{1}{2}\,\frac{N_+^{(n)} - N_-^{(n)}}{\sqrt{n}} + o_{\mathrm{P}}(1).
$$

Hence,

$$
\begin{aligned}
&\sqrt{n}\left\{\frac{N_-^{(n)}N_+^{(n)} - N_-^{(n)}\left(N_-^{(n)}-1\right)}{n(n-1)}\int_{-\infty}^{0} ef(e)de + \frac{N_+^{(n)}\left(N_+^{(n)}-1\right) - N_-^{(n)}N_+^{(n)}}{n(n-1)}\int_{0}^{\infty} ef(e)de\right\} \\
&= \frac{1}{2}\,\frac{N_+^{(n)} - N_-^{(n)}}{\sqrt{n}}\left\{\int_{-\infty}^{0} ef(e)de + \int_{0}^{\infty} ef(e)de\right\} + o_{\mathrm{P}}(1) \\
&= \frac{1}{2}\,\frac{N_+^{(n)} - N_-^{(n)}}{\sqrt{n}}\int_{-\infty}^{\infty} ef(e)de + o_{\mathrm{P}}(1) = \frac{1}{2}\,\frac{N_+^{(n)} - N_-^{(n)}}{\sqrt{n}}\mu_f + o_{\mathrm{P}}(1).
\end{aligned}
$$

In conclusion, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\mathbb{E}^{(n)}_{f;\boldsymbol{\theta}}\left[L^{(n)}_{3;f}\,\middle|\,\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varphi_f\left(H(e_i^{(n)})\right)F^{-1}\left(H(e_j^{(n)})\right)\left(G_{ij}^{(n)}-\frac{g^{(n)}}{n(n-1)}\right)+2f(0)\mu_f\frac{g^{(n)}}{n}\,\frac{N_+^{(n)}-N_-^{(n)}}{\sqrt{n}}+o_{\mathrm{P}}(1)$$

or still, applying (A.8),

$$\mathbb{E}^{(n)}_{f;\boldsymbol{\theta}}\left[L^{(n)}_{3;f}\,\middle|\,\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varphi_f\left(\widetilde{R}_i^{(n)}\right)F^{-1}\left(\widetilde{R}_j^{(n)}\right)\left(G_{ij}^{(n)}-\frac{g^{(n)}}{n(n-1)}\right)+2f(0)\mu_f\frac{g^{(n)}}{n}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^{(n)}+o_{\mathrm{P}}(1).$$

(v). The asymptotic equivalences (8) directly follow from the various results stated above in (i) to (iv).

(vi) Determination of the expression of matrix $\mathbf{I}_f^*(\boldsymbol{\theta})$ in Part (i) of Prop. 2. Let us consider

$$\boldsymbol{\Delta}^{(n)*}_{fh} = \begin{pmatrix}\boldsymbol{\Delta}^{(n)*}_{fh;\boldsymbol{\beta}}\\ \Delta^{(n)*}_{fh;\lambda}\end{pmatrix}$$

where

$$\boldsymbol{\Delta}^{(n)*}_{fh;\boldsymbol{\beta}} = \mathbf{B}^{(n)*}_{1;fh}+\mathbf{B}^{(n)*}_{2;fh} \tag{A.9}$$

with

$$\mathbf{B}^{(n)*}_{1;fh} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(H(e_i^{(n)})\right)\left(\mathbf{x}_i^{(n)}-\overline{\mathbf{x}}^{(n)}\right), \tag{A.10}$$

$$\mathbf{B}^{(n)*}_{2;fh} = 2f(0)\overline{\mathbf{x}}^{(n)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^{(n)}, \tag{A.11}$$

and

$$\Delta^{(n)*}_{fh;\lambda} = L^{(n)*}_{1;fh}+L^{(n)*}_{2;fh}+L^{(n)*}_{3;fh}+L^{(n)*}_{4;fh} \tag{A.12}$$

with

$$L^{(n)*}_{1;fh} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(H(e_i^{(n)})\right)\left(\mathbf{G}^{(n)}_{i\bullet}-\overline{\mathbf{G}}^{(n)}_{\bullet}\right)\mathbf{X}^{(n)}\boldsymbol{\beta}, \tag{A.13}$$

$$L^{(n)*}_{2;fh} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_f\left(H(e_i^{(n)})\right)F^{-1}\left(H(e_i^{(n)})\right)\left(G_{ii}^{(n)}-\frac{\mathrm{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right), \tag{A.14}$$

$$L^{(n)*}_{3;fh} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varphi_f\left(H(e_i^{(n)})\right)F^{-1}\left(H(e_j^{(n)})\right)\left(G_{ij}^{(n)}-\frac{g^{(n)}}{n(n-1)}\right), \tag{A.15}$$

$$L^{(n)*}_{4;fh} = 2f(0)\left(\overline{\mathbf{G}}^{(n)}_{\bullet}\mathbf{X}^{(n)}\boldsymbol{\beta}+\mu_f\frac{g^{(n)}}{n}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^{(n)}. \tag{A.16}$$

Under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, the random variables $H(e_i^{(n)})$, $i = 1, \ldots, n$, are i.i.d. $\mathcal{U}(0,1)$, which implies that, for all $i = 1, \ldots, n$ (see the notations introduced in Section Appendix A.1):

$$\mathbb{E}\left[F^{-1}\left(H(e_i^{(n)})\right)\right] = \mu_f, \qquad \mathbb{E}\left[\left(F^{-1}\left(H(e_i^{(n)})\right)\right)^2\right] = \nu_f,$$

$$\mathbb{E}\left[\varphi_f\left(H(e_i^{(n)})\right)\right] = 0, \qquad \mathbb{E}\left[\varphi_f^2\left(H(e_i^{(n)})\right)\right] = \mathcal{I}_f,$$

$$\mathbb{E}\left[\varphi_f\left(H(e_i^{(n)})\right) F^{-1}\left(H(e_i^{(n)})\right)\right] = 1, \qquad \mathbb{E}\left[\varphi_f^2\left(H(e_i^{(n)})\right) F^{-1}\left(H(e_i^{(n)})\right)\right] = \mathcal{K}_f,$$

$$\mathbb{E}\left[\varphi_f^2\left(H(e_i^{(n)})\right)\left(F^{-1}\left(H(e_i^{(n)})\right)\right)^2\right] = \mathcal{Q}_f.$$

Moreover, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, the signs $s_i^{(n)}$, $i = 1, \ldots, n$, are i.i.d. $\mathcal{U}\{-1, +1\}$. Since

$$\sum_{i=1}^n \left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}} - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}\right) = \mathbf{0}, \quad \sum_{i=1}^n \left(G^{(n)}_{ii} - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right) = 0, \quad \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(G^{(n)}_{ij} - \frac{g^{(n)}}{n(n-1)}\right) = 0,$$

we have that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, $\mathbb{E}\left[\boldsymbol{\Delta}^{(n)*}_{fh}\right] = \mathbf{0}$. Considering the decompositions (A.9) and (A.12), similar calculations than those summarized in Appendix Appendix A.2 provide the expressions of $\mathbb{E}\left[\boldsymbol{\Delta}^{(n)*}_{fh;\boldsymbol{\beta}}\left(\boldsymbol{\Delta}^{(n)*}_{fh;\boldsymbol{\beta}}\right)^{\mathrm{T}}\right]$, $\mathbb{E}\left[\left(\Delta^{(n)*}_{fh;\lambda}\right)^2\right]$ and $\mathbb{E}\left[\boldsymbol{\Delta}^{(n)*}_{fh;\boldsymbol{\beta}}\Delta^{(n)*}_{fh;\lambda}\right]$, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, and, at the same time, the expression of matrix $\mathbf{I}^*_f(\boldsymbol{\theta})$.

Appendix A.5. Proof of Proposition 2 - Part (ii)

Note first that, if density h belongs to $\mathcal{F}$, the sequence of parametric submodels $\mathcal{E}_h^{(n)}$, $n \in \mathbb{N}$, is ULAN, which implies that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned} \Lambda^{(n)}_{h;\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}/\boldsymbol{\theta}} &\stackrel{\text{def}}{=} \ln\left(\frac{L_h\left(\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{L_h\left(\boldsymbol{\theta}\,\middle|\mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}\right) \\ &= \boldsymbol{\tau}^{\mathrm{T}}\boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau} + o_{\mathrm{P}}(1) \end{aligned} \tag{A.17}$$

and

$$\Lambda^{(n)}_{h;\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}/\boldsymbol{\theta}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(-\frac{1}{2}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau}, \boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau}\right).$$

Let us now consider the sign-and-rank based central sequence $\boldsymbol{\Delta}^{(n)*}_{f;\mathrm{exact/appr}}(\boldsymbol{\theta})$ whose the *exact* or *approximate* score functions are associated with the reference density f. Using (A.17) and Proposition 2-Part (i) — namely the asymptotic equivalences (8) —, we obtain that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{pmatrix} \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{exact/appr}}(\boldsymbol{\theta}) \\ \Lambda^{(n)}_{h;\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}/\boldsymbol{\theta}} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ -\frac{1}{2}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau} \end{pmatrix}, \begin{pmatrix} \mathbf{I}^*_f(\boldsymbol{\theta}) & \mathbf{I}^*_{fh}(\boldsymbol{\theta})\boldsymbol{\tau} \\ \boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}^*_{fh}(\boldsymbol{\theta}) & \boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau} \end{pmatrix}\right).$$

It then follows from the Le Cam's Third Lemma that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}}$, as $n \to \infty$,

$$\boldsymbol{\Delta}^{(n)*}_{f;\text{exact/appr}}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\boldsymbol{\tau}, \mathbf{I}^*_f(\boldsymbol{\theta})\right),$$

which is what we wanted to prove.

Appendix A.6. Proof of Proposition 3

Let $\mathbf{B}^{(n)*}_{1;fh}(\boldsymbol{\theta})$, $L^{(n)*}_{1;fh}(\boldsymbol{\theta})$, $L^{(n)*}_{2;fh}(\boldsymbol{\theta})$ and $L^{(n)*}_{3;fh}(\boldsymbol{\theta})$ be defined by (A.10), (A.13), (A.14) and (A.15), respectively. Define the statistic

$$\mathbf{D}^{(n)}_{fh}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}) \\ S^{(n)}(\boldsymbol{\theta}) \end{pmatrix}$$

with

$$\mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{T}^{(n)}_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) \\ T^{(n)}_{fh;\lambda}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \mathbf{B}^{(n)*}_{1;fh}(\boldsymbol{\theta}) \\ L^{(n)*}_{1;fh}(\boldsymbol{\theta}) + L^{(n)*}_{2;fh}(\boldsymbol{\theta}) + L^{(n)*}_{3;fh}(\boldsymbol{\theta}) \end{pmatrix}$$

and

$$S^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s^{(n)}_i(\boldsymbol{\theta}).$$

Note that

$$\begin{aligned}
\boldsymbol{\Delta}^{(n)*}_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \mathbf{T}^{(n)}_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) + 2f(0)\overline{\mathbf{x}}^{(n)} S^{(n)}(\boldsymbol{\theta}), \\
\Delta^{(n)*}_{fh;\lambda}(\boldsymbol{\theta}) &= T^{(n)}_{fh;\lambda}(\boldsymbol{\theta}) + 2f(0)\left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n}\right) S^{(n)}(\boldsymbol{\theta}).
\end{aligned}$$

Under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, $\mathbf{D}^{(n)}_{fh}(\boldsymbol{\theta})$ and the parametric central sequence $\boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta})$ are jointly asymptotically normal. More precisely, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{pmatrix} \boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta}) \\ \mathbf{D}^{(n)}_{fh}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta}) \\ \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}) \\ S^{(n)}(\boldsymbol{\theta}) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{I}_h(\boldsymbol{\theta}) & \boldsymbol{\Gamma}_{fh}(\boldsymbol{\theta}) & \boldsymbol{\Omega}_h(\boldsymbol{\theta}) \\ \boldsymbol{\Gamma}^{\mathrm{T}}_{fh}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_f(\boldsymbol{\theta}) & \mathbf{0} \\ \boldsymbol{\Omega}^{\mathrm{T}}_h(\boldsymbol{\theta}) & \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix}\right),$$

where $\mathbf{I}_h(\boldsymbol{\theta})$ is defined in Proposition 1,

$$\boldsymbol{\Gamma}_{fh}(\boldsymbol{\theta}) = \lim_{n\to\infty} \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta})\left(\mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right] = \begin{pmatrix} \boldsymbol{\Gamma}_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{\Gamma}_{fh;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) \\ \boldsymbol{\Gamma}_{fh;\lambda,\boldsymbol{\beta}}(\boldsymbol{\theta}) & \Gamma_{fh;\lambda}(\boldsymbol{\theta}) \end{pmatrix} \tag{A.18}$$

with

$$\begin{aligned}
\boldsymbol{\Gamma}_{fh;\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{\mathcal{I}_{fh}\,\mathbf{C}_{\mathbf{x}}^{(n)*}\right\},\\
\boldsymbol{\Gamma}_{fh;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{\mathcal{I}_{fh}\,\mathbf{C}_{\mathbf{x},\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) + \mathcal{K}_{hf}\,\mathbf{C}_{\mathbf{x},\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\mu_f\,\mathbf{C}_{\mathbf{x},\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\right\},\\
\boldsymbol{\Gamma}_{fh;\lambda,\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{\mathcal{I}_{fh}\,\mathbf{C}_{\mathbf{x},\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) + \mathcal{K}_{fh}\,\mathbf{C}_{\mathbf{x},\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\mu_h\,\mathbf{C}_{\mathbf{x},\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\right\},\\
\Gamma_{fh;\lambda}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\Big\{\mathcal{I}_{fh}\,C_{\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) + (\mathcal{Q}_{fh}-1)\,C_{\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\nu_{fh}\,C_{\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\\
&\qquad + (\mathcal{K}_{fh}\mu_f + \mathcal{K}_{hf}\mu_h)\,C_{\mathbf{G},3}^{(n)*}(\boldsymbol{\theta}) + \mathcal{J}_{fh}\mathcal{J}_{hf}\,C_{\mathbf{G},4}^{(n)*}(\boldsymbol{\theta})\\
&\qquad + \mathcal{I}_{fh}\mu_f\mu_h\,C_{\mathbf{G},5}^{(n)*}(\boldsymbol{\theta})\\
&\qquad + (\mathcal{K}_{fh} + \mathcal{K}_{hf})\,C_{\mathbf{Gx},\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_{fh}\,(\mu_f + \mu_h)\,C_{\mathbf{Gx},\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\Big\},
\end{aligned}$$

$$\boldsymbol{\Sigma}_f(\boldsymbol{\theta}) = \lim_{n\to\infty}\mathbb{E}\left[\mathbf{T}_{fh}^{(n)}(\boldsymbol{\theta})\left(\mathbf{T}_{fh}^{(n)}(\boldsymbol{\theta})\right)^{\mathrm{T}}\right] = \begin{pmatrix}\boldsymbol{\Sigma}_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta})\\ \boldsymbol{\Sigma}_{f;\lambda,\boldsymbol{\beta}}(\boldsymbol{\theta}) & \Sigma_{f;\lambda}(\boldsymbol{\theta})\end{pmatrix} \tag{A.19}$$

with

$$\begin{aligned}
\boldsymbol{\Sigma}_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{\mathcal{I}_f\mathbf{C}_{\mathbf{x}}^{(n)*}\right\},\\
\boldsymbol{\Sigma}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{\mathcal{I}_f\,\mathbf{C}_{\mathbf{x},\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) + \mathcal{K}_f\,\mathbf{C}_{\mathbf{x},\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_f\mu_f\,\mathbf{C}_{\mathbf{x},\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\right\},\\
\boldsymbol{\Sigma}_{f;\lambda,\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \boldsymbol{\Sigma}_{f;\boldsymbol{\beta},\lambda}^{\mathrm{T}}(\boldsymbol{\theta}),\\
\Sigma_{f;\lambda}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\Big\{\mathcal{I}_f\,C_{\mathbf{Gx}}^{(n)*}(\boldsymbol{\theta}) + (\mathcal{Q}_f-1)\,C_{\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_f\nu_f\,C_{\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\\
&\qquad + 2\,\mathcal{K}_f\mu_f\,C_{\mathbf{G},3}^{(n)*}(\boldsymbol{\theta}) + C_{\mathbf{G},4}^{(n)*}(\boldsymbol{\theta}) + \mathcal{I}_f\mu_f^2\,C_{\mathbf{G},5}^{(n)*}(\boldsymbol{\theta})\\
&\qquad + 2\,\mathcal{K}_f\,C_{\mathbf{Gx},\mathbf{G},1}^{(n)*}(\boldsymbol{\theta}) + 2\,\mathcal{I}_f\mu_f\,C_{\mathbf{Gx},\mathbf{G},2}^{(n)*}(\boldsymbol{\theta})\Big\},
\end{aligned}$$

and

$$\boldsymbol{\Omega}_h(\boldsymbol{\theta}) = \lim_{n\to\infty}\mathbb{E}\left[\boldsymbol{\Delta}_h^{(n)}(\boldsymbol{\theta})S^{(n)}(\boldsymbol{\theta})\right] = \begin{pmatrix}\boldsymbol{\Omega}_{h;\boldsymbol{\beta}}(\boldsymbol{\theta})\\ \Omega_{h;\lambda}(\boldsymbol{\theta})\end{pmatrix} \tag{A.20}$$

with

$$\begin{aligned}
\boldsymbol{\Omega}_{h;\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{2h(0)\overline{\mathbf{x}}^{(n)}\right\},\\
\Omega_{h;\lambda}(\boldsymbol{\theta}) &= \lim_{n\to\infty}\left\{2h(0)\left(\overline{\mathbf{G}}_{\boldsymbol{\cdot}}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_h\frac{g^{(n)}(\lambda)}{n}\right)\right\}.
\end{aligned}$$

Consider now an arbitrary sequence of the form $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + \boldsymbol{\tau}^{(n)}/\sqrt{n} + O(1/\sqrt{n})$, with a bounded sequence $\boldsymbol{\tau}^{(n)} \to \boldsymbol{\tau}$ as $n \to \infty$. Under $\mathrm{P}_{h;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\begin{pmatrix}\mathbf{D}_{fh}^{(n)}(\boldsymbol{\theta})\\ \Lambda_{h;\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}/\boldsymbol{\theta}}^{(n)}\end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix}\mathbf{0}\\ -\frac{1}{2}\boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau}\end{pmatrix}, \begin{pmatrix}\mathbf{J}_f(\boldsymbol{\theta}) & \mathbf{K}_{fh}(\boldsymbol{\theta})\boldsymbol{\tau}\\ \boldsymbol{\tau}^{\mathrm{T}}\mathbf{K}_{fh}^{\mathrm{T}}(\boldsymbol{\theta}) & \boldsymbol{\tau}^{\mathrm{T}}\mathbf{I}_h(\boldsymbol{\theta})\boldsymbol{\tau}\end{pmatrix}\right)$$

with

$$\mathbf{J}_f(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}_f(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{K}_{fh}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Gamma}^{\mathrm{T}}_{fh}(\boldsymbol{\theta}) \\ \boldsymbol{\Omega}^{\mathrm{T}}_h(\boldsymbol{\theta}) \end{pmatrix}.$$

Moreover, we have the following *local asymptotic linearity property*: under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{pmatrix} \boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta}^{(n)}) \\ \mathbf{D}^{(n)}_{fh}(\boldsymbol{\theta}^{(n)}) \end{pmatrix} - \begin{pmatrix} \boldsymbol{\Delta}^{(n)}_h(\boldsymbol{\theta}) \\ \mathbf{D}^{(n)}_{fh}(\boldsymbol{\theta}) \end{pmatrix} + \begin{pmatrix} \mathbf{I}_h(\boldsymbol{\theta}) \\ \mathbf{K}_{fh}(\boldsymbol{\theta}) \end{pmatrix} \boldsymbol{\tau}^{(n)} = o_{\mathrm{P}}(1). \tag{A.21}$$

With respect to the central sequence $\boldsymbol{\Delta}^{(n)}_h(\cdot)$, expression (A.21) actually follows from the ULAN of the sequence of parametric submodels $\mathcal{E}^{(n)}_h$, $n \in \mathbb{N}$. The expression (A.21) holds for $\mathbf{T}^{(n)}_{fh}(\cdot)$ if $f \in \mathcal{F}$ in such a way that the sequence of parametric submodels $\mathcal{E}^{(n)}_f$, $n \in \mathbb{N}$, is ULAN. The asymptotic linearity of $S^{(n)}(\cdot)$ is classical in the literature on sign tests.

It follows from (A.21) that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned} \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}^{(n)}) &= \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}) - \boldsymbol{\Gamma}^{\mathrm{T}}_{fh}(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1), \\ S^{(n)}(\boldsymbol{\theta}^{(n)}) &= S^{(n)}(\boldsymbol{\theta}) - \boldsymbol{\Omega}^{\mathrm{T}}_h(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1). \end{aligned}$$

We then have that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned} \boldsymbol{\Delta}^{(n)*}_{fh}(\boldsymbol{\theta}^{(n)}) &= \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}^{(n)}) + \boldsymbol{\Omega}_f(\boldsymbol{\theta}^{(n)})S^{(n)}(\boldsymbol{\theta}^{(n)}) + o_{\mathrm{P}}(1) \\ &= \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}^{(n)}) + \boldsymbol{\Omega}_f(\boldsymbol{\theta})S^{(n)}(\boldsymbol{\theta}^{(n)}) + o_{\mathrm{P}}(1) \\ &= \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}) - \boldsymbol{\Gamma}^{\mathrm{T}}_{fh}(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + \boldsymbol{\Omega}_f(\boldsymbol{\theta})\left[S^{(n)}(\boldsymbol{\theta}) - \boldsymbol{\Omega}^{\mathrm{T}}_h(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)}\right] + o_{\mathrm{P}}(1) \\ &= \mathbf{T}^{(n)}_{fh}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_f(\boldsymbol{\theta})S^{(n)}(\boldsymbol{\theta}) - \left[\boldsymbol{\Gamma}^{\mathrm{T}}_{fh}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_f(\boldsymbol{\theta})\boldsymbol{\Omega}^{\mathrm{T}}_h(\boldsymbol{\theta})\right]\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \\ &= \boldsymbol{\Delta}^{(n)*}_{fh}(\boldsymbol{\theta}) - \left[\boldsymbol{\Gamma}^{\mathrm{T}}_{fh}(\boldsymbol{\theta}) + \boldsymbol{\Omega}_f(\boldsymbol{\theta})\boldsymbol{\Omega}^{\mathrm{T}}_h(\boldsymbol{\theta})\right]\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1) \\ &= \boldsymbol{\Delta}^{(n)*}_{fh}(\boldsymbol{\theta}) - \mathbf{I}^{*}_{fh}(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1). \end{aligned} \tag{A.22}$$

The assumption that the preliminary estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ is $\sqrt{n}$-consistent and locally discrete implies that it may be treated as if it were of the form $\boldsymbol{\theta}^{(n)} + \boldsymbol{\tau}^{(n)}/\sqrt{n}$ for some sequence $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta} + O(1/\sqrt{n})$ and some bounded and deterministic sequence $\boldsymbol{\tau}^{(n)}$, $n \in \mathbb{N}$. We then have, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned} \sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{f;\mathrm{R\&S}} - \boldsymbol{\theta}\right) &= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\widehat{\mathbf{I}}^{(n)*}_{f\widehat{h}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) \\ &= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^{*}_{fh}(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) + o_{\mathrm{P}}(1) \\ &= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^{*}_{fh}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) + o_{\mathrm{P}}(1), \end{aligned}$$

given that $\widehat{\mathbf{I}}^{(n)*}_{f\widehat{h}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})$ is a consistent estimate of $\mathbf{I}^{*}_{fh}(\widetilde{\boldsymbol{\theta}}^{(n)})$ and matrix $\mathbf{I}^{*}_{fh}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$.

Using (8) and (A.22), we obtain that, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{f;\mathrm{R\&S}} - \boldsymbol{\theta}\right) \\
&= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{fh}(\widetilde{\boldsymbol{\theta}}^{(n)}) + o_{\mathrm{P}}(1) \\
&= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\right)^{-1} \left[\boldsymbol{\Delta}^{(n)*}_{fh}(\boldsymbol{\theta}) - \mathbf{I}^*_{fh}(\boldsymbol{\theta})\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right)\right] + o_{\mathrm{P}}(1) \\
&= \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{fh}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1).
\end{aligned}
$$

Since, under $\mathrm{P}^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$, $\boldsymbol{\Delta}^{(n)*}_{fh}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}^*_f(\boldsymbol{\theta})\right)$, we may conclude that

$$
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{f;\mathrm{R\&S}} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\right)^{-1} \mathbf{I}^*_f(\boldsymbol{\theta}) \left(\mathbf{I}^*_{fh}(\boldsymbol{\theta})\right)^{-1}\right).
$$

Appendix A.7. Consistency condition for the estimators of the functions φ_f and ψ_f

Let $f \in \mathcal{F}$, $\varphi_f(u) = \phi_f\left(F^{-1}(u)\right)$ and $\psi_f(u) = F^{-1}(u)$ for $u \in (0,1)$. Let $\mathbf{e}^{(n)}_{(\cdot)} = (e^{(n)}_{(1)}, \ldots, e^{(n)}_{(n)})^{\mathrm{T}}$ be the vector of ordered residuals: we have $\mathbf{e}^{(n)}_{(\cdot)} = \left(\left(\mathbf{e}^{(N^{(n)}_-)}_{(\cdot)-}\right)^{\mathrm{T}}, \left(\mathbf{e}^{(N^{(n)}_+)}_{(\cdot)+}\right)^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\mathbf{e}^{(N^{(n)}_-)}_{(\cdot)-}$ is the vector of the $N^{(n)}_-$ negative ordered residuals and $\mathbf{e}^{(N^{(n)}_+)}_{(\cdot)+}$ is the vector of the $N^{(n)}_+$ positive ordered residuals.

Let $\widehat{\varphi}_{f,n}$ and $\widehat{\psi}_{f,n}$ be estimators of the functions φ_f and ψ_f satisfying the following condition.

Condition 1.

$$
\widehat{\varphi}_{f,n}(u) = \begin{cases} \widehat{\varphi}^-_{f,n}(u) & \text{if } u \in (0, 1/2] \\ \widehat{\varphi}^+_{f,n}(u) & \text{if } u \in (1/2, 1) \end{cases}, \quad \widehat{\psi}_{f,n}(u) = \begin{cases} \widehat{\psi}^-_{f,n}(u) & \text{if } u \in (0, 1/2] \\ \widehat{\psi}^+_{f,n}(u) & \text{if } u \in (1/2, 1) \end{cases}
$$

where $\widehat{\varphi}^-_{f,n}$ and $\widehat{\psi}^-_{f,n}$ are measurable with respect to $\mathbf{e}^{N^{(n)}_-}_{(\cdot)-}$ and consistent in the sense that, as $n \to \infty$,

$$
\mathbb{E}_f\left[\left\{\widehat{\varphi}^-_{f,n}\left(\frac{1}{2}\,\frac{R^{(n)}_1}{N^{(n)}_- + 1}\right) - \varphi_f\left(\frac{1}{2}\,\frac{R^{(n)}_1}{N^{(n)}_- + 1}\right)\right\}^2 \middle|\, \mathbf{e}^{(N^{(n)}_-)}_{(\cdot)-}, s^{(n)}_1 = -1\right] = o_{\mathrm{P}}(1)
$$

(idem for $\widehat{\psi}^-_{f,n}$ and ψ_f) and where $\widehat{\varphi}^+_{f,n}$ and $\widehat{\psi}^+_{f,n}$ are measurable with respect to $\mathbf{e}^{N^{(n)}_+}_{(\cdot)+}$ and consistent in the sense that, as $n \to \infty$,

$$
\mathbb{E}_f\left[\left\{\widehat{\varphi}^+_{f,n}\left(\frac{1}{2} + \frac{1}{2}\,\frac{R^{(n)}_1 - (n - N^{(n)}_+)}{N^{(n)}_+ + 1}\right) - \varphi_f\left(\frac{1}{2} + \frac{1}{2}\,\frac{R^{(n)}_1 - (n - N^{(n)}_+)}{N^{(n)}_+ + 1}\right)\right\}^2 \middle|\, \mathbf{e}^{(N^{(n)}_+)}_{(\cdot)+}, s^{(n)}_1 = +1\right] = o_{\mathrm{P}}(1)
$$

(idem for $\widehat{\psi}^{+}_{f,n}$ and ψ_f).

Using arguments similar to those presented in the proof of Proposition 3.4 of Hallin and Werker (2003) and in Vermandele (2000), it is possible to prove that Condition 1 is a *sufficient* condition for the asymptotic equivalence (18).

It is quite natural to define $\widehat{\psi}_{f,n}$ as follows: for $i = 1, \ldots, n$, if $s_i^{(n)} = -1$ and, consequently, $R_i^{(n)} \in \{1, \ldots, N_-^{(n)}\}$, take

$$\widehat{\psi}^{-}_{f,n}\left(\frac{1}{2}\,\frac{R_i^{(n)}}{N_-^{(n)}+1}\right) = e^{(N_-^{(n)})}_{\left(R_i^{(n)}\right)-} = e_i^{(n)}\,;$$

if $s_i^{(n)} = +1$ and, consequently, $R_i^{(n)} \in \{N_-^{(n)}+1, \ldots, n\}$, take

$$\widehat{\psi}^{+}_{f,n}\left(\frac{1}{2} + \frac{1}{2}\,\frac{R_i^{(n)} - (n - N_+^{(n)})}{N_+^{(n)}+1}\right) = e^{(N_+^{(n)})}_{\left(R_i^{(n)} - (n - N_+^{(n)})\right)-} = e_i^{(n)}.$$

On the other hand, the definition of an estimator $\widehat{\varphi}_{f,n}$ of φ_f that satisfies Condition 1 needs further development. Note that Hájek and Sidák (1967, Chapter VI, Section 1.5) propose an appropriately consistent estimator of φ_f in the context of rank tests that could probably be adapted in the context of rank and sign statistics, but their estimator converges to φ_f very slowly which makes its practical usefulness doubtful, as indicated by the authors themselves. We have not yet been able to investigate this issue further. Consequently, we simply use, in practice, a variable-bandwidth (Gaussian) kernel to estimate the density function $f(\cdot)$ of the error term and, hence, the score function $\phi_f(\cdot)$ involved in the definition of $\varphi(\cdot) = \phi_f\left(F^{-1}(\cdot)\right)$.

Appendix A.8. Proof of Proposition 4

Using (18) and given that $\widehat{\mathbf{I}}^{(n)*}_{\widehat{f}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})$ is a consistent estimate of $\mathbf{I}^*_f(\widetilde{\boldsymbol{\theta}}^{(n)})$ and matrix $\mathbf{I}^*_f(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, we may write that, under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\mathrm{R\&S}} - \boldsymbol{\theta}\right) &= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\widehat{\mathbf{I}}^{(n)*}_{\widehat{f}_n}(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \widehat{\boldsymbol{\Delta}}^{(n)*}_{\widehat{f}_n;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) \\
&= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^*_f(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) + o_{\mathrm{P}}(1) \\
&= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^*_f(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{f;\mathrm{appr}}(\widetilde{\boldsymbol{\theta}}^{(n)}) + o_{\mathrm{P}}(1).
\end{aligned}$$

Using (8) and (A.22), we then obtain that, under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)}_{\mathrm{R\&S}} - \boldsymbol{\theta}\right) \\
&= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^*_f(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{ff}(\widetilde{\boldsymbol{\theta}}^{(n)}) + o_{\mathrm{P}}(1) \\
&= \sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) + \left(\mathbf{I}^*_f(\boldsymbol{\theta})\right)^{-1} \left[\boldsymbol{\Delta}^{(n)*}_{ff}(\boldsymbol{\theta}) - \mathbf{I}^*_f(\boldsymbol{\theta})\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right)\right] + o_{\mathrm{P}}(1) \\
&= \left(\mathbf{I}^*_f(\boldsymbol{\theta})\right)^{-1} \boldsymbol{\Delta}^{(n)*}_{ff}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1).
\end{aligned}$$

Since, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$, $\boldsymbol{\Delta}_{ff}^{(n)*}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_f^*(\boldsymbol{\theta})\right)$, we may conclude that

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\mathrm{R\&S}}^{(n)} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{I}_f^*(\boldsymbol{\theta})\right)^{-1}\right).$$

Appendix A.9. R&S estimator with a specific reference density f

This section reports the formulas used to implement the R&S estimator presented in section 4.2 when the reference density is either the normal density (section Appendix A.9.1) or the logistic distribution (Section Appendix A.9.2). We first present all the expressions needed for a generic density, and we will then apply them to the two considered densities. Denote by f_1 the reference density with a median of 0 and a unit variance, and by f the reference density of the same distribution, but with a variance σ^2.
If $W \sim f$, then $Z = \frac{W}{\sigma} \sim f_1$ with $W = \sigma Z$. Further,

$$\begin{aligned}
F(\omega) &= P(W \leqslant \omega) = P\left(Z \leqslant \omega/\sigma\right) = F_1(\omega/\sigma), \\
f(\omega) &= \frac{dF(\omega)}{d\omega} = \frac{d}{d\omega}\left\{F_1(\omega/\sigma)\right\} = \frac{1}{\sigma} f_1\left(\frac{\omega}{\sigma}\right), \\
f'(\omega) &= \frac{df(\omega)}{d\omega} = \frac{d}{d\omega}\left\{\frac{1}{\sigma} f_1\left(\frac{\omega}{\sigma}\right)\right\} = \frac{1}{\sigma^2} f_1'\left(\frac{\omega}{\sigma}\right), \\
\phi_f(\omega) &= -\frac{f'(\omega)}{f(\omega)} = -\frac{\frac{1}{\sigma^2} f_1'(\omega/\sigma)}{\frac{1}{\sigma} f_1(\omega/\sigma)} = \frac{1}{\sigma}\phi_{f_1}(\omega/\sigma), \qquad \text{(A.23)} \\
F^{-1}(u) &= \sigma F_1^{-1}(u), \\
\varphi_f(u) &= \phi_f\left(F^{-1}(u)\right) = \phi_f\left(\sigma F_1^{-1}(u)\right) \\
&= \frac{1}{\sigma}\phi_{f_1}\left(F_1^{-1}(u)\right) = \frac{1}{\sigma}\varphi_{f_1}(u), \\
\varphi_f(u) F^{-1}(u) &= \varphi_{f_1}(u) F_1^{-1}(u).
\end{aligned}$$

The integrals that appear in the expression (15) are shown below:

$$\begin{aligned}
\mu_f &= \int_0^1 F^{-1}(u)du = \int_0^1 \sigma F_1^{-1}(u)du = \sigma\mu_{f_1},\\
\nu_f &= \int_0^1 \left(F^{-1}(u)\right)^2 du = \sigma^2 \int_0^1 \left(F_1^{-1}(u)\right)^2 du = \sigma^2\nu_{f_1},\\
\mathcal{I}_f &= \int_0^1 \varphi_f^2(u)du = \int_0^1 \frac{1}{\sigma^2}\varphi_{f_1}^2(u)du = \frac{\mathcal{I}_{f_1}}{\sigma^2},\\
\mathcal{K}_f &= \int_0^1 \varphi_f^2(u)F^{-1}(u)du = \int_0^1 \frac{1}{\sigma^2}\varphi_{f_1}^2(u)\sigma F_1^{-1}(u)du = \frac{\mathcal{K}_{f_1}}{\sigma},\\
\mathcal{Q}_f &= \int_0^1 \left(\varphi_f(u)F^{-1}(u)\right)^2 du = \int_0^1 \left(\varphi_{f_1}(u)F_1^{-1}(u)\right)^2 du = \mathcal{Q}_{f_1},\\
\nu_{fh} &= \int_0^1 F^{-1}(u)H^{-1}(u)du = \sigma\int_0^1 F_1^{-1}(u)H^{-1}(u)du = \sigma\nu_{f_1h},\\
\mathcal{I}_{fh} &= \int_0^1 \varphi_f(u)\varphi_h(u)du = \int_0^1 \frac{1}{\sigma}\varphi_{f_1}(u)\varphi_h(u)du = \frac{\mathcal{I}_{f_1h}}{\sigma},\\
\mathcal{J}_{fh} &= \int_0^1 \varphi_f(u)H^{-1}(u)du = \frac{\mathcal{J}_{f_1h}}{\sigma},\\
\mathcal{J}_{hf} &= \int_0^1 \varphi_h(u)F^{-1}(u)du = \sigma\mathcal{J}_{hf_1},\\
\mathcal{K}_{fh} &= \int_0^1 \varphi_f(u)\varphi_h(u)H^{-1}(u)du = \frac{\mathcal{K}_{f_1h}}{\sigma},\\
\mathcal{K}_{hf} &= \int_0^1 \varphi_h(u)\varphi_f(u)F^{-1}(u)du = \mathcal{K}_{hf_1},\\
\mathcal{Q}_{fh} &= \mathcal{Q}_{f_1h}.
\end{aligned} \tag{A.24}$$

The standard deviation σ can be estimated by the corrected (for small samples) standard deviation of the residuals $e_i(\widetilde{\theta}^{(n)})$.

Appendix A.9.1. Assuming a Normal reference density

Considering a centered Normal as the reference density in the rank-and-sign semiparametric procedure leads to a variant of the Van der Waerden scores. In the following, we present the explicit expressions to be plugged in (A.23) and (A.24) using the standard normal distribution f_1:

$$\begin{aligned}
f_1(z) &= \frac{1}{\sqrt{2\pi}}e^{-z^2/2},\\
\mu_{f_1} &= 0, \quad \sigma_{f_1}^2 = 1,\\
f_1'(z) &= -\frac{z}{\sqrt{2\pi}}e^{-z^2/2},\\
\phi_{f_1}(z) &= -\frac{f_1'(z)}{f_1(z)} = z,\\
\varphi_{f_1}(u) &= F_1^{-1}(u), \text{ i.e the quantile function of the } \mathcal{N}(0,1) \text{ distribution.}
\end{aligned}$$

Appendix A.9.2. Assuming a Logistic reference density

Considering a centered logistic density as the reference density in the rank-and-sign semiparametric procedure leads to a variant of the Wilcoxon scores. In the following, we present the explicit expressions to be plugged in (A.23) and (A.24) using the standard logistic distribution f_1:

$$
\begin{aligned}
f_1(z) &= \frac{\gamma e^{-\gamma z}}{\left(1+e^{-\gamma z}\right)^2} \qquad \text{with } \gamma = \frac{\pi}{\sqrt{3}}, \\
\mu_{f_1} &= 0, \quad \sigma^2_{f_1} = 1, \\
f_1'(z) &= \frac{-\gamma^2 e^{-\gamma z}\left(1+e^{-\gamma z}\right) + 2\gamma^2 e^{-2\gamma z}}{\left(1+e^{-\gamma z}\right)^3}, \\
\phi_{f_1}(z) &= \frac{\gamma(1-e^{-\gamma z})}{1+e^{-\gamma z}}, \\
F_1^{-1}(u) &= \frac{1}{\gamma}\ln\left(\frac{u}{1-u}\right), \\
\varphi_{f_1}(u) &= \gamma(2u-1).
\end{aligned}
$$