Document de travail

Discussion paper

N°2024-04

https://lem.univ-lille.fr

Semiparametrically Efficient Estimation of Linear Regression Models with Spillovers

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 Efficient Estimation of Linear Regression Models with Spillovers

Nicolas Debarsy[1], Vincenzo Verardi*[2], and Catherine Vermandele[3]

[1]Univ. Lille, CNRS, IESEG School of Management, UMR 9221 - LEM - Lille Économie Management, F-59000 Lille, France
[2]Université Catholique de Louvain. LouRIM. Place des Doyens 1/L2.01.01. B-1348 Louvain-la-Neuve.
[3]Université libre de Bruxelles, LMTD, 50 Av. F. D. Roosevelt, B-1050 Brussels.

Abstract

Linear regression models with spillover effects generally cannot be estimated using ordinary least squares given the simultaneity that results from interactions among individuals. Instead, they are fitted using two-stage least squares (Kelejian & Prucha 1998, Bramoullé et al. 2009), generalized method of moments (Liu et al. 2010), (quasi-) maximum likelihood typically under the normality assumption (Lee 2004) or adaptive estimation (Robinson 2010).

In this article, we propose a semiparametrically efficient estimator, based on the Local Asymptotic Normality theory of Le Cam (1960) and on the work of Hallin et al. (2006, 2008) on residuals ranks-and-signs, that only requires strong unimodality of errors' distribution as a distributional assumption. Monte Carlo simulations show that the suggested estimator performs well in comparison to competing estimators. A trade regression from Behrens et al. (2012) is used to illustrate how empirical findings might greatly change when the Gaussian distribution is not imposed.

Keywords: Spillovers, Efficiency, Local Asymptotic Normality, Semiparametric estimation
JEL: C14, C21, C51

*This research was initiated when Vincenzo Verardi was Associated Researcher of the FNRS (Fonds National de la Recherche Scientifique) at the University of Namur and he gratefully acknowledges their financial support.

1 Introduction

It is well known that models that explicitly account for endogenous spillover effects, such as Spatial Autoregressive (SAR) models in the spatial econometrics literature, cannot be estimated using 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), Maximum Likelihood (ML), and adaptive estimations (ADP) (see Kelejian & Prucha 1998, 1999, Bramoullé et al. 2009, Lee 2004, 2007, Liu et al. 2010, Robinson 2010, Lee & Robinson 2020).

ML estimation yields the most efficient estimator if the distribution of the error term coincides with the assumed one (typically the Gaussian). However, if the error term's distribution is unknown, the ML estimator cannot be computed.

Lee (2004) develops a quasi-ML (QML) estimator 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.[1]

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 performs generally better than the Gaussian QML estimator when the normality assumption is not satisfied.

Robinson (2010) proposes a different methodology, using an adaptive estimator of the parameters of interest, that relies on series estimates of the score function. This method does not need to specify a parametric distribution for the error term and, under a set of assumptions detailed later, leads to an efficient estimator. Similarly, Lee & Robinson (2020) present an adaptive estimator designed specifically for pure spatial models that do not include explanatory variables. This estimator is useful for a wide variety of interaction models, including spatial autoregressive, spatial moving average, and matrix exponential spatial specifications.

In this paper, we propose a *semiparametric* approach, where the innovation density is viewed as an infinite-dimensional nuisance parameter in the regression model. In such a semiparametric context, one can define an estimator for the regression model's vector of parameters that asymptotically approaches the semiparametric efficiency bound, provided that a suitable function of the residuals is used.

More precisely, relying on the concept of Local Asymptotic Normality (LAN) introduced by Le Cam (1960) and on the work of Hallin et al. (2006, 2008), we build a ranks-and-signs-based (R&S) semiparametric estimator for linear models with spillovers which is asymptotically semiparametrically efficient.

We then perform Monte Carlo experiments to evaluate the behavior of the proposed estimator in finite samples and observe that it performs well compared to existing alternatives.

[1]The consistency of QML under Normality comes from the fact that the Gaussian distribution belongs to the linear exponential family (see Gourieroux et al. 1984).

Finally, using a trade regression proposed by Behrens et al. (2012), we illustrate the usefulness of the developed estimator in applied research. In one of their intermediate empirical results, relying on QML, the authors show that the endogenous (spillover) effect is non-statistically significant. Using the R&S estimator, we get a point estimate that is 2.5 times higher and strongly statistically different from 0.

The rest of the paper is organized as follows. Section 2 presents the Local Asymptotic Normality property on which the estimator is built. Section 3 presents the model to be estimated and details how the semiparametric estimation procedure is obtained. This section also derives the fully semiparametrically efficient estimator for the model under study. Section 4 presents the practical implementation of the proposed estimator. Section 5 is dedicated to Monte Carlo experiments which compare the R&S estimator with TSLS, QML, GMM, and ADP. The efficiency of the R&S 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 a substantially higher efficiency compared to the alternative estimators. Section 6 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 by relying on the Gaussian QML of Lee (2004) as well as to GMM and ADP estimators. Finally, section 7 concludes.

2 LAN property

In this section we rely on Hallin (1996) and Van der Vaart (1998) to briefly describe the LAN property.[2] The notations and definitions adopted are those of Hallin (1996) (from page 129 onwards). Let $\mathbf{y}^{(n)} = \left(y_1^{(n)}, \ldots, y_n^{(n)}\right)^{\mathrm{T}}$, $n \in \mathbb{N}_0$, be a sequence of observations described by the sequence of statistical models $\mathcal{E}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, where $\mathcal{P}^{(n)} = \left\{\mathrm{P}_{\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} \in \Theta\right\}$ is a parametric family of probability distributions defined on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and indexed by the parameter vector $\boldsymbol{\theta} \in \Theta$ (with Θ an open set of $\mathbb{R}^K$); observation $\mathbf{y}^{(n)}$ is a random vector, of distribution $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$.

Consider the sequences of probability distributions $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$ and $P_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}}^{(n)}$, where $\boldsymbol{\nu}(n)$ is a $(K \times K)$ non singular matrix such that $\|\boldsymbol{\nu}(n)\| \to 0$ for $n \to \infty$ ($\|\cdot\|$ is the matrix norm induced by the euclidean norm[3]), and $\boldsymbol{\tau}^{(n)}$ is a $(K \times 1)$ real vector such that $\sup_{n \in \mathbb{N}_0} (\boldsymbol{\tau}^{(n)})^{\mathrm{T}} \boldsymbol{\tau}^{(n)} < \infty$. The logarithm of the likelihood ratio is shown in equation (1):

$$\Lambda_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta}}^{(n)} = \ln\left(\frac{d\mathrm{P}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}}^{(n)}}{d\mathrm{P}_{\boldsymbol{\theta}}^{(n)}}\right). \tag{1}$$

[2]The interested reader may also consult Le Cam (1986) and Le Cam & Yang (2000).

[3]$\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = (\max\{\text{eigenvalues of } \mathbf{A}\mathbf{A}^{\mathrm{T}}\})^{1/2}$.

Le Cam (1986) highlighted that a very general structure characterized by the behavior of $\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta}}$ is sufficient (and almost necessary) for the study of the asymptotic performances of almost all statistical inferential procedures for $\boldsymbol{\theta}$. This is called the LAN property.

Definition. (cf. Definition 4.1, page 131 in Hallin (1996)) The sequence of parametric statistical models $\mathcal{E}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$ is said to be *locally asymptotically normal* if, for all $\boldsymbol{\theta} \in \Theta$, there exists a sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ of K-dimensional and $(\mathbf{y}^{(n)}, \boldsymbol{\theta})$-measurable random vectors, and a $(K \times K)$ symmetric positive semi-definite matrix $\mathbf{I}(\boldsymbol{\theta})$, such that, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, as $n \to \infty$:

(i) for every sequence $\boldsymbol{\tau}^{(n)}$ such that $\sup_{n\in\mathbb{N}_0}(\boldsymbol{\tau}^{(n)})^{\mathrm{T}}\boldsymbol{\tau}^{(n)} < \infty$,

$$\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}(n)\boldsymbol{\tau}^{(n)}/\boldsymbol{\theta}} = (\boldsymbol{\tau}^{(n)})^{\mathrm{T}}\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{\tau}^{(n)})^{\mathrm{T}}\mathbf{I}(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}}(1)\,; \tag{2}$$

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

The vector $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta})$ is called the *central sequence*. It is only defined up to $o_{\mathrm{P}}(1)$ (under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, as $n \to \infty$).

The LAN property of the sequence of statistical models $\mathcal{E}^{(n)}$ implies namely that, if $\widetilde{\boldsymbol{\theta}}^{(n)}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}$, then the one-step estimator

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

is an *asymptotically efficient* estimator of $\boldsymbol{\theta}$ (see, for instance, Hallin et al. 2008, p.399). In other words, $\widehat{\boldsymbol{\theta}}^{(n)}$ is asymptotically equivalent to the ML estimator of $\boldsymbol{\theta}$: under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, (\mathbf{I}(\boldsymbol{\theta}))^{-1}\right). \tag{3}$$

According to Le Cam (1970), the conditions for LAN to hold are less restrictive than the conventional differentiability conditions needed for maximum likelihood. He has shown that LAN only requires the density function to be differentiable in quadratic mean. Broadly speaking, quadratic mean differentiability (QMD) requires a density function to be differentiable almost everywhere. The Laplace distribution, for example, is not differentiable at all points but nevertheless exhibits the QMD property. This makes it theoretically unsuitable for ML but appropriate for a one-step estimator based on the LAN property.

3 Efficient estimation of the semiparametric linear model with spillovers

3.1 The model

Consider the following linear model with endogenous spillover effects.[4] For $i = 1, \dots, n$,

$$y_i^{(n)} = (\mathbf{x}_i^{(n)})^{\mathrm{T}} \boldsymbol{\beta} + \lambda \sum_{\substack{j=1 \\ j \neq i}}^{n} w_{ij}^{(n)} y_j^{(n)} + \varepsilon_i^{(n)}, \tag{4}$$

where n is the considered sample size, $\mathbf{x}_i^{(n)} = \left(1, x_{i1}^{(n)}, \dots, x_{iK}^{(n)}\right)^{\mathrm{T}}$ is the vector of explanatory variables for individual i and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)^{\mathrm{T}} \in \mathbb{R}^{K+1}$ is the associated vector of regression parameters, $\sum_{\substack{j=1 \\ j \neq i}}^{n} w_{ij}^{(n)} y_j^{(n)}$ 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. Finally, $\varepsilon_1^{(n)}, \dots, \varepsilon_n^{(n)}$ are i.i.d. error terms with unknown distribution function F and density f assumed, without any loss of generality, to have a median of zero (this is needed to ensure the identification of the intercept β_0 of the model).

The definition and properties of the ranks-and-signs-based estimator proposed here require some regularity conditions detailed below. Assumption 1 relates to the interaction terms $w_{ij}^{(n)}$ and on the endogenous effects parameter λ, while Assumption 2 concerns the covariate vectors $\mathbf{x}_i^{(n)}$; these first two assumptions come from Lee (2004). Assumption 3 specifies regularity conditions for the unknown distribution of the error terms.

Assumption 1.

(i) 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 $n \to \infty$.[5] As a normalization, $w_{ii}^{(n)} = 0$ for all i.

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

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

(ii) Let $\mathbf{I}_n$ be the $(n \times n)$-identity matrix. In model (4), the matrix $\mathbf{I}_n - \lambda \mathbf{W}^{(n)}$ is nonsingular. Moreover, the sequences $\left\{\mathbf{W}^{(n)}\right\}$ and $\left\{\left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)^{-1}\right\}$ are uniformly bounded in both row and column sums (Horn & Johnson 1985).

(iii) 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 (4) 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 $\left(1/\omega_{\text{max}}^{(n)}, 1/\omega_{\text{min}}^{(n)}\right)$, where $\omega_{\text{min}}^{(n)}$ and $\omega_{\text{max}}^{(n)}$ are respectively the minimal and maximal eigenvalues of $\mathbf{W}^{(n)}$. To ensure the same parameter space for λ for different connectivity matrices, it is most of the time normalized. Kelejian & Prucha (2010) proposes two matrix norms, namely the spectral radius and the minimum between the absolute row and column sum norms, which allow to restrict Λ to be the open subset $(-1, 1)$.[6]

Assumption 2. The elements of $\mathbf{x}_i^{(n)}$ are uniformly bounded constants for all n. Besides, the $\lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} (\mathbf{x}_i^{(n)})^{\mathrm{T}}/n$ exists and is non-singular.

Remark. By writing model (4) for the whole sample, we compute its reduced form as:

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

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

$$\mathbf{W}^{(n)} \mathbf{y}^{(n)} = \mathbf{G}^{(n)}(\lambda) \left(\mathbf{X}^{(n)} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^{(n)}\right), \tag{5}$$

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

$$\begin{aligned}
e_i^{(n)}(\boldsymbol{\theta}) &= y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}} \boldsymbol{\beta} - \lambda \mathbf{W}_{i\bullet}^{(n)} \mathbf{y}^{(n)} && (6)\\
&= y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}} \boldsymbol{\beta} - \lambda \mathbf{G}_{i\bullet}^{(n)}(\lambda) \left(\mathbf{X}^{(n)} \boldsymbol{\beta} + \mathbf{e}^{(n)}(\boldsymbol{\theta})\right) \\
&= y_i^{(n)} - (\mathbf{x}_i^{(n)})^{\mathrm{T}} \boldsymbol{\beta} - \lambda \mathbf{G}_{i\bullet}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} - \lambda \sum_{j=1}^{n} G_{ij}^{(n)}(\lambda) e_j^{(n)}(\boldsymbol{\theta}). && (7)
\end{aligned}$$

[6] The row-normalization is also widely used in applied work. However, unless it is theoretically grounded (see, for instance, Patacchini & 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 & Plümper 2016).

Assumption 3. 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:
(i) $F(0) = \int_{-\infty}^{0} f(e)de = 1/2$; (ii) $\mu_f = \int_{-\infty}^{\infty} ef(e)de < \infty$ and $0 < \nu_f = \int_{-\infty}^{\infty} e^2 f(e)de < \infty$; (iii) f is absolutely continuous, strictly positive for all points in $\mathbb{R}$, with (almost everywhere) derivative f' and finite Fisher information for location $\mathcal{I}_f = \int_{\infty}^{\infty} \phi_f^2(e)f(e)de$, where $\phi_f(\cdot) = -\frac{f'(\cdot)}{f(\cdot)}$; (iv) f is strongly unimodal, i.e. function ϕ_f is non-decreasing[7]; (v) $\mathcal{K}_f = \int_{-\infty}^{\infty} \phi_f^2(e)ef(e)de < \infty$ and $0 < \mathcal{Q}_f = \int_{-\infty}^{\infty} \phi_f^2(e)e^2 f(e)de < \infty$.

Let $\mathcal{F}_0 = \{f : \mathbb{R} \to [0, \infty)$ such that f satisfies Assumption 3$\}$. Since the density of errors in model (4) is unknown but assumed to belong to $\mathcal{F}_0$, it plays the role of a nonparametric (infinite dimensional) nuisance. Hence, specification (4) defines a *semiparametric* model

$$\mathcal{E}_0^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)} : f \in \mathcal{F}_0, \boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}} \in \mathbb{R}^{K+1} \times \Lambda\right\}\right).$$

Under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, the residuals $e_i^{(n)}(\boldsymbol{\theta})$ $(i = 1, \ldots, n)$ defined by (6) are i.i.d. with (marginal) density $f \in \mathcal{F}_0$.

It is important to emphasize that the median, not the mean, is used as the location parameter of the innovation density function, and is assumed to be 0.

The rationale behind selecting the zero-median over the traditional zero-mean assumption is motivated by the fact that the former allows us to identify a simple group of transformations of $\mathbb{R}^n$ that "generates" our semiparametric SAR 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 3.4.

Let us finally introduce some additional notations that will be used throughout the text:

- For a probability density function $f \in \mathcal{F}_0$: and for $u \in (0, 1)$, $\varphi_f(u) = \phi_f\left(F^{-1}(u)\right)$.
- For a square $(n \times n)$-matrix $\mathbf{A}^{(n)}$: $\mathbf{A}_{i\cdot}^{(n)}$ is the i^{th} row of $\mathbf{A}^{(n)}$, $\overline{\mathbf{A}}_{\cdot}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_{i\cdot}^{(n)}$ is the average $(1 \times n)$-vector of the n rows of $\mathbf{A}^{(n)}$, $\overline{A}_{\cdot\cdot}^{(n)} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}$ is the average of the n^2 components of matrix $\mathbf{A}^{(n)}$, and $\mathrm{tr}(\mathbf{A}^{(n)})$ is the trace of $\mathbf{A}^{(n)}$.

3.2 Semiparametric efficiency

If we consider that the error term density is known and equal to a specific f, the problem of the estimation of the parameters vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}}$ occurs in the context of the *parametric* submodel

$$\mathcal{E}_{0;f}^{(n)} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)} : \boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}} \in \mathbb{R}^{K+1} \times \Lambda\right\}\right)$$

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

of $\mathcal{E}_0^{(n)}$. In this parametric context, maximum likelihood estimation of $\boldsymbol{\theta}$ is straightforward. The log-likelihood function associated to $\mathcal{E}_{0;f}^{(n)}$ is

$$\ln L\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right) = \ln \left|\det\left(\mathbf{I}_n - \lambda \mathbf{W}^{(n)}\right)\right| + \sum_{i=1}^{n} \ln f\left(e_i^{(n)}(\boldsymbol{\theta})\right). \tag{8}$$

Estimating $\boldsymbol{\theta}$ efficiently in a semiparametric setting is more difficult. Keep in mind that when the error distribution is treated as an unknown nuisance component in the model, there is usually a loss of precision in the estimation of the relevant parameters. One possible intuitive explanation for this efficiency loss is that small changes to both the relevant parameters and the model's nuisance component can have a comparable effect on the distribution of the observations $y_i^{(n)}$ $(i = 1, \ldots, n)$ and hence, cannot be distinguished, even asymptotically. Knowing the efficiency bounds for the estimation of the parameters of interest in semiparametric models is of fundamental importance (see for instance Newey 1990, Bickel et al. 1993) as they offer a benchmark against which to evaluate the asymptotic efficiency of any semiparametric estimator.

As explained below, since any parametric submodel $\mathcal{E}_{0;f}^{(n)}$ is LAN, asymptotically efficient inference for $\boldsymbol{\theta}$ in the semiparametric model $\mathcal{E}_0^{(n)}$ can be conducted on the basis of a ranks-and-signs based central sequence for $\boldsymbol{\theta}$.

3.3 Parametrically efficient estimation of $\boldsymbol{\theta}$ under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$

For every density function $f \in \mathcal{F}_0$, the sequence of parametric submodels

$$\mathcal{E}^{(n)}_{0;f} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}} : \boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \lambda)^{\mathrm{T}} \in \mathbb{R}^{K+1} \times \Lambda\right\}\right)$$

is LAN.[8] Hence, classical likelihood inference for $\boldsymbol{\theta}$ in the parametric submodel $\mathcal{E}^{(n)}_{0;f}$ can be based on the *central sequence*

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

which can be decomposed as

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

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

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

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

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

where

$$\mathbf{I}_{f;\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathcal{I}_f \lim_{n\to\infty} \left\{\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^{(n)} (\mathbf{x}_i^{(n)})^{\mathrm{T}}\right\},$$

[8] Taking $\boldsymbol{\tau}^{(n)} = \boldsymbol{\tau}$ for all n, and $\boldsymbol{\nu}^{(n)} = \frac{1}{\sqrt{n}} \mathbf{I}_K$, where $\mathbf{I}_K$ is the identity matrix of dimension K, we obtain the decomposition (2) of the logarithm of the likelihood ratio, $\Lambda^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n}/\boldsymbol{\theta}} = \ln\left(\frac{L\left(\boldsymbol{\theta}+\boldsymbol{\tau}/\sqrt{n} \middle| \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}{L\left(\boldsymbol{\theta} \middle| \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right)}\right)$, using a second order Taylor expansion of $\ln L\left(\boldsymbol{\theta} + \boldsymbol{\tau}/\sqrt{n} \left| \mathbf{y}^{(n)}, \mathbf{W}^{(n)}, \mathbf{X}^{(n)}\right.\right)$ around $\boldsymbol{\theta}$.

$$\mathbf{I}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) = \mathcal{I}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} \left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} \right) \right\}$$

$$+ \mathcal{K}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{(n)} G_{ii}^{(n)}(\lambda) \right\} + \mathcal{I}_f \mu_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbf{x}_i^{(n)} G_{ij}^{(n)}(\lambda) \right\},$$

and

$$\mathbf{I}_{f;\lambda}(\boldsymbol{\theta}) = \mathcal{I}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} \right)^2 \right\}$$

$$+ (\mathcal{Q}_f - 1) \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(G_{ii}^{(n)}(\lambda) \right)^2 \right\} + \mathcal{I}_f \nu_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(G_{ij}^{(n)}(\lambda) \right)^2 \right\}$$

$$+ 2\,\mathcal{K}_f \mu_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} G_{ii}^{(n)}(\lambda) G_{ij}^{(n)}(\lambda) \right\} + \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} G_{ij}^{(n)}(\lambda) G_{ji}^{(n)}(\lambda) \right\}$$

$$+ \mathcal{I}_f \mu_f^2 \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{k=1 \\ k\neq i,j}}^{n} G_{ij}^{(n)}(\lambda) G_{ik}^{(n)}(\lambda) \right\}$$

$$+ 2\,\mathcal{K}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} \right) G_{ii}^{(n)}(\lambda) \right\}$$

$$+ 2\,\mathcal{I}_f \mu_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(\mathbf{G}_{i\boldsymbol{.}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} \right) G_{ij}^{(n)}(\lambda) \right\},$$

with μ_f, ν_f, $\mathcal{I}_f$, $\mathcal{K}_f$, and $\mathcal{Q}_f$ defined in Assumption 3.[9]

In particular, if $\widetilde{\boldsymbol{\theta}}^{(n)}$ is a $\sqrt{n}$-consistent preliminary estimator of $\boldsymbol{\theta}$, then

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

[9] Appendix A.1 contains some details on the derivation of the different terms.

is an *asymptotically parametrically efficient* estimator of $\boldsymbol{\theta}$: under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$,

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

3.4 Semiparametrically efficient estimation of $\boldsymbol{\theta}$ under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$

As $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$ is in general not properly centered under density $h \neq f$ (hence, it does not exhibit central limit behavior), inference based on this central sequence is not valid when density f used for the score function $\phi_f(\cdot)$ does not coincide with the true error density; the estimator $\widehat{\boldsymbol{\theta}}^{(n)}_f$ is no longer $\sqrt{n}$-consistent.

Typically, as explained in Hallin et al. (2008), semiparametric theory optimally solves this issue by projecting the central sequence $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$ onto the so-called *tangent spaces* that are related to the variations of the error term density f (see Bickel et al. 1993). These projections produce score functions that are semiparametrically efficient, defining a semiparametric central sequence denoted as $\boldsymbol{\Delta}^{(n)*}_f(\boldsymbol{\theta})$. However, Hallin & Werker (2003) show 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 is more readily and intuitively obtained, with the added benefit of not being dependent on any particular distribution, by conditioning $\boldsymbol{\Delta}^{(n)}_f(\boldsymbol{\theta})$ on the *maximal invariant* for this group of transformations. In this paper, we rely on this second approach to determine $\boldsymbol{\Delta}^{(n)*}_f(\boldsymbol{\theta})$.

Consider the fixed-$\boldsymbol{\theta}$ submodel $\mathcal{E}^{(n)}_{0;\boldsymbol{\theta}} = \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \left\{\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}} : f \in \mathcal{F}_0\right\}\right)$ of $\mathcal{E}^{(n)}_0$ as in Hallin et al. (2006). This submodel is characterized by (i) the *residual function* defined by (6) and (ii) a concept of *white noise* with (marginal) density f such that the one defined in Assumption 3: $\mathbf{y}^{(n)}$ has distribution $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$ if and only if $r^{(n)}_{\boldsymbol{\theta}}(\mathbf{y}^{(n)})$ is white noise with (marginal) density f. Denote by $\mathbf{R}^{(n)}(\boldsymbol{\theta}) = \left(R^{(n)}_1(\boldsymbol{\theta}), \ldots, R^{(n)}_n(\boldsymbol{\theta})\right)^{\mathrm{T}}$ and by $\mathbf{s}^{(n)}(\boldsymbol{\theta}) = \left(s^{(n)}_1(\boldsymbol{\theta}), \ldots, s^{(n)}_n(\boldsymbol{\theta})\right)^{\mathrm{T}}$ the vector of ranks and the vector of signs associated with the residuals $e^{(n)}_1(\boldsymbol{\theta}), \ldots, e^{(n)}_n(\boldsymbol{\theta})$. Define $N^{(n)}_+(\boldsymbol{\theta}) = \sharp\left\{i : s^{(n)}_i(\boldsymbol{\theta}) = +1\right\}$ and $N^{(n)}_-(\boldsymbol{\theta}) = \sharp\left\{i : s^{(n)}_i(\boldsymbol{\theta}) = -1\right\}$ as the numbers of positive and negative residuals, respectively. Clearly, under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$ (for any $f \in \mathcal{F}_0$), $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ is uniformly distributed on 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 both binomial random variables $\mathrm{Bin}(n, 1/2)$. 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 monotone increasing transformations $t : \mathbb{R} \to \mathbb{R}$ such that $\lim_{e \to \pm\infty} t(e) = \pm\infty$ and $t(0) = 0$. Defining the transformation $t^{(n)} : \mathbb{R}^n \to \mathbb{R}^n$ by $t^{(n)}(e_1, \ldots, e_n) = \left(t(e_1), \ldots, t(e_n)\right)$ for $t \in T_0$, we have that the *group of order preserving transfor-*

mations (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 $\mathcal{E}_{0;\boldsymbol{\theta}}^{(n)}$. This generating group has for *maximal invariant* the vectors of residuals ranks $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ and signs $\mathbf{s}^{(n)}(\boldsymbol{\theta})$, or, equivalently, the vectors $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{N}^{(n)}(\boldsymbol{\theta})$.

In this context, following the conditioning argument of Hallin & Werker (2003), we get a semiparametric central sequence, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, by taking the expectation of $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})$ conditionally to $\mathbf{R}^{(n)}(\boldsymbol{\theta})$ and $\mathbf{N}^{(n)}(\boldsymbol{\theta})$, as stated in Proposition 1 below.[10]

Proposition 1. Under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) \middle| \mathbf{N}^{(n)}(\boldsymbol{\theta}), \mathbf{R}^{(n)}(\boldsymbol{\theta})\right] &= \boldsymbol{\Delta}_f^{(n)*}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1) \\ &= \widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1),\end{aligned}$$

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

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

and, defining $g^{(n)}(\lambda) = n^2 \overline{G}_{\cdot\cdot}^{(n)}(\lambda) - \mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)$,

$$\begin{aligned}\Delta_{f;\lambda}^{(n)*}(\boldsymbol{\theta}) = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e_i^{(n)}(\boldsymbol{\theta}) \right) \left(\mathbf{G}_{i\cdot}^{(n)}(\lambda) - \overline{\mathbf{G}}_{\cdot}^{(n)}(\lambda) \right) \mathbf{X}^{(n)}\boldsymbol{\beta} \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e_i^{(n)}(\boldsymbol{\theta}) \right) e_i^{(n)}(\boldsymbol{\theta}) \left(G_{ii}^{(n)}(\lambda) - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n} \right) \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \phi_f \left(e_i^{(n)}(\boldsymbol{\theta}) \right) e_j^{(n)}(\boldsymbol{\theta}) \left(G_{ij}^{(n)}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)} \right) \\ &+ 2f(0) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i^{(n)}(\boldsymbol{\theta}) \left(\overline{\mathbf{G}}_{\cdot}^{(n)}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n} \right),\end{aligned} \tag{10}$$

whereas $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta}) = \begin{pmatrix} \widetilde{\boldsymbol{\Delta}}_{f;\boldsymbol{\beta}}^{(n)*}(\boldsymbol{\theta}) \\ \widetilde{\Delta}_{f;\lambda}^{(n)*}(\boldsymbol{\theta}) \end{pmatrix}$ with

$$\widetilde{\boldsymbol{\Delta}}_{f;\boldsymbol{\beta}}^{(n)*}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f \left(\widetilde{R}_i^{(n)}(\boldsymbol{\theta}) \right) \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)} \right) + 2f(0)\overline{\mathbf{x}}^{(n)} \frac{1}{\sqrt{n}} \left(N_+^{(n)}(\boldsymbol{\theta}) - N_-^{(n)}(\boldsymbol{\theta}) \right) \tag{11}$$

[10]Appendix A.2 presents the most important steps to prove Proposition 1.

and

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

where, as defined in Hallin et al. (2006), for $i = 1, \ldots, n$,

$$
\begin{aligned}
\widetilde{R}_i^{(n)}(\boldsymbol{\theta}) = & \mathbb{I}\left[s_i^{(n)}(\boldsymbol{\theta}) = -1 \right] \left\{ \frac{1}{2} \frac{R_i^{(n)}(\boldsymbol{\theta})}{N_-^{(n)}(\boldsymbol{\theta}) + 1} \right\} \\
& + \mathbb{I}\left[s_i^{(n)}(\boldsymbol{\theta}) = +1 \right] \left\{ \frac{1}{2} + \frac{1}{2} \frac{R_i^{(n)}(\boldsymbol{\theta}) - (n - N_+^{(n)}(\boldsymbol{\theta}))}{N_+^{(n)}(\boldsymbol{\theta}) + 1} \right\}.
\end{aligned}
$$

$\boldsymbol{\Delta}_f^{(n)*}(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta})$ are two versions of the semiparametric (under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$) central sequence for $\boldsymbol{\theta}$ in the semiparametric model $\mathcal{E}_0^{(n)}$. From now on, we will focus our attention to the ranks-and-signs version $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta})$, because of its distribution freeness stated in Proposition 2.

Proposition 2. Under $\mathrm{P}_{h;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$, for any $h \in \mathcal{F}_0$,

$$
\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_f^*(\boldsymbol{\theta}) \right), \tag{13}
$$

where $\mathbf{I}_f^*(\boldsymbol{\theta})$ is the information matrix for $\boldsymbol{\theta}$, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, in the semiparametric model $\mathcal{E}_0^{(n)}$, that is, $\mathbf{I}_f^*(\boldsymbol{\theta})^{-1}$ coincides with the semiparametric efficiency bound for the estimation of $\boldsymbol{\theta}$, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$; $\mathbf{I}_f^*(\boldsymbol{\theta})$ is given by

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

where

$$
\mathbf{I}_{f;\boldsymbol{\beta}}^*(\boldsymbol{\theta}) = \mathcal{I}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)} \right) \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)} \right)^{\mathrm{T}} \right\} + (2f(0))^2 \lim_{n\to\infty} \left\{ \overline{\mathbf{x}}^{(n)} \left(\overline{\mathbf{x}}^{(n)} \right)^{\mathrm{T}} \right\},
$$

$$\begin{aligned}
\mathbf{I}^{*}_{f;\boldsymbol{\beta},\lambda}(\boldsymbol{\theta}) = {} & \mathcal{I}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)} \right) \left(\mathbf{G}_{i\boldsymbol{\cdot}}^{(n)}(\lambda) - \overline{\mathbf{G}}_{\boldsymbol{\cdot}}^{(n)}(\lambda) \right) \mathbf{X}^{(n)} \boldsymbol{\beta} \right\} \\
& + \mathcal{K}_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)} \right) \left(G_{ii}^{(n)}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda) \right)}{n} \right) \right\} \\
& + \mathcal{I}_f \mu_f \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)} \right) \left(G_{ij}^{(n)}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)} \right) \right\} \\
& + (2f(0))^2 \lim_{n\to\infty} \left\{ \left(\overline{\mathbf{G}}_{\boldsymbol{\cdot}}^{(n)}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n} \right) \overline{\mathbf{x}}^{(n)} \right\},
\end{aligned}$$

and

$$
\begin{aligned}
\mathbf{I}^*_{f;\lambda}(\boldsymbol{\theta}) = \mathcal{I}_f \lim_{n\to\infty} & \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda) - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda) \right) \mathbf{X}^{(n)} \boldsymbol{\beta} \right]^2 \right\} \\
& + (\mathcal{Q}_f - 1) \lim_{n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(G^{(n)}_{ii}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n} \right)^2 \right\} \\
& + \mathcal{I}_f \nu_f \lim_{n\to\infty} \left\{ \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 \right\} \\
& + 2\,\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} \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) \right\} \\
& + \lim_{n\to\infty} \left\{ \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) \right\} \\
& + \mathcal{I}_f \mu_f^2 \lim_{n\to\infty} \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} \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) \right\} \\
& + 2\,\mathcal{K}_f \lim_{n\to\infty} \left\{ \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) \right\} \\
& + 2\,\mathcal{I}_f \mu_f \lim_{n\to\infty} \left\{ \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) \right\} \\
& + (2f(0))^2 \lim_{n\to\infty} \left[\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda) \mathbf{X}^{(n)} \boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n} \right]^2,
\end{aligned}
$$

with μ_f, ν_f, $\mathcal{I}_f$, $\mathcal{K}_f$, and $\mathcal{Q}_f$ defined in Assumption 3.[11]

Note that, in accordance with the invariance properties of ranks and signs, the limiting distribution (13) depends only on the reference density f, and not on the true density h.

[11] Appendix A.3 presents the main computation steps of $\mathbf{I}^*_f(\boldsymbol{\theta})$.

3.5 Fully semiparametrically efficient estimation of $\boldsymbol{\theta}$ in $\mathcal{E}_0^{(n)}$

It follows from Proposition 2 that the ranks-and-signs-based central sequence $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta})$ allows to construct inference procedures for $\boldsymbol{\theta}$ that are semiparametrically efficient if the density function f used to define the score function coincides with the true underlying error density. For instance, if $\widetilde{\boldsymbol{\theta}}^{(n)}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}$, then the one-step estimator

$$\widehat{\boldsymbol{\theta}}_f^{(n)*} = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}} \left(\mathbf{I}_f^*(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\widetilde{\boldsymbol{\theta}}^{(n)}) \tag{14}$$

is, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, asymptotically normal with zero mean and covariance matrix $\left(\mathbf{I}_f^*(\boldsymbol{\theta})\right)^{-1}$ and, consequently, $\widehat{\boldsymbol{\theta}}_f^{(n)*}$ is an asymptotically efficient estimator of $\boldsymbol{\theta}$, under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, in the semiparametric model $\mathcal{E}_0^{(n)}$.

But our objective is to define a *uniformly* semiparametrically efficient estimator of $\boldsymbol{\theta}$ in $\mathcal{E}_0^{(n)}$, i.e. an estimator that is semiparametrically efficient, regardless of the true error density. For this, we need a fully semiparametric central sequence that is asymptotically equivalent to $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta})$ under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, but that no longer depends on the unknown density f.

A way to obtain a completely semiparametric central sequence $\widetilde{\boldsymbol{\Delta}}^{(n)*}(\boldsymbol{\theta})$ consists in replacing the density function f in $\widetilde{\boldsymbol{\Delta}}_f^{(n)*}(\boldsymbol{\theta})$ by a kernel estimate computed from the residuals $e_i^{(n)}(\widetilde{\boldsymbol{\theta}}^{(n)})$, $i = 1, \ldots, n$. Then, replacing f by its estimate in the central sequence and the information matrix involved in (14), we define the one-step estimator

$$\widehat{\boldsymbol{\theta}}^{(n)*} = \widetilde{\boldsymbol{\theta}}^{(n)} + \frac{1}{\sqrt{n}} \left(\widehat{\mathbf{I}}^{(n)*}(\widetilde{\boldsymbol{\theta}}^{(n)})\right)^{-1} \widetilde{\boldsymbol{\Delta}}^{(n)*}(\widetilde{\boldsymbol{\theta}}^{(n)}); \tag{15}$$

for any $f \in \mathcal{F}_0$, $\widehat{\boldsymbol{\theta}}^{(n)*} = \widehat{\boldsymbol{\theta}}_f^{(n)*} + o_{\mathrm{P}}(1)$ under $\mathrm{P}_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$, i.e., $\widehat{\boldsymbol{\theta}}^{(n)*}$ is a *fully* semiparametrically efficient estimator of $\boldsymbol{\theta}$ in $\mathcal{E}_0^{(n)}$.

As mentioned in the introduction, Robinson (2010) develops two adaptive estimators, based on series approximations of the score function, designed to be efficient even when the distribution of the error term is unknown. Several important differences exist between these adaptive estimators and the approach proposed in this paper.

To start with, the first estimator (A) proposed by Robinson (2010) has been derived within the framework of interaction matrices which satisfy conditions similar to those imposed by Lee (2002). By contrast, the second estimator (B) requires that either the distribution of the error term or the row-normalized interaction matrix be symmetric. The method proposed here merely requires the density function to be strongly unimodal without other assumption on W than those described in Assumption 1. Furthermore, Robinson (2010) assumes that the density function is differentiable everywhere. In our approach, assumption 3 only requires the distribution to be differentiable in quadratic mean (hence differentiable once almost everywhere).

Finally, the series approximation of the score function necessitates choosing the function that serves as the basis for the series, which Robinson (2010) limits to two alternatives. When the true density distribution is not known, both the basis function and its power must be selected. Our method for estimating the density function is based on a data-dependent variable bandwidth kernel density, which does not require a specific parametrization of the scoring function.

4 Practical implementation

Implementation of the R&S approach we propose requires first to select a preliminary estimator. Next, we have to estimate the unknown error density function f relying on the preliminary residuals and compute the central sequence and information matrix to obtain a one-step efficient estimator. Despite the complexity of the formulas presented in section 3, they are explicitly defined and the integrals can be approximated numerically. In this section, we thus focus on the first two points and then present a refinement procedure used in finite samples.

4.1 The preliminary estimator $\widetilde{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}$

The only condition imposed on the preliminary estimator of $\boldsymbol{\theta}$ is to be $\sqrt{n}$-consistent. Therefore, we start with the TSLS estimator of Kelejian & Prucha (1998), Bramoullé et al. (2009).[12] We also correct the preliminary estimated intercept to ensure that residuals have zero-median.

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

We use a variable-bandwidth (Gaussian) kernel to estimate the density function of the error term. 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.

The formula used for the variable-bandwidth used is $bw_i = bw \times \left\{\mathbf{M}_{geom}/\widehat{f}_{prel}\left(e_i\right)\right\}^{0.5}$, where $\mathbf{M}_{geom}$ is the geometric mean of a preliminary fixed bandwidth (bw) density estimate $\widehat{f}_{prel}$ evaluated at each point (see Abramson 1982, Van Kerm 2003). The bandwidth of the preliminary density estimator is chosen according a rule of thumb due to Silverman, namely $bw = 0.9 \min\left(\hat{\sigma}, \frac{IQR}{1.349}\right) n^{-\frac{1}{5}}$

[12]We could also start from a GMM estimator (Lee 2007) or even from the QML estimator of Lee (2004) but the initial gain in precision by relying on GMM or QML instead of TSLS does not have any effect on our one-step estimator. Naturally, if no exogenous regressors are present, TSLS cannot be used since there is no internal instrument available and QML or GMM should be considered as a starting point.

where $\hat{\sigma}$ corresponds to the estimated standard deviation and IQR is the fitted interquartile range.[13]

4.3 Iterative procedure

To improve the one-step estimator $\widehat{\boldsymbol{\theta}}^{(n)*}$ presented in (15), we propose a refinement procedure. The residuals $e_i^{(n)}(\widehat{\boldsymbol{\theta}}^{(n)*})$ $(i = 1, \ldots, n)$ are computed and used to estimate once again the underlying density function f and to evaluate the log-likelihood function in (8) at the parameter value $\widehat{\boldsymbol{\theta}}^{(n)*}$.

Then, we update $\widehat{\boldsymbol{\theta}}^{(n)*}$ by applying (15) in which $\widehat{\boldsymbol{\theta}}^{(n)*}$ acts as the preliminary estimator, and we evaluate (8) for this new estimate of $\boldsymbol{\theta}$. This iterative process stops (usually very fast) when the log-likelihood value stops increasing.

5 Simulations

The experimental design considered is

$$y_i = \beta_0 + \beta_1 x_i + \lambda \mathbf{W}_{i\cdot}\mathbf{y} + \varepsilon_i^{(n)}, \quad i = 1, \ldots, n \tag{16}$$

where the x_i's are generated once (and kept constant over all the simulations) from a standard normal. We also have that $\beta_0 = \beta_1 = 1$, and λ spans values from -0.7 to 0.7, increasing in steps of 0.2, and also includes the value 0. We consider two different connection patterns between observations, both based on random coordinates from two $U(0, 10)$ distributions (also kept constant across the simulations). The first interaction scheme is binary and considers the 10 nearest neighbors constructed from Euclidian distance. The second connectivity scheme is constructed from the inverse distance truncated to the 15 nearest data points. Finally, these 2 matrices have been normalized using the spectral radius norm of Kelejian & Prucha (2010).

Six alternative probability distributions are considered for the error term which could be encountered in practice:

(a) Standard normal distribution;

(b) Student distribution with two degrees of freedom;

(c) Median-centered Lognormal distribution, with $\mu = 0$ and $\sigma^2 = 1$;

(d) Mixed Distribution of a (zero median) shifted Beta(2,2) and a Student distribution with two degrees of freedom;

(e) Standard Laplace distribution;

(f) Bimodal mixture normal $f(e) = \frac{0.5}{\sqrt{2\pi}} \exp\left(-\frac{(e-3)^2}{2}\right) + \frac{0.5}{\sqrt{2\pi}} \exp\left(-\frac{(e+3)^2}{2}\right)$.

[13] In Silverman (1986, p.48), IQR is divided by 1.34 instead of 1.349.

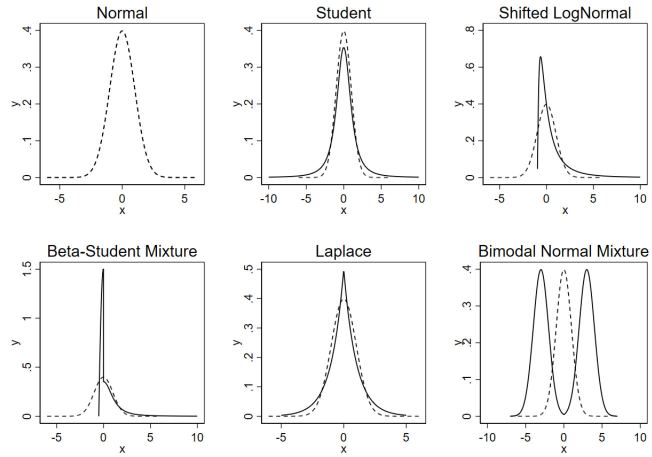

Figure 1: Distributions used in simulations

The last distribution, also used in Robinson (2010), is considered to see how the R&S estimator behaves when the strong unimodality assumption is severely violated. Figure 1 presents the shape of all considered distributions with the normal distribution reproduced in the dashed line to serve as a benchmark.

In total, 108 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 $n = 900$.[14]

For each setup, we assess the performance of $\widehat{\lambda}$ and $\widehat{\beta}_1$ for five alternative estimators: TSLS, QML, efficient GMM, ADP, and the R&S semiparametric estimator proposed here.[15] The summary measures considered to assess the performances of estimators are the median difference of the estimated coefficients to the true values as a measure of their bias and the interquartile range (divided by 1.349 to guarantee Gaussian consistency towards the standard deviation) of the point estimates as a measure of dispersion.

Only simulations related to the largest sample size ($n = 900$) and the inverse distance truncated to the 15 closest data points connectivity matrix are presented in the core of the paper.[16] As far as the constant term is concerned, since it cannot be compared across estimation methods, we do not present the graphs related to the simulations here. However, generally speaking, its bias and

[14] 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.

[15] For the normal distribution setup, we use the ML estimator of Lee (2004).

[16] All remaining simulations are presented in a supplementary file, which also contains, for the R&S estimator, the comparison of the dispersion over repeated samples (computed as the interquartile range normalized by 1.349) of the point estimates of the parameters with the median of the fitted standard errors, to evaluate the bias of the standard errors.

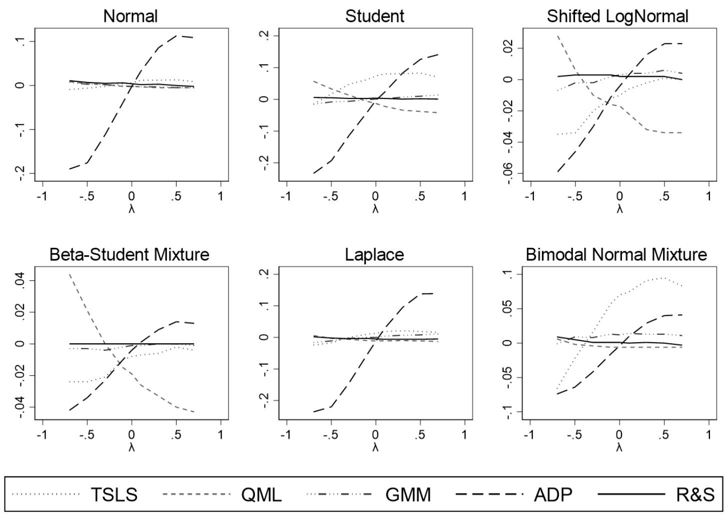

Figure 2: Bias of $\widehat{\lambda}$, $n = 900$ and truncated inverse distance matrix

dispersion over repeated samples are small for the R&S (except for the case of the normal bimodal distribution which is not surprising as the basic assumptions of the procedure are not met).

The TSLS estimator is computed using the 2 first-order neighborhood's characteristics as instruments for the endogenous effects (Kelejian & 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. Computing the ADP estimator of Robinson (2010) requires selecting the series function and power used to approximate the scoring function. Relying on the Monte Carlo estimation results presented in Robinson (2010) we have decided to use expression (2.29) and power $L = 4$ for the series function, specifically a polynomial of fourth degree. We also have decided to focus on estimator A rather than B as, according to Robinson (2010, pp. 13–14), the latter does not show a clear superiority with respect to the former.

Finally, the TSLS approach presented above serves as preliminary estimator for both the ADP and the R&S estimators.

5.1 Bias of the coefficients estimators

The bias of the R&S estimator of λ is negligible for all setups and is less sensitive to the values of λ than the other estimators (see Figure 2). The bias in the estimation of β_1 turns out to be minimal for all the estimators and all the setups, whatever the value of λ (see Figure 3).

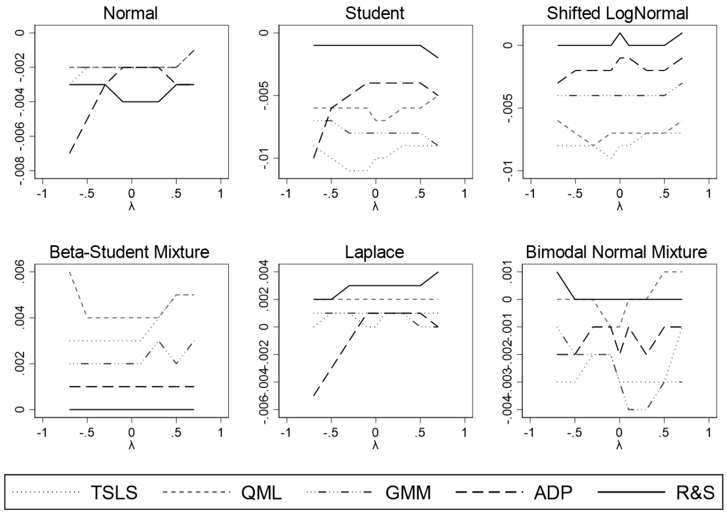

Figure 3: Bias of $\widehat{\beta}_1$, $n = 900$ and truncated inverse distance matrix

5.2 Relative dispersion

Figures 4 and 5 compare the dispersion over repeated samples of the point estimates of the parameters λ and β_1, using the five estimation method presented above.

The R&S estimator of λ is the one with the lowest dispersion in all cases except, as anticipated, when the errors are normal. This result comes from the fact that the dispersion of a semiparametrically efficient estimator will generally be higher than for a parametric efficient one (under the true distribution). Nevertheless, we do not observe a large difference between ML (and GMM) and the R&S estimator. The ADP performs worse, but this is likely due to the choice of a fourth-order polynomial (while the normal score is a linear polynomial).

The dispersion of the R&S estimator of λ is much less sensitive to the true value of this parameter than the other estimators. We also note that the estimator with the second lowest dispersion varies depending on the error distribution. For instance, the ADP estimator is the second most efficient for the case of mixtures distributions while it is dominated by QML and GMM when the error term is distributed according to the Laplace distribution.

Finally, the R&S estimator is the most efficient for the bimodal mixture of normal distributions, even if this distribution does not satisfy the strong unimodality assumption. This finding is interesting as it indicates some robustness of the proposed procedure with respect to the violation of its core assumption.

The ADP estimator for this distribution exhibits good performance, even if its dispersion is on average (over the values of λ) 61% larger than for the proposed R&S estimator.

For parameter β_1, the dispersion is lowest for R&S in all cases with non-normal errors.

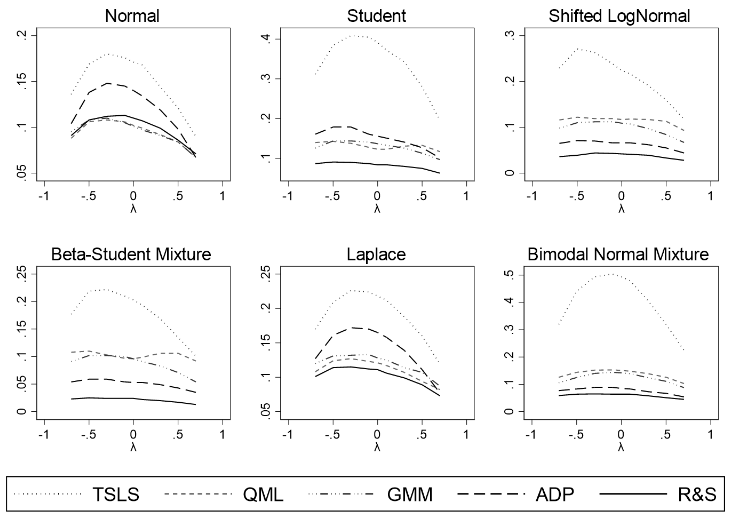

Figure 4: Dispersion of $\widehat{\lambda}$, $n = 900$ and truncated inverse distance matrix

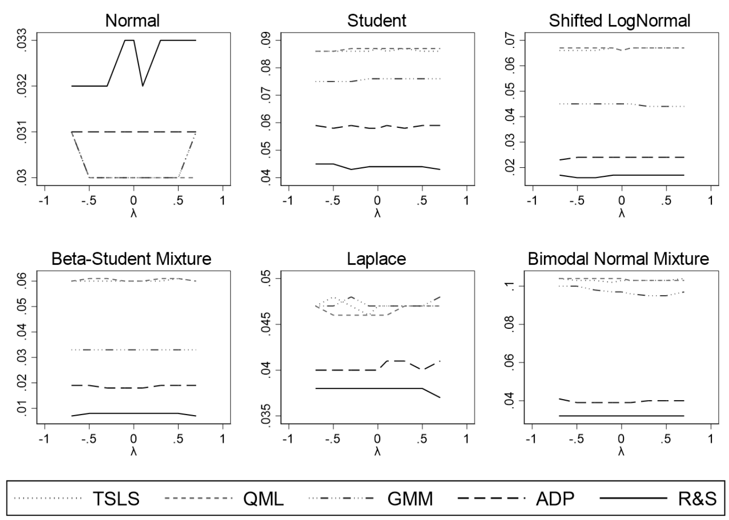

Figure 5: Dispersion of $\widehat{\beta}_1$, $n = 900$ and truncated inverse distance matrix

6 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 econometrics techniques to assess the effect of the Canada-U.S. border on trade flows. Their sample includes 30 US states and 10 Canadian regions, which leads 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 (17):

$$\ln(Z_{ij}) = \beta_0 + \beta_1 d_{ij} + \beta_2 \ln(w_i) + \beta_3 b_{ij} + \lambda \sum_{\substack{k=1 \\ k \neq i}}^{n} \frac{L_k}{L} \ln(Z_{kj}) + \varepsilon_{ij}, \tag{17}$$

where Z_{ij} is the GDP-standardized manufacturing exports from region i to j, d_{ij} is the great circle distance (in kilometers) between regional and provincial capitals. The internal distance is measured as one-fourth the distance of a region's capital from the nearest capital of another region (see Anderson & van Winkoop 2003).[17] 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 belongs to Canada and j is part of the U.S. or vice-versa and 0 otherwise. [18] 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), 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 comparison with the GMM estimator of Liu et al. (2010) and the A adaptive estimator of Robinson (2010), which are potentially more efficient than QML.

The first column in Table 1 presents the estimation results reported in BEK, based on QML under normality. We observe that the interaction effect λ is not statistically different from 0. In Figure 6, we show the qqplot 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. Columns 2 and 3 of Table 1 present estimation results obtained by efficient GMM and by ADP. GMM is slightly more precise than QML, even though qualitatively similar. We also note that λ becomes significant for ADP. The last column of Table 1 reports the results of the R&S estimator.

[17] 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.

[18] To account for zero flow observations in their logarithmic bilateral export model, BEK add value one 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.

Table 1: Comparison of estimation results

Dependent variable: $\ln(Z_{ij})$

	QML	GMM	ADP	R&S
Constant	-13.890	-12.274		-12.230
	(0.713)	(0.691)	-	(0.473)
	[-19.496]	[-17.767]		[-25.871]
d_{ij}	-1.223	-1.280	-1.209	-1.208
	(0.034)	(0.033)	(0.030)	(0.024)
	[-35.984]	[-39.278]	[-40.508]	[-50.780]
$\ln(w_i)$	-1.173	-1.759	-1.203	-1.263
	(0.180)	(0.170)	(0.155)	(0.124)
	[-6.631]	[-10.370]	[-7.759]	[-10.210]
b_{ij}	-1.052	-0.804	-1.074	-1.197
	(0.066)	(0.063)	(0.058)	(0.046)
	[-15.961]	[-12.726]	[-18.647]	[-25.930]
λ	0.030	0.045	0.051	0.108
	(0.030)	(0.029)	(0.023)	(0.019)
	[1.012]	[1.577]	[2.164]	[5.697]
Rel. eff	1	1.084	1.358	2.107

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

The R&S point estimate of the spillover effect is significantly greater compared to the original paper (0.108 vs 0.03), while its standard error decreases substantially (0.019 vs 0.03). This suggests that the absence of significant spillover effect in the original results might come from the QML estimation method inefficiency. We also note that the standard errors of all the other regression coefficients are much smaller as well when using the R&S estimator.

Finally, the bottom panel of Table 1 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):

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

with $\mathbf{I}(\boldsymbol{\theta})^{-1}_{q}$ being the asymptotic covariance matrix of the q^{th} estimator and K_{comp} the number of estimates to be compared. We do not include the constant here as it is viewed as a nuisance parameter by Robinson (2010).

In this empirical application, GMM is slightly more efficient than QML (around 8%), while ADP improves the efficiency by around 36%. The R&S estimator is approximately two times more efficient than QML.

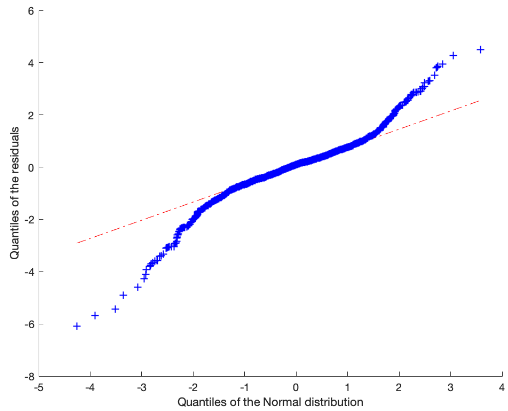

Figure 6: QQ plot of the structural residuals

7 Conclusions

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

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, albeit not efficient.

In this paper, we develop a new estimator based on the concept of Local Asymptotic Normality and previous research by Hallin & 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, is asymptotically *semiparametrically* efficient. Monte Carlo experiments show that it performs generally better than the other methods considered, once the assumption of a normal error distribution is relaxed.

When applied to the trade regression model developed by Behrens et al. (2012), this new approach produces more accurate point estimates than those obtained in the original paper, and provides a statistically significant spillover effect that was not identified in the original paper, based on quasi-maximum likelihood.

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

Acknowledgements

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, University of 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. We also would like to thank Arthur Hemdat for having implemented the Matlab code in R. All remaining errors are our responsibility.

References

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

Anderson, J. & van Winkoop, E. (2003), 'Gravity with gravitas: a solution to the border puzzle', *American Economic Review* **93**, 170–192.

Behrens, K., Ertur, C. & 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. & Wellner, J. A. (1993), *Efficient and Adaptive Statistical Inference for Semiparametric Models*, John Hopkins University Press, Baltimore.

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

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

Hallin, M. (1996), Eléments de la théorie asymptotique des expériences statistiques, *in* J.-J. Droesbeke & J. Fine, eds, 'Inférence non paramétrique fondée sur les Rangs', Editions del'Université' de Bruxelles et Ellipses, pp. 129–166.

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

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

Kelejian, H. H. & 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. & 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. & 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. (1970), 'On the assumptions used to prove asymptotic normality of maximum likelihood estimates', *Annals of Mathematical Statistics* **41**, 802–828.

Le Cam, L. (1986), *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag.

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

Lee, J. & Robinson, P. M. (2020), 'Adaptive inference on pure spatial models', *Journal of Econometrics* **216**, 375–393.

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. (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. & Bollinger, C. R. (2010), 'An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* **159**, 302–319.

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

Newey, W. K. (1990), 'Semiparametric efficiency bounds', *Journal of Applied Econometrics* **5**(2), 99–135.

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

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.

Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Van der Vaart, A. W. (1998), *Asymptotic Statistics, first edition*, Cambridge University Press.

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

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

A Proofs

A.1 Parametric Fisher information matrix for $\boldsymbol{\theta}$ under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$

As shown in Section 3.3, the central sequence for $\boldsymbol{\theta}$ in the parametric submodel $\mathcal{E}^{(n)}_{0;f}$ is

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

with

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

and

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

Since, in view of (5),

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

we have that

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

with

$$L^{(n)}_{1;f}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e^{(n)}_i(\boldsymbol{\theta}) \right) \mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta},$$

$$L^{(n)}_{2;f}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e^{(n)}_i(\boldsymbol{\theta}) \right) e^{(n)}_i(\boldsymbol{\theta}) G^{(n)}_{ii}(\lambda),$$

$$L^{(n)}_{3;f}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \phi_f \left(e^{(n)}_i(\boldsymbol{\theta}) \right) e^{(n)}_j(\boldsymbol{\theta}) G^{(n)}_{ij}(\lambda),$$

$$L^{(n)}_{4;f}(\boldsymbol{\theta}) = -\frac{1}{\sqrt{n}} \mathrm{tr} \left(\mathbf{G}^{(n)}(\lambda) \right).$$

Under $P^{(n)}_{f;\boldsymbol{\theta}}$, the error terms $e^{(n)}_1(\boldsymbol{\theta}), \dots, e^{(n)}_n(\boldsymbol{\theta})$ are i.i.d. with density function f and we have, for all $i = 1, \dots, n$:

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

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

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

and

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

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

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

where

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

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

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

where

$$
\begin{aligned}
\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) \\
&\quad + \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{\cdot}}^{(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{\cdot}}^{(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,
\end{aligned}
$$

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.

A.2 Semiparametric central sequence for $\boldsymbol{\theta}$ under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$

Throughout this section, 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 $e_i^{(n)}(\boldsymbol{\theta})$, $s_i^{(n)}(\boldsymbol{\theta})$, $R_i^{(n)}(\boldsymbol{\theta})$, $\mathbf{N}^{(n)}(\boldsymbol{\theta})$, $\mathbf{R}^{(n)}(\boldsymbol{\theta})$, $\mathbf{G}^{(n)}(\lambda)$, ...

A.2.1 Component associated with $\boldsymbol{\beta}$

Let us first consider

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

By Proposition 3.2 of Hallin et al. (2006), we have that, under $P^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{f;\boldsymbol{\beta}}\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right] - \mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{f;\boldsymbol{\beta}}\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]\right] \\
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(F(e_i^{(n)})\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) \\
+ \sqrt{n}\,\overline{\mathbf{x}}^{(n)}\left\{2\frac{N_-^{(n)}}{n}\mu^-_{\varphi_f} + 2\frac{N_+^{(n)}}{n}\mu^+_{\varphi_f} - \mu_{\varphi_f}\right\} + o_{\mathrm{P}}(1),
\end{aligned}$$

with

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{f;\boldsymbol{\beta}}\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]\right] &= \mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{f;\boldsymbol{\beta}}\right] = \mathbf{0}, \\
\mu_{\varphi_f} &= \int_0^1 \varphi_f(u)du = \int_0^1 \phi_f\left(F^{-1}(u)\right)du = \int_{-\infty}^{\infty} \phi_f(e)f(e)de = 0, \\
\mu^-_{\varphi_f} &= \int_0^{1/2} \varphi_f(u)du = \int_{-\infty}^{0} \phi_f(e)f(e)de = -\int_{-\infty}^{0} f'(e)de \\
&= -\left[f(e)\right]_{-\infty}^{0} = -f(0), \\
\mu^+_{\varphi_f} &= \int_{1/2}^{1} \varphi_f(u)du = \int_0^{\infty} \phi_f(e)f(e)de = -\int_0^{\infty} f'(e)de \\
&= -\left[f(e)\right]_0^{\infty} = f(0).
\end{aligned}$$

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

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

Moreover, Lemma 3.1 — more precisely, relation (3.7) — of Hallin et al. (2006) implies that, under $P^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$,

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f\left(e_i^{(n)}\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(F(e_i^{(n)})\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(\widetilde{R}_i^{(n)}\right)\left(\mathbf{x}_i^{(n)} - \overline{\mathbf{x}}^{(n)}\right) + o_{\mathrm{P}}(1),
\end{aligned}
$$

where

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

Hence,

$$
\mathbb{E}\left[\boldsymbol{\Delta}^{(n)}_{f;\boldsymbol{\beta}} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] = \widetilde{\boldsymbol{\Delta}}^{(n)*}_{f;\boldsymbol{\beta}} + o_{\mathrm{P}}(1),
$$

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

A.2.2 Component associated with λ

Consider now

$$
\begin{aligned}
\Delta^{(n)}_{f;\lambda} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f\left(e_i^{(n)}\right)\mathbf{W}^{(n)}_{i\boldsymbol{.}}\mathbf{y}^{(n)} - \frac{1}{\sqrt{n}}\mathrm{tr}\left(\mathbf{G}^{(n)}\right) \\
&= L^{(n)}_{1;f} + L^{(n)}_{2;f} + L^{(n)}_{3;f} - \frac{1}{\sqrt{n}}\mathrm{tr}\left(\mathbf{G}^{(n)}\right),
\end{aligned}
$$

with

$$L_{1;f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e_i^{(n)} \right) \mathbf{G}_{i\cdot}^{(n)} \mathbf{X}^{(n)} \boldsymbol{\beta},$$

$$L_{2;f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e_i^{(n)} \right) e_i^{(n)} G_{ii}^{(n)},$$

$$L_{3;f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \phi_f \left(e_i^{(n)} \right) e_j^{(n)} G_{ij}^{(n)}.$$

(i) Following similar developments of Section A.2.1, we get that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\begin{aligned} &\mathbb{E} \left[L_{1;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)} \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_f \left(e_i^{(n)} \right) \left(\mathbf{G}_{i\cdot}^{(n)} - \overline{\mathbf{G}}_{\cdot}^{(n)} \right) \mathbf{X}^{(n)} \boldsymbol{\beta} + 2f(0) \overline{\mathbf{G}}_{\cdot}^{(n)} \mathbf{X}^{(n)} \boldsymbol{\beta} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i^{(n)} + o_{\mathrm{P}}(1) \qquad (18) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f \left(\widetilde{R}_i^{(n)} \right) \left(\mathbf{G}_{i\cdot}^{(n)} - \overline{\mathbf{G}}_{\cdot}^{(n)} \right) \mathbf{X}^{(n)} \boldsymbol{\beta} + 2f(0) \overline{\mathbf{G}}_{\cdot}^{(n)} \mathbf{X}^{(n)} \boldsymbol{\beta} \frac{1}{\sqrt{n}} \left(N_{+}^{(n)} - N_{-}^{(n)} \right) + o_{\mathrm{P}}(1), \qquad (19) \end{aligned}$$

where $\overline{\mathbf{G}}_{\cdot}^{(n)} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{G}_{i\cdot}^{(n)}$.

(ii) Applying Proposition 3.2 of Hallin et al. (2006) once again, we have that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$\begin{aligned} &\mathbb{E} \left[L_{2;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)} \right] - \mathbb{E} \left[\mathbb{E} \left[L_{2;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)} \right] \right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f \left(F(e_i^{(n)}) \right) F^{-1} \left(F(e_i^{(n)}) \right) \left(G_{ii}^{(n)} - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)} \right)}{n} \right) \\ &\qquad + \frac{\mathrm{tr}\left(\mathbf{G}^{(n)} \right)}{\sqrt{n}} \left\{ 2 \frac{N_{-}^{(n)}}{n} \mu_{\psi_f}^{-} + 2 \frac{N_{+}^{(n)}}{n} \mu_{\psi_f}^{+} - \mu_{\psi_f} \right\} + o_{\mathrm{P}}(1), \end{aligned}$$

with

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left[L_{2;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right]\right] &= \mathbb{E}\left[L_{2;f}^{(n)}\right] = \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{\sqrt{n}}, \\
\psi_f(u) &= \varphi_f(u) F^{-1}(u), \\
\mu_{\psi_f} &= \int_0^1 \psi_f(u) du = \int_0^1 \varphi_f\,(u)\, F^{-1}(u) du = \int_0^1 \phi_f\left(F^{-1}(u)\right) F^{-1}(u) du \\
&= \int_{-\infty}^{\infty} \phi_f(e) e f(e) de = 1, \\
\mu_{\psi_f}^{-} &= \int_0^{1/2} \psi_f(u) du = \int_{-\infty}^{0} \phi_f(e) e f(e) de = -\int_{-\infty}^{0} f'(e) e de \\
&= -\left[f(e)e\right]_{-\infty}^{0} + \int_{-\infty}^{0} f(e) de = 0 + \frac{1}{2} = \frac{1}{2}, \\
\mu_{\psi_f}^{+} &= \int_{1/2}^{1} \psi_f(u) du = \int_{0}^{\infty} \phi_f(e) e f(e) de = -\int_{0}^{\infty} f'(e) e de \\
&= -\left[f(e)e\right]_{0}^{\infty} + \int_{0}^{\infty} f(e) de = 0 + \frac{1}{2} = \frac{1}{2}.
\end{aligned}
$$

Hence, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

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

Since, under $P_{f;\boldsymbol{\theta}}^{(n)}$, $N_{-}^{(n)} + N_{+}^{(n)} \stackrel{\text{a.s.}}{=} n$, we have that

$$
\begin{aligned}
&\mathbb{E}\left[L_{2;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] \\
&= \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(F(e_i^{(n)})\right) F^{-1}\left(F(e_i^{(n)})\right)\left(G_{ii}^{(n)} - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right) + o_{\mathrm{P}}(1).
\end{aligned} \tag{20}
$$

Applying again result (3.7) of Hallin et al. (2006), we write that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$
\begin{aligned}
&\mathbb{E}\left[L_{2;f}^{(n)} \middle| \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] \\
&= \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(\widetilde{R}_i^{(n)}\right) F^{-1}\left(\widetilde{R}_i^{(n)}\right)\left(G_{ii}^{(n)} - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}\right)}{n}\right) + o_{\mathrm{P}}(1).
\end{aligned} \tag{21}
$$

(iii) Let us now consider the third term of $\Delta^{(n)}_{f;\lambda}$:

$$\begin{aligned} L^{(n)}_{3;f} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \phi_f \left(e^{(n)}_i\right) e^{(n)}_j G^{(n)}_{ij} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varphi_f \left(F(e^{(n)}_i)\right) F^{-1} \left(F(e^{(n)}_j)\right) G^{(n)}_{ij}. \end{aligned}$$

Define the linear "serial" sign-and-rank statistics of order 2, based on the so-called *exact* and *approximate* serial score functions,

$$\begin{aligned} S^{(n)}_{\text{exact}} &= \mathbb{E}\left[L^{(n)}_{3;f} \,\middle|\, \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left[\phi_f \left(e^{(n)}_i\right) e^{(n)}_j \,\middle|\, \mathbf{N}^{(n)}, \mathbf{R}^{(n)}\right] G^{(n)}_{ij} \end{aligned}$$

and

$$S^{(n)}_{\text{appr}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varphi_f \left(\widetilde{R}^{(n)}_i\right) F^{-1} \left(\widetilde{R}^{(n)}_j\right) G^{(n)}_{ij}.$$

Under a straightforward generalisation of Lemma 4.1 of Hallin et al. (2006), we have that, under $P^{(n)}_{f;\boldsymbol{\theta}}$, as $n \to \infty$,

$$S^{(n)}_{\text{exact}} = L^{(n)}_{3;f} - \mathbb{E}\left[L^{(n)}_{3;f} \,\middle|\, \mathbf{e}^{(n)}_{(\cdot)}\right] + \mathbb{E}\left[S^{(n)}_{\text{exact}} \,\middle|\, \mathbf{N}^{(n)}\right] + o_{\mathrm{P}}(1) \tag{22}$$

$$= S^{(n)}_{\text{appr}} - \mathbb{E}\left[S^{(n)}_{\text{appr}} \,\middle|\, \mathbf{N}^{(n)}\right] + \mathbb{E}\left[S^{(n)}_{\text{exact}} \,\middle|\, \mathbf{N}^{(n)}\right] + o_{\mathrm{P}}(1), \tag{23}$$

where $\mathbf{e}^{(n)}_{(\cdot)} = \left(e^{(n)}_{(1)}, \ldots, e^{(n)}_{(n)}\right)^{\mathrm{T}}$ is the vector of order statistics associated with $\mathbf{e}^{(n)}$.

Note first that

$$\begin{aligned} \mathbb{E}\left[L^{(n)}_{3;f} \,\middle|\, \mathbf{e}^{(n)}_{(\cdot)}\right] &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left[\phi_f \left(e^{(n)}_i\right) e^{(n)}_j \,\middle|\, \mathbf{e}^{(n)}_{(\cdot)}\right] G^{(n)}_{ij} \\ &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} G^{(n)}_{ij}\right) \left(\frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{\ell=1 \\ \ell\neq k}}^{n} \phi_f \left(e^{(n)}_{(k)}\right) e^{(n)}_{(\ell)}\right) \\ &= \left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} G^{(n)}_{ij}\right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \phi_f \left(e^{(n)}_i\right) e^{(n)}_j\right). \end{aligned}$$

Since

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G_{ij}^{(n)} = \sum_{i=1}^{n}\sum_{j=1}^{n} G_{ij}^{(n)} - \sum_{i=1}^{n} G_{ii}^{(n)} = n^2\overline{G}_{\bullet\bullet} - \operatorname{tr}\left(\mathbf{G}^{(n)}\right) \stackrel{\text{def}}{=} g^{(n)},$$

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

$$L_{3;f}^{(n)} - \mathbb{E}\left[L_{3;f}^{(n)}\middle|\mathbf{e}_{(\cdot)}^{(n)}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} \phi_f\left(e_i^{(n)}\right)e_j^{(n)}\left(G_{ij}^{(n)} - \frac{g^{(n)}}{n(n-1)}\right). \tag{24}$$

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

$$\begin{aligned}
\mathbb{E}\left[S_{\text{exact}}^{(n)}\middle|\mathbf{N}^{(n)}\right] &= \mathbb{E}\left[\mathbb{E}\left[L_{3;f}^{(n)}\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]\middle|\mathbf{N}^{(n)}\right] = \mathbb{E}\left[L_{3;f}^{(n)}\middle|\mathbf{N}^{(n)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[L_{3;f}^{(n)}\middle|\mathbf{s}^{(n)}\right]\middle|\mathbf{N}^{(n)}\right] \\
&= \mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}\left[\phi_f\left(e_i^{(n)}\right)e_j^{(n)}\middle|s_i^{(n)},s_j^{(n)}\right]G_{ij}^{(n)}\middle|\mathbf{N}^{(n)}\right],
\end{aligned}$$

with

$$\begin{aligned}
\mathbb{E}\left[\phi_f\left(e_i^{(n)}\right)e_j^{(n)}\middle|s_i^{(n)},s_j^{(n)}\right] &= \mathbb{I}\left[s_i^{(n)}=-1, s_j^{(n)}=-1\right]\int_{-\infty}^{0}\phi_f(e)2f(e)de\int_{-\infty}^{0}e2f(e)de \\
&\quad + \mathbb{I}\left[s_i^{(n)}=-1, s_j^{(n)}=+1\right]\int_{-\infty}^{0}\phi_f(e)2f(e)de\int_{0}^{\infty}e2f(e)de \\
&\quad + \mathbb{I}\left[s_i^{(n)}=+1, s_j^{(n)}=-1\right]\int_{0}^{\infty}\phi_f(e)2f(e)de\int_{-\infty}^{0}e2f(e)de \\
&\quad + \mathbb{I}\left[s_i^{(n)}=+1, s_j^{(n)}=+1\right]\int_{0}^{\infty}\phi_f(e)2f(e)de\int_{0}^{\infty}e2f(e)de.
\end{aligned}$$

Since $\int_{-\infty}^{0}\phi_f(e)2f(e)de = -2f(0)$ and $\int_{0}^{\infty}\phi_f(e)2f(e)de = 2f(0)$, we have that

$$\begin{aligned}
&\mathbb{E}\left[\phi_f\left(e_i^{(n)}\right)e_j^{(n)}\middle|s_i^{(n)},s_j^{(n)}\right] \\
&= 4f(0)s_i^{(n)}\left\{\mathbb{I}\left[s_j^{(n)}=-1\right]\int_{-\infty}^{0}ef(e)de + \mathbb{I}\left[s_j^{(n)}=+1\right]\int_{0}^{\infty}ef(e)de\right\}
\end{aligned}$$

and, consequently,

$$\begin{aligned}
\mathbb{E}\left[S^{(n)}_{\text{exact}}\,\middle|\,\mathbf{N}^{(n)}\right] = 4f(0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} G^{(n)}_{ij}\Bigg(&-\mathbb{P}\left[s^{(n)}_i=-1, s^{(n)}_j=-1\,\middle|\,\mathbf{N}^{(n)}\right]\int_{-\infty}^{0} ef(e)de\\
&+\mathbb{P}\left[s^{(n)}_i=+1, s^{(n)}_j=-1\,\middle|\,\mathbf{N}^{(n)}\right]\int_{-\infty}^{0} ef(e)de\\
&-\mathbb{P}\left[s^{(n)}_i=-1, s^{(n)}_j=+1\,\middle|\,\mathbf{N}^{(n)}\right]\int_{0}^{\infty} ef(e)de\\
&+\mathbb{P}\left[s^{(n)}_i=+1, s^{(n)}_j=+1\,\middle|\,\mathbf{N}^{(n)}\right]\int_{0}^{\infty} ef(e)de\Bigg)\\
= 4f(0)\frac{g^{(n)}}{\sqrt{n}}\Bigg(&-\frac{N^{(n)}_-(N^{(n)}_- -1)}{n(n-1)}\int_{-\infty}^{0} ef(e)de+\frac{N^{(n)}_+N^{(n)}_-}{n(n-1)}\int_{-\infty}^{0} ef(e)de\\
&-\frac{N^{(n)}_-N^{(n)}_+}{n(n-1)}\int_{0}^{\infty} ef(e)de+\frac{N^{(n)}_+(N^{(n)}_+ -1)}{n(n-1)}\int_{0}^{\infty} ef(e)de\Bigg)\\
= 4f(0)\frac{g^{(n)}}{n}\sqrt{n}\Bigg(&\frac{N^{(n)}_+N^{(n)}_- - N^{(n)}_-(N^{(n)}_- -1)}{n(n-1)}\int_{-\infty}^{0} ef(e)de\\
&+\frac{N^{(n)}_+(N^{(n)}_+ -1) - N^{(n)}_-N^{(n)}_+}{n(n-1)}\int_{0}^{\infty} ef(e)de\Bigg).
\end{aligned}$$

But

$$\begin{aligned}
\sqrt{n}\frac{N^{(n)}_+N^{(n)}_- - N^{(n)}_-(N^{(n)}_- -1)}{n(n-1)} &= \frac{N^{(n)}_-\left(N^{(n)}_+ - N^{(n)}_- + 1\right)}{\sqrt{n}(n-1)}\\
&= \frac{N^{(n)}_-}{n-1}\,\frac{N^{(n)}_+ - N^{(n)}_-}{\sqrt{n}}+\frac{N^{(n)}_-}{\sqrt{n}(n-1)}\\
&= \frac{1}{2}\,\frac{N^{(n)}_+ - N^{(n)}_-}{\sqrt{n}}+o_{\mathrm{P}}(1)
\end{aligned}$$

and, similarly,

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

Hence,

$$
\begin{aligned}
\mathbb{E}\left[S_{\text{exact}}^{(n)}\,\middle|\mathbf{N}^{(n)}\right] &= 2f(0)\frac{g^{(n)}}{n}\,\frac{N_+^{(n)}-N_-^{(n)}}{\sqrt{n}}\int_{-\infty}^{\infty} ef(e)de + o_{\mathrm{P}}(1) \\
&= 2f(0)\mu_f\frac{g^{(n)}}{n}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i^{(n)} + o_{\mathrm{P}}(1). \qquad (25)
\end{aligned}
$$

Combining (24) and (25), equation (22) gives that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n \to \infty$,

$$
\begin{aligned}
\mathbb{E}\left[L_{3;f}^{(n)}\,\middle|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right] &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\phi_f\left(e_i^{(n)}\right)e_j^{(n)}\left(G_{ij}^{(n)}-\frac{g^{(n)}}{n(n-1)}\right) \\
&\quad + 2f(0)\mu_f\frac{g^{(n)}}{n}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_i^{(n)} + o_{\mathrm{P}}(1). \qquad (26)
\end{aligned}
$$

Finally, consider $\mathbb{E}\left[S_{\text{appr}}^{(n)}\,\middle|\mathbf{N}^{(n)}\right]$. We have

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

For $i \neq j$,

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

Since

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

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

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 $i \neq j$,

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

So,

$$\mathbb{E}\left[S_{\mathrm{appr}}^{(n)}\Big|\mathbf{N}^{(n)}\right]=\frac{g^{(n)}}{\sqrt{n}}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varphi_f\left(\widetilde{R}_i^{(n)}\right)F^{-1}\left(\widetilde{R}_j^{(n)}\right). \tag{27}$$

In conclusion, using (23), (27), and (25), we derive that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n\to\infty$,

$$\begin{aligned}
\mathbb{E}\left[L_{3;f}^{(n)}\Big|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\varphi_f\left(\widetilde{R}_i^{(n)}\right)F^{-1}\left(\widetilde{R}_j^{(n)}\right)\left(G_{ij}^{(n)}-\frac{g^{(n)}}{n(n-1)}\right)\\
&\quad+2f(0)\frac{1}{\sqrt{n}}\left(N_+^{(n)}-N_-^{(n)}\right)\mu_f\frac{g^{(n)}}{n}+o_{\mathrm{P}}(1).
\end{aligned}\tag{28}$$

(iv) Now, from (18), (20), and (26) we may conclude that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n\to\infty$,

$$\mathbb{E}\left[\Delta_{f;\lambda}^{(n)}\Big|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]=\Delta_{f;\lambda}^{(n)*}+o_{\mathrm{P}}(1),$$

with $\Delta_{f;\lambda}^{(n)*}$ given by (10). Similarly, we may conclude from (19), (21), and (28) that, under $P_{f;\boldsymbol{\theta}}^{(n)}$, as $n\to\infty$,

$$\mathbb{E}\left[\Delta_{f;\lambda}^{(n)}\Big|\mathbf{N}^{(n)},\mathbf{R}^{(n)}\right]=\widetilde{\Delta}_{f;\lambda}^{(n)*}+o_{\mathrm{P}}(1),$$

with $\widetilde{\Delta}_{f;\lambda}^{(n)*}$ given by (12).

A.3 Semiparametric Fisher information matrix for $\boldsymbol{\theta}$ under $\mathrm{P}^{(n)}_{f;\boldsymbol{\theta}}$

Consider the central sequence $\widetilde{\boldsymbol{\Delta}}^{(n)*}_{f}(\boldsymbol{\theta})$ defined in Proposition 1. Note first that, under $P^{(n)}_{h;\boldsymbol{\theta}}$, as $n \to \infty$,

$$\widetilde{\boldsymbol{\Delta}}^{(n)*}_{f}(\boldsymbol{\theta}) = \boldsymbol{\Delta}^{(n)*}_{f,h}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1),$$

where

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

with, for H being the distribution function associated with density function h,

$$\boldsymbol{\Delta}^{(n)*}_{f,h;\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{B}^{(n)*}_{1;f,h}(\boldsymbol{\theta}) + \mathbf{B}^{(n)*}_{2;f}(\boldsymbol{\theta}), \tag{29}$$

where

$$\mathbf{B}^{(n)*}_{1;f,h}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(H(e^{(n)}_i(\boldsymbol{\theta}))\right)\left(\mathbf{x}^{(n)}_i - \overline{\mathbf{x}}^{(n)}\right),$$

$$\mathbf{B}^{(n)*}_{2;f}(\boldsymbol{\theta}) = 2f(0)\overline{\mathbf{x}}^{(n)} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s^{(n)}_i(\boldsymbol{\theta}),$$

and

$$\Delta^{(n)*}_{f,h;\lambda}(\boldsymbol{\theta}) = L^{(n)*}_{1;f,h}(\boldsymbol{\theta}) + L^{(n)*}_{2;f,h}(\boldsymbol{\theta}) + L^{(n)*}_{3;f,h}(\boldsymbol{\theta}) + L^{(n)*}_{4;f}(\boldsymbol{\theta}), \tag{30}$$

where

$$L^{(n)*}_{1;f,h}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(H(e^{(n)}_i(\boldsymbol{\theta}))\right)\left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda) - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\right)\mathbf{X}^{(n)}\boldsymbol{\beta},$$

$$L^{(n)*}_{2;f,h}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_f\left(H(e^{(n)}_i(\boldsymbol{\theta}))\right) F^{-1}\left(H(e^{(n)}_i(\boldsymbol{\theta}))\right)\left(G^{(n)}_{ii}(\lambda) - \frac{\mathrm{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n}\right),$$

$$L^{(n)*}_{3;f,h}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \varphi_f\left(H(e^{(n)}_i(\boldsymbol{\theta}))\right) F^{-1}\left(H(e^{(n)}_j(\boldsymbol{\theta}))\right)\left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)}\right),$$

$$L^{(n)*}_{4;f}(\boldsymbol{\theta}) = 2f(0)\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s^{(n)}_i(\boldsymbol{\theta})\left(\overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda)\mathbf{X}^{(n)}\boldsymbol{\beta} + \mu_f \frac{g^{(n)}(\lambda)}{n}\right).$$

Under $P^{(n)}_{h;\boldsymbol{\theta}}$, the terms $H(e_i^{(n)}(\boldsymbol{\theta}))$ $(i = 1, \ldots, n)$ are i.i.d. $\mathcal{U}(0,1)$, which implies that, for all $i = 1, \ldots, n$:

$$\begin{aligned}
\mathbb{E}\left[F^{-1}\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right] &= \int_0^1 F^{-1}(u)du = \int_{-\infty}^{\infty} ef(e)de \stackrel{\text{def}}{=} \mu_f; \\
\mathbb{E}\left[\left(F^{-1}\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right)^2\right] &= \int_0^1 \left(F^{-1}(u)\right)^2 du = \int_{-\infty}^{\infty} e^2 f(e)de \stackrel{\text{def}}{=} \nu_f; \\
\mathbb{E}\left[\varphi_f\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right] &= \int_0^1 \varphi_f(u)du = \int_0^1 \phi_f\left(F^{-1}(u)\right)du \\
&= \int_{-\infty}^{\infty} \phi_f(e)f(e)de = 0; \\
\mathbb{E}\left[\varphi_f\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)F^{-1}\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right] &= \int_0^1 \varphi_f(u)F^{-1}(u)du = \int_{-\infty}^{\infty} \phi_f(e)ef(e)de = 1; \\
\mathbb{E}\left[\varphi_f^2\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right] &= \int_0^1 \varphi_f^2(u)du = \int_{-\infty}^{\infty} \phi_f^2(e)f(e)de \stackrel{\text{def}}{=} \mathcal{I}_f; \\
\mathbb{E}\left[\varphi_f^2\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)F^{-1}\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right] &= \int_0^1 \varphi_f^2(u)F^{-1}(u)du = \int_{-\infty}^{\infty} \phi_f^2(e)ef(e)de \stackrel{\text{def}}{=} \mathcal{K}_f; \\
\mathbb{E}\left[\varphi_f^2\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\left(F^{-1}\left(H(e_i^{(n)}(\boldsymbol{\theta}))\right)\right)^2\right] &= \int_0^1 \varphi_f^2(u)\left(F^{-1}(u)\right)^2 du \\
&= \int_{-\infty}^{\infty} \phi_f^2(e)e^2 f(e)de \stackrel{\text{def}}{=} \mathcal{Q}_f.
\end{aligned}$$

Moreover, under $P^{(n)}_{h;\boldsymbol{\theta}}$, the signs $s^{(n)}_i(\boldsymbol{\theta})$ $(i = 1, \ldots, n)$ are i.i.d. $\mathcal{U}\{-1, 1\}$. In addition,

$$\sum_{i=1}^{n} \left(\mathbf{G}^{(n)}_{i\boldsymbol{\cdot}}(\lambda) - \overline{\mathbf{G}}^{(n)}_{\boldsymbol{\cdot}}(\lambda) \right) = \mathbf{0},$$

$$\sum_{i=1}^{n} \left(G^{(n)}_{ii}(\lambda) - \frac{\operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)}{n} \right) = 0,$$

$$\begin{aligned}
\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(G^{(n)}_{ij}(\lambda) - \frac{g^{(n)}(\lambda)}{n(n-1)} \right) \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} G^{(n)}_{ij}(\lambda) - n(n-1)\frac{g^{(n)}(\lambda)}{n(n-1)} \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} G^{(n)}_{ij}(\lambda) - \left(n^2 \overline{G}^{(n)}_{\boldsymbol{\cdot\cdot}}(\lambda) - \operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right)\right) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} G^{(n)}_{ij}(\lambda) - \sum_{i=1}^{n} G^{(n)}_{ii}(\lambda) - n^2 \overline{G}^{(n)}_{\boldsymbol{\cdot\cdot}}(\lambda) + \operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) \\
&= n^2 \overline{G}^{(n)}_{\boldsymbol{\cdot\cdot}}(\lambda) - \operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) - n^2 \overline{G}^{(n)}_{\boldsymbol{\cdot\cdot}}(\lambda) + \operatorname{tr}\left(\mathbf{G}^{(n)}(\lambda)\right) = 0.
\end{aligned}$$

It follows that, under $P^{(n)}_{h;\boldsymbol{\theta}}$,

$$\mathbb{E}\left[\boldsymbol{\Delta}^{(n)*}_{f,h}(\boldsymbol{\theta}) \right] = \mathbf{0}.$$

Using the decompositions (29) and (30), similar calculations than those summarized in Appendix A.1 provide the expressions of $\mathbb{E}\left[\boldsymbol{\Delta}^{(n)*}_{f,h;\boldsymbol{\beta}}(\boldsymbol{\theta}) \left(\boldsymbol{\Delta}^{(n)*}_{f,h;\boldsymbol{\beta}}(\boldsymbol{\theta}) \right)^{\mathrm{T}} \right]$, $\mathbb{E}\left[\left(\Delta^{(n)*}_{f,h;\lambda}(\boldsymbol{\theta}) \right)^2 \right]$ and $\mathbb{E}\left[\boldsymbol{\Delta}^{(n)*}_{f,h;\boldsymbol{\beta}}(\boldsymbol{\theta}) \Delta^{(n)*}_{f,h;\lambda}(\boldsymbol{\theta}) \right]$, under $P^{(n)}_{h;\boldsymbol{\theta}}$, and, at the same time, the expression of matrix $\mathbf{I}^*_f(\boldsymbol{\theta})$.