CentER

Discussion Paper

No. 2004–11

SEMIPARAMETRICALLY EFFICIENT INFERENCE BASED ON SIGNS AND RANKS FOR MEDIAN RESTRICTED MODELS

By M. Hallin, C. Vermandele, B.J.M. Werker

January 2004

ISSN 0924-7815

Tilburg University

Semiparametrically Efficient Inference Based on Signs and Ranks for Median Restricted Models

Marc Hallin[*], Catherine Vermandele[†], and Bas J.M. Werker[‡]

Université Libre de Bruxelles and Tilburg University

January 20, 2004

Abstract

Since the pioneering work of Koenker and Bassett (1978), econometric models involving median and quantile rather than the classical mean or conditional mean concepts have attracted much interest. Contrary to the traditional models where the noise is assumed to have mean zero, median-restricted models enjoy a rich group-invariance structure. In this paper, we exploit this invariance structure in order to obtain semiparametrically efficient inference procedures for these models. These procedures are based on residual signs and ranks, and therefore insensitive to possible misspecification of the underlying innovation density, yet semiparametrically efficient at correctly specified densities. This latter combination is a definite advantage of these procedures over classical quasi-likelihood methods. The techniques we propose can be applied, without additional technical difficulties, to both cross-sectional and time-series models. They do not require any explicit tangent space calculation nor any projections on these.

JEL code C14.

Keywords and phrases. Invariance structure, Local asymptotic normality, Median regression, Quantile restrictions.

---

[*]I.S.R.O., E.C.A.R.E.S., and Département de Mathématique, Université Libre de Bruxelles, CP210, B-1050, Bruxelles, Belgique.

Research supported by an A.R.C. of the Communauté française de Belgique and a P.A.I. contract of the Belgian federal Government

[†]Laboratoire de Méthodologie du Traitement des Données (LMTD) and I.S.R.O., Université libre de Bruxelles, CP210, B-1050, Bruxelles, Belgique.

[‡]Econometrics and Finance Group, CentER, Tilburg University, P.O.Box 90153, 5000 LE, Tilburg, The Netherlands.

1 Introduction

In their seminal 1978 paper, Koenker and Bassett provided an extension to L_1-estimators for median-regression models of the classical results on L_2-estimation in the standard conditional mean setup. Since then, an enormous literature has been devoted to inference problems in median (and other quantile) restricted regression models. The median-regression model considered in Koenker and Bassett (1978) is

$$Y = \mathbf{X}^T\boldsymbol{\theta} + \varepsilon, \tag{1.1}$$

where ε has zero median density f and is independent of $\mathbf{X}$. This specification (1.1) however is still too restrictive in a number of econometric applications, and has been extended in several semi- and nonparametric directions. Restricting attention to models with i.i.d. observations on $(Y, \mathbf{X})$, the fully nonparametric median regression model aims at estimating the conditional median $m(\mathbf{X})$ (m unspecified) of Y given $\mathbf{X}$. The first results in this direction are in Stute (1986) where, via a Donsker-type invariance property, asymptotic normality of a nonparametric conditional quantile estimator of the nearest-neighbor type is established. Other approaches have been used since (for instance, Bhattacharya and Gangopadhyay, 1990), but the nonparametric specification leads to nonparametric, i.e., slower than $\sqrt{n}$, rates of convergence.

In an intermediate semiparametric specification, one imposes

$$Y = m(\mathbf{X};\boldsymbol{\theta}) + \varepsilon, \tag{1.2}$$

where ε, conditionally on $\mathbf{X}$, has median zero, and m is specified. The parameter of interest is $\boldsymbol{\theta}$, and the nuisance parameter is the unknown conditional density of ε given $\mathbf{X}$. In these semiparametric models, $\sqrt{n}$-rate inference, in general, is possible for the parameter $\boldsymbol{\theta}$. For instance, Sherman (1993) proves $\sqrt{n}$-consistency and asymptotic normality for Han (1987)'s maximum correlation estimator. These results even extend to the case where one does not observe Y, but some monotone increasing (not necessarily bijective) transformation $D(Y)$. In a hypothesis testing context, Horowitz and Spokoiny (2002) provide a rate-optimal test for the hypothesis that a conditional median is linear in the explanatory variables.

The advantage of the aforementioned semiparametric approaches is that they allow for arbitrary dependence (e.g., heteroskedasticity) between the innovations ε and the explanatory variables $\mathbf{X}$. However, as usual for non-adaptive models, this generality comes at a cost of reduced efficiency: semiparametric efficiency is strictly smaller than parametric efficiency. Moreover, classical semiparametric inference procedures in these models, as far as we know, all need some form of nonparametric estimation in order to attain semiparametric efficiency bounds. In the present paper, we impose the more restrictive condition that either the innovations ε are independent of the explanatory variables $\mathbf{X}$, or that some parametric form of heteroskedasticity can be specified, i.e., $\varepsilon = \sigma(\mathbf{X};\boldsymbol{\theta})\eta$ with η independent of $\mathbf{X}$ and $\sigma(\cdot,\cdot)$ known. We show that these models have in common a strong group-invariance structure, which allows to base semiparametric inference procedures (estimators and tests) on residual signs and ranks.

These sign-and-rank procedures enjoy $\sqrt{n}$-consistency rates and many other desirable properties. First, they can be constructed in order to achieve semiparametric optimality (i.e., attain the semiparametric efficiency bound) at some preselected density f. Second (but no less important), they are distribution-free, so that their distributional properties are the same under any density g as under $g = f$. In a hypothesis testing context, this means that the resulting tests, while reaching semiparametric efficiency under density f, remain valid (same size) under arbitrary density g. Highly desirable robustness and efficiency properties thus are combined within a single statistic. A third advantage of our procedures is their simplicity: they do not require smoothing of any form, nor techniques such as sample splitting. This again is a consequence of the fact that we specify statistics for a fixed preselected reference density f, rather than estimating the actual density. We will come back to this point shortly. A fourth advantage of our procedures is that they do not require any regularity conditions for the actual underlying density, and very little regularity conditions for the reference density f—mainly, a Local Asymptotic Normality (LAN) condition. This LAN condition is needed since we rely on the convergence under f of our statistical experiments to standard Gaussian shifts in order to substantiate our claims of semiparametric efficiency (in the Hájek-Le Cam framework) at f; this LAN condition is widely accepted in the statistics and econometrics literature as a condition for "regularity" of models. This assumption excludes models with non-smooth parametrization as threshold models and models with non-stationary data. Extension of our results to these cases does not seem trivial. Our results do readily extend to time series models. Zhou and Liang (2000 and 2003) contain results on nonparametric kernel-type estimates of conditional medians for dependent processes. As mentioned before, we are interested in semiparametric inference about $\boldsymbol{\theta}$ and observations may come from standard time series models like ARMA and GARCH models.

Our approach appears as a natural alternative to quasi-likelihood methods. Both our method and the QML method rely on the choice of a reference density f which needs not be the actual density g. However, while the QML method is restricted to a very particular choice of the reference density f (the double-exponential density often will do the job in median restricted location models, but no other density will), our newly proposed method can be based on any zero median reference density f, subject to weak regularity conditions. To illustrate this point, consider an applied researcher facing a median restricted model, for which a QML method based on a double-exponential reference density indeed provides $\sqrt{n}$-consistent inference under a wide class of densities g. If she thinks that another density, h, say, gives a better description of reality, the researcher faces an unpleasant choice. One possibility is to stick to the double-exponential reference density f with the advantage of $\sqrt{n}$-consistent inference, even if the actual density g does not coincide with f, but with reduced efficiency. Alternatively, she may use the density h which is thought to provide a more accurate description of reality, but which generally leads to inconsistent inference in case the actual density g differs from h. (a notable exception is the case where the model is adaptive, but none of the models we will discuss in the present paper is.) However, using our sign-and-rank based method, she can combine the best of two worlds. The inference procedure can safely be applied preselecting the density h as reference density. If this density is indeed correctly specified, semiparametric efficiency is achieved, while the inference procedure remains distribution-free and, consequently, robust to possible misspecification of the density.

This paper mainly deals with the construction of sign-and-rank statistics with the aforementioned properties. In fact, we show how to reconstruct the central sequences in the so-called most difficult parametric submodel, based on signs and ranks of the innovations. These versions of the central sequences can subsequently be used to build inference procedures (estimates and tests) as in Le Cam and Yang (1990), Section 5.3, or in Bickel et al. (1993), Section 2.5.

The remainder of this paper is organized as follows. In the next section, we introduce, in a general setting, the innovation structure that is used in models specified through median restrictions. We give some elementary statistical properties and extend a representation theorem for sign-and-rank statistics of Hallin, Vermandele, and Werker (2003). In Section 3 we study the behavior of sign-and-rank statistics that are based on *parametrically* efficient inference procedures. We show how these statistics behave, both when the underlying innovation density is correctly specified (Theorem 3.1) and when it is not (Theorem 3.2). These results are subsequently used in Section 4 to show that sign-and-rank statistics at correctly specified innovation density attain the *semiparametric* efficiency bound. We illustrate our results throughout by means of a simple median regression and a median autoregression model. In Section 5 we give several other examples. Proofs are gathered in the Appendix.

2 Sign-and-rank statistics

Let $\mathcal{P}_0$ denote the set of all probability distributions on the real line $\mathbb{R}$ that are absolutely continuous with respect to the Lebesgue measure and have median zero. We denote by $\mathcal{F}_0$ the corresponding set of densities, i.e.,

$$\mathcal{F}_0 := \left\{ f : \mathbb{R} \to [0, \infty) : \int_{-\infty}^{0} f(z)\mathrm{d}z = \int_{0}^{\infty} f(z)\mathrm{d}z = \frac{1}{2} \right\}. \tag{2.1}$$

Writing $\mathbb{R}^n$ for the n-dimensional Euclidean space and $\mathcal{B}^n$ for the (completed) Borel sigma-field on this space, we consider the sequence of statistical experiments

$$\mathcal{E}_\varepsilon^{(n)} := \left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}_\varepsilon^{(n)} := \{ \mathbb{P}^n : \mathbb{P} \in \mathcal{P}_0 \} \right), \quad n \in \mathbb{N}, \tag{2.2}$$

of n i.i.d. random variables $\boldsymbol{\varepsilon}^{(n)} := \left(\varepsilon_t^{(n)}\right)_{t=1}^{n}$ with density $f \in \mathcal{F}_0$. Writing $\mathbb{P}_f$ for the distribution on the real line with density f, we may also write $\mathcal{P}_\varepsilon^{(n)} = \left\{ \mathbb{P}_f^n : f \in \mathcal{F}_0 \right\}$.

At the moment, we consider observations $\boldsymbol{\varepsilon}^{(n)} := \left(\varepsilon_t^{(n)}\right)_{t=1}^{n}$ from $\mathcal{E}_\varepsilon^{(n)}$. In Section 3, we will consider the more common and more relevant situation where $\boldsymbol{\varepsilon}^{(n)} = \left(\varepsilon_t^{(n)}\right)_{t=1}^{n}$ are unobserved innovations in some model for observables $\mathbf{Y}^{(n)} := \left(Y_t^{(n)}\right)_{t=1}^n$.

Let $\mathbf{R}^{(n)} := \left(R_1^{(n)}, \ldots, R_n^{(n)}\right)^T$ denote the vector of ranks associated with $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ and let $s_1^{(n)}, \ldots, s_n^{(n)}$ denote their signs, i.e., $s_t^{(n)} := \operatorname{sign}(\varepsilon_t^{(n)})$. Moreover, define $N_+^{(n)} := \#\{t : s_t^{(n)} = 1\}$ as the number of positive variables among $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ and $N_-^{(n)} := \#\{t : s_t^{(n)} = -1\}$ as the number of negative ones. Given our assumption that the distribution of $\varepsilon_t^{(n)}$ is absolutely continuous, we clearly have $N_+^{(n)} + N_-^{(n)} = n$ $(a.s.)$. It is well-known that the vector of ranks $\left(R_1^{(n)}, \ldots, R_n^{(n)}\right)$ is stochastically independent of $\mathbf{N}^{(n)} := (N_+^{(n)}, N_-^{(n)})$. The following result is easily established and provides a theoretical justification for considering signs and ranks in this context (see, e.g., Lehmann 1986, page 315).

Lemma 2.1 *The σ-field*

$$\mathcal{SR}^{(n)} := \sigma\left(\left(R_t^{(n)}, s_t^{(n)}\right)_{t=1}^n\right) = \sigma\left((R_t)_{t=1}^n\right) \vee \sigma\left(N_+^{(n)}\right) \tag{2.3}$$

generated by the ranks and the signs is maximal invariant for the group

$$\mathcal{G}^{(n)}, \circ := \left\{ \mathcal{g}_h^{(n)} \,\middle|\, h : \mathbb{R} \to \mathbb{R} \text{ continuous, monotone} \uparrow,\ h(0) = 0,\ \lim_{x \to \pm\infty} h(x) = \pm\infty \right\}, \circ$$

of order-preserving transformations $(x_1, \ldots, x_n) \longmapsto \mathcal{g}_h^{(n)}(x_1, \ldots, x_n) := (h(x_1), \ldots, h(x_n))$.

The central message of the present paper is that semiparametrically efficient inference procedures, in median restricted models, can be obtained easily (i.e., without, for instance, calculating tangent spaces and imposing many regularity conditions) by using statistics that are based on signs and ranks, i.e., that are $\mathcal{SR}^{(n)}$-measurable. Note that the sigma-field generated by $R_t^{(n)}$ and $s_t^{(n)}$, for given and fixed t, is not the same as the sigma-field generated by $R_t^{(n)}$ and $N_+^{(n)}$. Indeed, the knowledge of $R_t^{(n)}$ and $N_+^{(n)}$ implies the knowledge of $s_t^{(n)}$, but the knowledge of $R_t^{(n)}$ and $s_t^{(n)}$ does not imply the knowledge of $N_+^{(n)}$. Let us define the statistics of interest in this paper.

Definition 2.2 A (linear) sign-and-rank statistic of order k, $k = 0, 1, \ldots, n-1$, for the experiment $\mathcal{E}_\varepsilon^{(n)}$, is a statistic of the form

$$\mathbf{S}_k^{(n)} := \frac{1}{n-k} \sum_{t=k+1}^{n} \mathbf{C}_t^{(n)} \mathbf{a}_k^{(n)} \left(N_+^{(n)}; R_t^{(n)}, \ldots, R_{t-k}^{(n)}\right), \tag{2.4}$$

where $\left(\mathbf{C}_t^{(n)}\right)_{t=1}^n$ are given regression matrices and $\mathbf{a}_k^{(n)}$ is a given $\mathbb{R}^p$-valued score function defined over the set $\{0,\ldots,n\} \times \{\text{all } (k+1)\text{-tuples of distinct integers in } \{1,\ldots,n\}\}$.

Such statistics are called *serial* or *non-serial* according as $k > 0$ or $k = 0$. Hallin et al. (2003) give a detailed representation theorem for the asymptotic behavior of statistics of the form (2.4) in the *non-serial* case $k = 0$ and the *serial* case with $\mathbf{C}_t^{(n)} = \mathbf{I}$, the identity matrix, $t = 1,\ldots,n$. The purpose of this section is to establish an asymptotic representation results for statistics of the general form (2.4) which will be essential in constructing semiparametrically efficient inference procedures in Section 3. The key ingredient in this representation theorem is the notion of a *score-generating function* which we introduce now.

Let $\mathrm{E}_f^{(n)}$ denote expectation under $\mathbb{P}_f^{(n)}$ (expectation under $\mathbb{P}_f^n$ of distribution-free quantities however will be denoted by $\mathrm{E}^{(n)}$ rather than by $\mathrm{E}_f^{(n)}$), and write F for the cumulative distribution function corresponding to f. Finally, let $\boldsymbol{\varepsilon}_{(\cdot)}^{(n)} := \left(\varepsilon_{(t)}^{(n)}\right)_{t=1}^n$ denote the vector of order statistics of $\boldsymbol{\varepsilon}^{(n)}$.

Definition 2.3 A square-integrable function $\boldsymbol{\varphi}_k : (0,1)^{k+1} \to \mathbb{R}^p$ is a score-generating function for the score function $\mathbf{a}_k^{(n)}$ in the sign-and-rank statistic (2.4) if, for all $f \in \mathcal{F}_0$,

$$\mathrm{E}_f^{(n)}\left[\left\|\mathbf{a}_k^{(n)}\left(N_+^{(n)}; R_1^{(n)},\ldots,R_{k+1}^{(n)}\right) - \boldsymbol{\varphi}_k\left(F(\varepsilon_1^{(n)}),\ldots,F(\varepsilon_{k+1}^{(n)})\right)\right\|^2 \middle| \boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right] = o_{\mathbf{P}}(1), \qquad (2.5)$$

under $\mathbb{P}_f^{(n)}$, as $n \to \infty$.

Observe that, by the rule of iterated expectations, a sufficient condition for (2.5) to hold is that

$$\mathrm{E}_f^{(n)}\left[\left\|\mathbf{a}_k^{(n)}\left(N_+^{(n)}; R_1^{(n)},\ldots,R_{k+1}^{(n)}\right) - \boldsymbol{\varphi}_k\left(F(\varepsilon_1^{(n)}),\ldots,F(\varepsilon_{k+1}^{(n)})\right)\right\|^2 \middle| N_+^{(n)}\right] = o_{\mathbf{P}}(1)$$

under $\mathbb{P}_f^{(n)}$, as $n \to \infty$.

The following two propositions are in the vein of Hájek's Projection Theorem, and extend some of the detailed representation results of Hallin et al. (2003). Note that they cover both the *serial* and the *non-serial* case. This generality will be necessary when considering dynamic models with exogenous explanatory variables, as we will see in Section 5.

Proposition 2.4 *Let $\boldsymbol{\varphi}_k : (0,1)^{k+1} \to \mathbb{R}^p$ be a score-generating function for the score function $\mathbf{a}_k^{(n)}$ in the sign-and-rank statistic $\mathbf{S}_k^{(n)}$ in (2.4). Assume that the regression matrices $\mathbf{C}_t^{(n)}$ satisfy*

$$\bar{\mathbf{C}}^{(n)} := \frac{1}{n-k}\sum_{t=k+1}^{n} \mathbf{C}_t^{(n)} = O(1) \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n} \mathbf{C}_t^{(n)} \left[\mathbf{C}_t^{(n)}\right]^T = O(1),$$

as $n \to \infty$. Define the statistic

$$\mathbf{T}^{(n)}_{\varphi_k;f;k} := \frac{1}{n-k}\sum_{t=k+1}^{n} \mathbf{C}_t^{(n)} \boldsymbol{\varphi}_k\left(F(\varepsilon_t^{(n)}), \ldots, F(\varepsilon_{t-k}^{(n)})\right), \tag{2.6}$$

and let

$$\begin{aligned} \boldsymbol{\varphi}_k^*(u_0, \ldots, u_k) \;:=\; & \boldsymbol{\varphi}_k(u_0, \ldots, u_k) \\ & -\sum_{l=0}^{k} \mathrm{E}\left[\boldsymbol{\varphi}_k(U_0, \ldots, U_k) \middle| U_l = u_0\right] + k\mathrm{E}\left[\boldsymbol{\varphi}_k(U_0, \ldots, U_k)\right], \end{aligned} \tag{2.7}$$

where $U_0, \ldots, U_k$ are i.i.d. random variables uniformly distributed over $[0,1]$. *Then, under $\mathbf{P}_f^n$, as $n \to \infty$,*

$$\mathbf{S}_k^{(n)} - \mathrm{E}^{(n)}\left[\mathbf{S}_k^{(n)} \middle| N_+^{(n)}\right] = \mathbf{T}^{(n)}_{\boldsymbol{\varphi}_k;f;k} - \mathrm{E}_f^{(n)}\left[T^{(n)}_{\boldsymbol{\varphi}_k;f;k} \middle| \boldsymbol{\varepsilon}^{(n)}_{(\cdot)}\right] + o_{\mathbf{P}}(n^{-1/2}), \tag{2.8}$$

and $\mathbf{S}_k^{(n)}$ admits the asymptotic representation

$$\begin{aligned} \mathbf{S}_k^{(n)} - \mathrm{E}^{(n)}\left[\mathbf{S}_k^{(n)} \middle| N_+^{(n)}\right] \;=\; & \frac{1}{n-k}\sum_{t=k+1}^{n} \left(\mathbf{C}_t^{(n)} - \bar{\mathbf{C}}^{(n)}\right) \boldsymbol{\varphi}_k\left(F(\varepsilon_t^{(n)}), \ldots, F(\varepsilon_{t-k}^{(n)})\right) \\ & + \frac{\bar{\mathbf{C}}^{(n)}}{n-k}\sum_{t=k+1}^{n} \boldsymbol{\varphi}_k^*\left(F(\varepsilon_t^{(n)}), \ldots, F(\varepsilon_{t-k}^{(n)})\right) + o_{\mathbf{P}}(n^{-1/2}). \end{aligned} \tag{2.9}$$

This proposition allows for studying the asymptotic behavior of the sign-and-rank statistic $\mathbf{S}_k^{(n)}$ conditional on the number $N_+^{(n)}$ of positive signs in the vector $\boldsymbol{\varepsilon}^{(n)}$. The asymptotic results we need, however, are the unconditional ones. If, in addition to the conditions of Proposition 2.4, we assume that the scores $\mathbf{a}_k^{(n)}$ are the so-called *exact scores* associated with $\boldsymbol{\varphi}_k$, that is, if

$$\mathbf{a}_k^{(n)}(n_+; r_0, \ldots, r_k) := \mathrm{E}_f^{(n)}\left[\boldsymbol{\varphi}_k\left(F(\varepsilon_t^{(n)}), \ldots, F(\varepsilon_{t-k}^{(n)})\right) \middle| N_+^{(n)} = n_+; R_t^{(n)} = r_0, \ldots, R_{t-k}^{(n)} = r_k\right], \tag{2.10}$$

the following proposition establishes the unconditional behavior of $\mathbf{S}_k^{(n)}$. In order not to overload the paper, we only consider this exact score case here. Under extra regularity conditions on the function $\boldsymbol{\varphi}_k$, similar results also hold under the weaker assumption that $\boldsymbol{\varphi}_k$ is simply a score-generating function for $\mathbf{a}_k^{(n)}$ (the *approximate scores* case); these results are easily derived by combining the results below with those of Hallin et al. (2003).

Proposition 2.5 *In addition to the conditions of Proposition 2.4, assume that the scores $\mathbf{a}_k^{(n)}$ are the exact scores defined in (2.10). Define the function $\bar{\boldsymbol{\varphi}}_k : \{-1,1\}^{k+1} \to \mathbb{R}^p$ by*

$$\begin{aligned}\bar{\boldsymbol{\varphi}}_k(s_0,\ldots,s_k) &:= \mathrm{E}_f^{(n)}\left[\boldsymbol{\varphi}_k\left(F(\varepsilon_t^{(n)}),\ldots,F(\varepsilon_{t-k}^{(n)})\right)\Big|\,\mathrm{sign}(\varepsilon_t^{(n)})=s_0,\ldots,\mathrm{sign}(\varepsilon_{t-k}^{(n)})=s_k\right]\\ &= 2^{k+1}\int_{u_0=(1+s_0)/4}^{(3+s_0)/4}\cdots\int_{u_k=(1+s_k)/4}^{(3+s_k)/4}\boldsymbol{\varphi}_k(u_0,\ldots,u_k)\,du_0\cdots du_k,\end{aligned}$$

and the constant vector $\bar{\boldsymbol{\varphi}}_k$ by

$$\bar{\boldsymbol{\varphi}}_k := 4\sum_{(s_0,\ldots,s_k)\in\{-1,1\}^{k+1}}\bar{\boldsymbol{\varphi}}_k(s_0,\ldots,s_k)\,\frac{\#\{l:\ s_l=1\}-(k+1)/2}{2^{k+1}}.$$

Then, under $\mathbf{P}_f^n$, as $n\to\infty$, we have

$$\mathrm{E}^{(n)}\left[\mathbf{S}_k^{(n)}\Big|\,N_+^{(n)}\right]-\mathrm{E}^{(n)}\left[\mathbf{S}_k^{(n)}\right]=\left(\frac{N_+^{(n)}}{n}-\frac{1}{2}\right)\bar{\mathbf{C}}^{(n)}\bar{\boldsymbol{\varphi}}_k+o_{\mathrm{P}}(n^{-1/2}),\tag{2.11}$$

and, consequently,

$$\begin{aligned}\mathbf{S}_k^{(n)}-\mathrm{E}^{(n)}\left[\mathbf{S}_k^{(n)}\right] &= \frac{1}{n-k}\sum_{t=k+1}^{n}\left(\mathbf{C}_t^{(n)}-\bar{\mathbf{C}}^{(n)}\right)\boldsymbol{\varphi}_k\left(F(\varepsilon_t^{(n)}),\ldots,F(\varepsilon_{t-k}^{(n)})\right)\\ &\quad+\frac{\bar{\mathbf{C}}^{(n)}}{n-k}\sum_{t=k+1}^{n}\boldsymbol{\varphi}_k^*\left(F(\varepsilon_t^{(n)}),\ldots,F(\varepsilon_{t-k}^{(n)})\right)\\ &\quad+\left(\frac{N_+^{(n)}}{n}-\frac{1}{2}\right)\bar{\mathbf{C}}^{(n)}\bar{\boldsymbol{\varphi}}_k+o_{\mathrm{P}}(n^{-1/2}).\end{aligned}\tag{2.12}$$

Proposition 2.5 gives a representation of sign-and-rank statistics in terms of sums of i.i.d. random variables (clearly, $(N_+^{(n)}/n-1/2)=(2n)^{-1}\sum_{t=1}^n\mathrm{sign}(\varepsilon_t^{(n)})$). Under suitable conditions on the regression matrices, one easily derives a normal limiting distribution for these statistics. Note, however, that the representation as such is obtained under minimal conditions. To illustrate Proposition 2.5, consider the so-called *non-serial* case, i.e., $k=0$, for $p=1$. In this case we write $\varphi=\varphi_0$ and impose $\int_0^1\varphi(u)\mathrm{d}u=0$ so that $\bar{\varphi}(-1)=2\int_0^{1/2}\varphi(u)\mathrm{d}u=-\bar{\varphi}(1)$. Consequently, $\bar{\varphi}=\bar{\varphi}_0=-\bar{\varphi}(-1)+\bar{\varphi}(1)=2\bar{\varphi}(1)$ and, since

$\varphi^* = 0$,

$$S_0^{(n)} = \frac{1}{n}\sum_{t=1}^{n}\left(c_t^{(n)} - \bar{c}^{(n)}\right)\varphi\left(F\left(\varepsilon_t^{(n)}\right)\right) + 2\bar{c}^{(n)}\bar{\varphi}(1)\left(\frac{N_+^{(n)}}{n} - \frac{1}{2}\right) + o_{\mathrm{P}}(n^{-1/2}).$$

The key assumption underlying Propositions 2.4 and 2.5 is the existence of a score-generating function $\boldsymbol{\varphi}_k$ for $\mathbf{a}_k^{(n)}$. Often, however, one is interested in finding a function $\mathbf{a}_k^{(n)}$ with a given score generating function $\boldsymbol{\varphi}_k$. This occurs, e.g., if one is interested in constructing sign-and-rank statistics with a particular asymptotic behavior. More specifically, we propose below to use the *parametrically* efficient score function (i.e., the derivative of the log-likelihood contribution of a single observation in some time-series or cross-sectional model) as the score-generating function. The main contribution of the present paper is that such a choice leads to semiparametrically efficient inference at correctly specified reference density f in the semiparametric model $\mathcal{E}_y^{(n)}$ below, without any tangent space calculation, and with the additional benefits of distribution-free sign-and-rank-based inference. This main result is formalized in Theorem 4.1 below.

3 Sign-and-rank statistics based on parametric scores

Hallin and Werker (2003) show, in a general setting, that there is an intimate relationship between group invariance and semiparametric efficiency. That paper focusses on rank-based statistics in semiparametric models where some completely unrestricted innovation density plays the role of a nuisance. However, their results suggest that the fact that $\mathcal{SR}^{(n)}$ constitutes the maximal invariant σ-field for the median-restricted model $\mathcal{E}_\varepsilon^{(n)}$ (Lemma 2.1) can be used to construct semiparametrically efficient inference procedures in these models. In the present and the next section we investigate this in detail and conclude that sign-and-rank statistics based on *parametrically* efficient (i.e., likelihood-based) scores, automatically yield *semiparametrically* efficient inference procedures. More precisely, such procedures are robust against misspecification of the innovation density, while attaining the semiparametric efficiency bound when the density is correctly specified. For expository reasons, we consider a simple regression model as a running example. Other more interesting and more relevant examples are discussed in the next section. We start by describing the *semiparametric* model of interest.

The general class of models we consider is constructed using an invertible transformation $\mathcal{T}_{\boldsymbol{\theta}}^{(n)}$, depending on a Euclidean parameter of interest $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, and possibly also on some initial conditions or exogenous variables $Y_0^{(n)}$, of an i.i.d. sequence of innovations $\left(\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}\right)$. Formally, we introduce the mapping

$$\mathcal{T}_{\boldsymbol{\theta}}^{(n)} : \ \left(\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}; Y_0^{(n)}\right) \longmapsto \left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right), \tag{3.1}$$

and we assume that we observe $Y_0^{(n)}$ and $\left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right) = \mathcal{T}_{\boldsymbol{\theta}}^{(n)} \left(\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}; Y_0^{(n)}\right)$. The parameter of interest is $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, but the density $f \in \mathcal{F}_0$ of the innovations $(\varepsilon_1^{(n)}, \ldots \varepsilon_n^{(n)})$ constitutes an infinite dimensional nuisance parameter. Our model is thus semiparametric in nature. Formally, the sequence of experiments we are considering is

$$\mathcal{E}_y^{(n)} := \left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}_y^{(n)} = \left\{\mathbb{P}_{\boldsymbol{\theta},f}^{(n)} := \mathbb{P}_f^n \, \mathcal{T}_{\boldsymbol{\theta}}^{(n)\leftarrow} \ : \ \boldsymbol{\theta} \in \boldsymbol{\Theta}, \ f \in \mathcal{F}_0\right\}\right), \tag{3.2}$$

where $\mathcal{T}^{\leftarrow}$ denotes the inverse of the transformation $\mathcal{T}$. The measures $\mathbb{P}_{\boldsymbol{\theta},f}^{(n)}$ are the conditional probability measures of $\left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right)$ given $Y_0^{(n)}$, where the fact that $Y_0^{(n)}$ contains only initial values and/or exogenous variables is formalized by the condition that the distribution of $Y_0^{(n)}$ does not depend on either $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ or $f \in \mathcal{F}_0$. Finally, we assume that the transformation $\mathcal{T}_{\boldsymbol{\theta}}^{(n)}$ is invertible (given $Y_0^{(n)}$), i.e., given observed values $Y_0^{(n)}, Y_1^{(n)}, \ldots, Y_n^{(n)}$ and given a parameter value $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we may calculate

$$\left(\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})\right) := \mathcal{T}_{\boldsymbol{\theta}}^{(n)\leftarrow} \left(Y_1^{(n)}, \ldots, Y_n^{(n)}; Y_0^{(n)}\right). \tag{3.3}$$

Example 3.1 Consider a sequence $X_1^{(n)}, \ldots, X_n^{(n)}$ of real-valued exogenous variables and a sequence $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ of i.i.d. zero median random innovations. We observe $\left(Y_1^{(n)}, \ldots, Y_n^{(n)}\right)$, where

$$Y_t^{(n)} = \theta X_t^{(n)} + \varepsilon_t^{(n)}, \ t = 1, \ldots, n. \tag{3.4}$$

If we put $Y_0^{(n)} = \left(X_1^{(n)}, \ldots, X_n^{(n)}\right)$ and

$$\mathcal{T}_\theta^{(n)}(\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}; Y_0^{(n)}) := \left(\theta X_1^{(n)} + \varepsilon_1^{(n)}, \ldots, \theta X_n^{(n)} + \varepsilon_n^{(n)}\right),$$

hence

$$\mathcal{T}_{\boldsymbol{\theta}}^{(n)\leftarrow} \left(Y_1^{(n)}, \ldots, Y_n^{(n)}; Y_0^{(n)}\right) = \left(Y_1^{(n)} - \theta X_1^{(n)}, \ldots, Y_n^{(n)} - \theta X_n^{(n)}\right),$$

the model fits exactly in the setup of this section. □

Example 3.2 Autoregressive models are another example fitting into the above setup. For simplicity, we restrict to AR(1) models; more general dynamic models are treated in Section 5. In the AR(1) model, one observes $\left(Y_0^{(n)}, \dots, Y_n^{(n)}\right)$, where

$$Y_t^{(n)} = \theta Y_{t-1}^{(n)} + \varepsilon_t^{(n)}, \ t = 1, \dots, n, \tag{3.5}$$

$\theta \in \Theta := (-1, 1)$, $\left(\varepsilon_1^{(n)}, \dots, \varepsilon_n^{(n)}\right)$ are i.i.d. innovations with density $f \in \mathcal{F}_0$, and $Y_0^{(n)}$ is exogenous. Then, letting

$$\mathcal{T}_\theta^{(n)}(\varepsilon_1^{(n)}, \dots, \varepsilon_n^{(n)}; Y_0^{(n)}) := \left(\theta Y_0^{(n)} + \varepsilon_1^{(n)}, \dots, \sum_{i=0}^{t-1} \theta^i \varepsilon_{t-i}^{(n)} + \theta^t Y_0^{(n)}, \dots, \sum_{i=0}^{n-1} \theta^i \varepsilon_{n-i}^{(n)} + \theta^n Y_0^{(n)}\right),$$

we have

$$\mathcal{T}_{\boldsymbol{\theta}}^{(n)\leftarrow}\left(Y_1^{(n)}, \dots, Y_n^{(n)}; Y_0^{(n)}\right) = \left(Y_1^{(n)} - \theta Y_0^{(n)}, \dots, Y_n^{(n)} - \theta Y_{n-1}^{(n)}\right),$$

□

As mentioned before, the goal of the present paper is to show that suitably constructed sign-and-rank statistics in the model described are *semiparametrically efficient*. We will use the local and asymptotic efficiency concept as introduced by Hájek and Le Cam. Following their approach, we impose that the *parametric* model associated with given innovation density f satisfies the Uniform Local Asymptotic Normality (ULAN) condition:

Condition (ULAN). We assume that, for some fixed $f \in \mathcal{F}_0$, the sequence of (parametric) experiments $\mathcal{E}_y^{(n)}(f) = \left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}_y^{(n)}(f) = \left\{\mathbb{P}_{\boldsymbol{\theta},f}^{(n)} = \mathbb{P}_f^n \mathcal{T}_{\boldsymbol{\theta}}^{(n)\leftarrow} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\right\}\right)$ is *Uniformly Locally Asymptotically Normal* (ULAN) in the parameter of interest $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, with a central sequence of the form

$$\boldsymbol{\Delta}_{\boldsymbol{\theta},f}^{(n)} = \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathbf{C}_t^{(n)}(Y_0^{(n)}) \boldsymbol{\varphi}_f^{(k)}\left(F\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right), \dots, F\left(\varepsilon_{t-k}^{(n)}(\boldsymbol{\theta})\right)\right), \tag{3.6}$$

where $\boldsymbol{\varphi}_f^{(k)}$ is a square-integrable function and $\mathbf{C}_t^{(n)}$ are given functions, and with Fisher information $\mathbf{I}_f(\boldsymbol{\theta})$. Hence, under $\mathbb{P}_{\boldsymbol{\theta},f}^{(n)}$ and as $n \to \infty$, $\boldsymbol{\Delta}_{\boldsymbol{\theta},f}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{I}_f(\boldsymbol{\theta}))$. Moreover, we assume the central sequence (3.6) to form a martingale difference sequence, i.e., for all $u_1, \dots, u_k \in (0, 1)$,

$$\int_0^1 \boldsymbol{\varphi}_f^{(k)}(u_0, u_1, \dots, u_k) \mathrm{d}u_0 = 0, \tag{3.7}$$

and the regression matrices are assumed to satisfy, as $n \to \infty$, for some non-stochastic $\bar{\mathbf{C}}$,

$$\bar{\mathbf{C}}^{(n)} := \frac{1}{n-k} \sum_{t=k+1}^{n} \mathbf{C}_t^{(n)}(Y_0^{(n)}) \xrightarrow{\mathbf{P}} \bar{\mathbf{C}}.$$

□

This local asymptotic normality condition is by now widely accepted as the standard framework for the asymptotic analysis of "regular" statistical models. It essentially implies that the model under study is close (in the appropriate, local and asymptotic, sense) to a simple model where one observes a single observation from a multivariate normal distribution with known variance, the unknown mean of which is the parameter of interest. This model is known as the *Gaussian shift model*. Jeganathan (1995) gives an accessible summary of the main results in this literature. Note that we have imposed the so-called uniform LAN condition (ULAN). This condition imposes uniformity in the LAN condition over $\sqrt{n}$-neighborhoods. For the derivation of asymptotic Cramér-Rao type lower bounds on the behavior of estimators via the convolution theorem, this uniformity is not required. However, for the construction of inference procedures with desirable properties, this uniformity is necessary. It is important to note that ULAN is equivalent to LAN plus some asymptotic linearity property of the central sequence (see, e.g., Bickel et al., 1993).

As for the martingale difference condition (3.7), it is not standardly imposed in (U)LAN conditions. However, to the best of our knowledge, it is satisfied in essentially all locally and asymptotically normal models studied so far in the literature. We use this condition later to get simple asymptotic representations of the sign-and-rank statistics based on the parametric score function $\boldsymbol{\varphi}_f^{(k)}$.

Example 3.1 (continued). In the regression model we consider, Condition (ULAN) holds if the innovation density f is absolutely continuous with finite Fisher information for location, i.e., if $I_f := \int (f'/f)^2 \mathrm{d}F < \infty$ and $\frac{1}{n}\sum_{t=1}^{n} \left(X_t^{(n)}\right)^2 \xrightarrow{\mathbf{P}} m_X^2 > 0$. Under these conditions, the central sequence takes the form

$$\Delta_{\theta,f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{-f'}{f} \left(\varepsilon_t^{(n)}(\theta)\right) X_t^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_f m_X^2), \tag{3.8}$$

where $\varepsilon_t^{(n)}(\theta) = Y_t^{(n)} - \theta X_t^{(n)}$. Consequently, we have $k = 0$, $\varphi_f^{(0)}(u) = ((-f'/f)\circ F^{-1})(u)$, and the matrices $\mathbf{C}_t^{(n)}(Y_0^{(n)})$ reduce to scalars: $c_t^{(n)}(Y_0^{(n)}) = X_t^{(n)}$. The martingale difference condition on the terms in the central sequence is clear satisfied, since

$$\int_0^1 \varphi_f^{(0)}(u)\mathrm{d}u = -\int_{-\infty}^{\infty} f'(z)\mathrm{d}z = 0.$$

□

Standard as it is, Condition (ULAN) with central sequence of the form (3.6) is too restrictive for autoregressive models, unless the "memory" or "lag" parameter k is allowed to increase with n.

Example 3.2 (continued). Assuming, as in Example 3.1, that f is absolutely continuous with finite Fisher information for location I_f, but assuming also that the variance σ_ε^2 of the innovation is finite, it is well known (Swensen 1985, Kreiss 1987, or Drost et al. 1997; note that the fact that $\mathrm{E}_f(\varepsilon_t^{(n)}) \neq 0$ does not play any role in that respect), that Condition (ULAN) holds with (univariate) central sequence

$$\Delta_{\theta,f}^{(n)} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{-f'}{f}(\varepsilon_t^{(n)}(\theta))Y_{t-1}^{(n)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n-1}\theta^{i-1}\sum_{t=i+1}^{n}\frac{-f'}{f}(\varepsilon_t^{(n)}(\theta))\varepsilon_{t-i}^{(n)}(\theta), \tag{3.9}$$

where $\varepsilon_t^{(n)}(\theta) := Y_t^{(n)} - \theta Y_{t-1}^{(n)}$ (for notational simplicity we put $Y_0^{(n)} = 0$), and with information $\frac{I_f}{1-\theta^2}m_\varepsilon^2$ with $m_\varepsilon^2 := \mathrm{E}_f(\varepsilon_t^{(n)})^2$. It follows that, for any sequence $(k(n))$ such that $k(n) < n-1$ and $k(n) \uparrow \infty$,

$$\Delta_{k(n),\theta,f}^{(n)} := \frac{1}{\sqrt{n}}\sum_{i=1}^{k(n)}\theta^{i-1}\sum_{t=i+1}^{n}\frac{-f'}{f}(\varepsilon_t^{(n)}(\theta))\varepsilon_{t-i}^{(n)}(\theta), \tag{3.10}$$

which under $\mathbf{P}_{\theta,f}^{(n)}$ differs from (3.9) by a $o_{\mathbf{P}}(1)$ quantity, is still a central sequence.

This central sequence involves unbounded lags, and therefore does not have the required form (3.6). However, observe that, for all n, $\Delta_{k(n),\theta,f}^{(n)}$ is a linear combination, with exponentially decreasing coefficients $m_\varepsilon I_f^{1/2}\theta^{i-1}\sqrt{(n-i)/n}$, of $k(n)$ mutually uncorrelated statistics of the form $\sqrt{n-i}\; r_{f;i}^{(n)}$, where

$$r_{f;i}^{(n)} = r_{f;i}^{(n)}(\theta) := \frac{1}{(n-i)m_\varepsilon I_f^{1/2}}\sum_{t=i+1}^{n}\frac{-f'}{f}(\varepsilon_t^{(n)}(\theta))\varepsilon_{t-i}^{(n)}(\theta). \tag{3.11}$$

Each statistic $\sqrt{n-i}\ r^{(n)}_{f;i}$ under $\mathrm{P}^{(n)}_{\theta,f}$ has mean zero and unit variance, and is of the required form (3.6), with constants $c^{(n)}_t = 1$, and a score function $\varphi^{(i)}_f(u_0, u_1, \ldots, u_i) := \left((-f'/f)\circ F^{-1}\right)(u_0)F^{-1}(u_i)$; it admits the same interpretation, and plays the same role, as traditional residual autocorrelations. The martingale difference condition is clearly satisfied. As we shall see, all the results which, for the sake of simplicity, we are deriving under Condition (ULAN) still hold under this more general setting. □

Very roughly, the main consequence of the Hájek-Le Cam theory is that in the *parametric* model $\mathcal{E}^{(n)}_y(f)$, locally and asymptotically optimal inference can (and should) be based on the central sequence $\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f}$, treating it as if it were the single observation from a Gaussian shift model (see, once more, e.g., Jeganathan, 1995, for details). This implies that the score function $\boldsymbol{\varphi}^{(k)}_f$, which makes up the central sequence, plays a crucial role in the parametric model associated with density f. However, it is not possible to put this score function to immediate good use in the *semiparametric* model $\mathcal{E}^{(n)}_y$. The problem lies in the fact that the score-function $\boldsymbol{\varphi}^{(k)}_f$ is not appropriately centered anymore under innovation densities $g \neq f$. More precisely, it is generally not true that we have $\mathrm{E}^{(n)}_g \boldsymbol{\varphi}^{(k)}_f \left(F(\varepsilon^{(n)}_t), \ldots, F(\varepsilon^{(n)}_{t-k})\right) = 0$ for $g \in \mathcal{F}_0$ and $g \neq f$. A variation in the underlying density thus has the same shift effect on the central sequence $\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f}$ as certain variations in the parameter of interest. Semiparametric theory usually palliates this *confounding effect*, in an optimal way, by projecting $\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f}$ along the *tangent spaces* associated with the variations of innovation densities. These projections yield *semiparametrically efficient score* (or *influence*) *functions*, defining *semiparametrically efficient central sequences.* This approach in general requires nontrivial tangent space calculation.

However, general results in Hallin and Werker (2003) suggest that a version of the same semiparametrically efficient central sequence can be obtained, in the presence of a group-invariance structure of the type we have here, by simply considering $\boldsymbol{\varphi}^{(k)}_f$ as a score-generating function in the sense of Definition 2.3. The first result of the present section (Theorem 3.1) gives the asymptotic behavior of the resulting sign-and-rank statistic. Our main result (Theorem 4.1 in the next section) shows that this statistic indeed provides a version of the semiparametrically efficient central sequence, hence leads to *semiparametrically* efficient inference.

It is important to insist that this semiparametric efficiency is obtained automatically, due to the use of sign-and-rank statistics, and does not need any explicit calculation of tangent spaces or efficient score functions. Moreover, the resulting statistic is also distribution-free over $\mathcal{F}_0$. As explained in the introduction, this is in sharp contrast with the more standard quasi-likelihood approaches.

Let us first introduce the sign-and-rank statistic based on the parametric central sequence $\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f}$. Given a value for the parameter of interest $\boldsymbol{\theta} \in \mathbf{\Theta}$, we are able to calculate (3.3) the residuals $\varepsilon^{(n)}_t(\boldsymbol{\theta})$ from the observations $Y^{(n)}_0, Y^{(n)}_1, \ldots, Y^{(n)}_n$. Denote by $(R^{(n)}_t(\boldsymbol{\theta}))^n_{t=1}$ the ranks of these residuals, and by $N^{(n)}_+(\boldsymbol{\theta})$ the number of positive ones. We consider the following sign-and-rank statistic, of the form $\sqrt{n-k}\mathbf{S}^{(n)}_k$:

$$\begin{aligned}
\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\boldsymbol{\theta},f} &:= \mathrm{E}^{(n)}_{\boldsymbol{\theta},f}\left\{\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f}\middle| N^{(n)}_+(\boldsymbol{\theta}); R^{(n)}_1(\boldsymbol{\theta}), \ldots, R^{(n)}_n(\boldsymbol{\theta}); Y^{(n)}_0\right\} \qquad (3.12)\\
&\qquad\qquad\qquad \Big| N^{(n)}_+(\boldsymbol{\theta}); R^{(n)}_1(\boldsymbol{\theta}), \ldots, R^{(n)}_n(\boldsymbol{\theta}); Y^{(n)}_0\Big\}\\
&= \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathbf{C}^{(n)}_t(Y^{(n)}_0)\mathrm{E}^{(n)}_{\boldsymbol{\theta},f}\left\{\boldsymbol{\varphi}^{(k)}_f\left(F(\varepsilon^{(n)}_t(\boldsymbol{\theta})), \ldots, F(\varepsilon^{(n)}_{t-k}(\boldsymbol{\theta}))\right)\right.\\
&\qquad\qquad\qquad \Big| N^{(n)}_+(\boldsymbol{\theta}); R^{(n)}_t(\boldsymbol{\theta}), \ldots, R^{(n)}_{t-k}(\boldsymbol{\theta}); Y^{(n)}_0\Big\}\\
&= \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathbf{C}^{(n)}_t(Y^{(n)}_0)\mathbf{a}^{(n)}_k(N^{(n)}_+(\boldsymbol{\theta}); R^{(n)}_t(\boldsymbol{\theta}), \ldots, R^{(n)}_{t-k}(\boldsymbol{\theta})),
\end{aligned}$$

with (the so-called *exact* scores)

$$\mathbf{a}^{(n)}_k(n_+; r_0, \ldots, r_k) := \mathrm{E}\left\{\boldsymbol{\varphi}^{(k)}_f\left(U^{(n)}_t, \ldots, U^{(n)}_{t-k}\right) \Big| N^{(n)}_+ = n_+; R^{(n)}_t = r_0, \ldots, R^{(n)}_{t-k} = r_k\right\},$$

where $U^{(n)}_1, \ldots, U^{(n)}_n$ are i.i.d. $U(0,1)$ random variables, $R^{(n)}_1, \ldots, R^{(n)}_n$ their ranks, and $N^{(n)}_+ := \#\{t :\ U^{(n)}_t > 1/2\} \sim \mathrm{Bin}(n, 1/2)$ the number of such U's that exceed 1/2.

Although the score-function $\mathbf{a}^{(n)}_k$ is based on the *parametrically* efficient score-function $\boldsymbol{\varphi}^{(k)}_f$, it is not the case inference based on $\mathbf{a}^{(n)}_k$ is asymptotically as efficient (at f) as inference based on $\boldsymbol{\varphi}^{(k)}_f$. Indeed, $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\boldsymbol{\theta},f} - \mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f}$ in general is not $o_{\mathrm{P}}(1)$ under $\mathbb{P}^{(n)}_{\boldsymbol{\theta},f}$, as $n \to \infty$. Theorem 3.1, based on Proposition 2.5, below makes this precise.

Theorem 3.1 *Consider the sign-and-rank statistic* $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\boldsymbol{\theta},f}$ *as defined in (3.12) in the experiment* $\mathcal{E}^{(n)}_y$ *satisfying Condition (ULAN). Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$ *and as* $n\to\infty$*,*

$$\begin{aligned}\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\boldsymbol{\theta},f} &= \frac{1}{\sqrt{n-k}}\sum_{t=k+1}^{n}\Big\{\mathbf{C}^{(n)}_t(\mathrm{Y}^{(n)}_0)\boldsymbol{\varphi}^{(k)}_f\left(F\left(\varepsilon^{(n)}_t(\boldsymbol{\theta})\right),\ldots,F\left(\varepsilon^{(n)}_{t-k}(\boldsymbol{\theta})\right)\right) \\ &\qquad - \bar{\mathbf{C}}^{(n)}\mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,U_1,\ldots,U_k)\Big|\,U_0=F\left(\varepsilon^{(n)}_t(\boldsymbol{\theta})\right)\right] \\ &\qquad + \bar{\mathbf{C}}^{(n)}\mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,U_1,\ldots,U_k)\Big|\,U_0>1/2\right]\mathrm{sign}\left(\varepsilon^{(n)}_t(\boldsymbol{\theta})\right)\Big\} \\ &\quad + o_{\mathrm{P}}(1) \\ &= \mathbf{\Delta}^{(n)}_{\boldsymbol{\theta},f} - \bar{\mathbf{C}}^{(n)}\frac{1}{\sqrt{n-k}}\sum_{t=k+1}^{n}\Big\{\mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,U_1,\ldots,U_k)\Big|\,U_0=F\left(\varepsilon^{(n)}_t(\boldsymbol{\theta})\right)\right] \\ &\qquad + \mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,U_1,\ldots,U_k)\Big|\,U_0\le 1/2\right]\mathrm{sign}\left(\varepsilon^{(n)}_t(\boldsymbol{\theta})\right)\Big\} \\ &\quad + o_{\mathrm{P}}(1).\end{aligned} \tag{3.13}$$

Moreover, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},f}$*,* $\underset{\sim}{\mathbf{\Delta}}^{(n)}_{\boldsymbol{\theta},f}\xrightarrow{\mathcal{L}} N(0,\mathbf{V})$*, with*

$$\begin{aligned}\mathbf{V} &:= \mathbf{I}_f(\boldsymbol{\theta}) - \bar{\mathbf{C}}\Big[\mathrm{Var}\left\{\mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,\ldots,U_k)\Big|\,U_0\right]\right\} - \\ &\quad \mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,\ldots,U_k)\Big|\,U_0>1/2\right]\mathrm{E}\left[\boldsymbol{\varphi}^{(k)}_f(U_0,\ldots,U_k)\Big|\,U_0>1/2\right]^T\Big]\bar{\mathbf{C}}^T,\end{aligned} \tag{3.14}$$

as $n\to\infty$*.*

Example 3.1 (continued). In our regression example, we immediately obtain

$$\mathrm{E}\left[\varphi^{(0)}_f(U_0)\Big|\,U_0=F\left(\varepsilon^{(n)}_t(\theta)\right)\right] = -\frac{f'}{f}\left(\varepsilon^{(n)}_t(\theta)\right),$$

and

$$\mathrm{E}\left[\varphi^{(0)}_f(U_0)\Big|\,U_0\le 1/2\right] = 2\int_0^{1/2}\varphi_f(u)\mathrm{d}u = 2\int_0^{\infty}f'(z)\mathrm{d}z = -2f(0),$$

since the finiteness of $I_f=\int(f'/f)^2\mathrm{d}F$ implies $\lim_{z\to\infty}f(z)=0$. As a result, letting $a^{(n)}_0(n_+;r)=\mathrm{E}\left\{-\frac{f'}{f}\left(F^{-1}\left(U^{(n)}_t\right)\right)\Big|N^{(n)}_+=n_+;R^{(n)}_t=r\right\}$ and $\bar{X}^{(n)}:=\frac{1}{n}\sum_{t=1}^{n}X^{(n)}_t$ we obtain, for $k=0$,

$$\begin{aligned}\underset{\sim}{\Delta}^{(n)}_{\theta,f} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}X^{(n)}_t a^{(n)}_0(N^{(n)}_+;R^{(n)}_t) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{-f'}{f}\left(\varepsilon^{(n)}_t(\theta)\right)X^{(n)}_t - \frac{\bar{X}^{(n)}}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{-f'}{f}\left(\varepsilon^{(n)}_t(\theta)\right)+2f(0)\mathrm{sign}\left(\varepsilon^{(n)}_t(\theta)\right)\right)+o_{\mathrm{P}}(1).\end{aligned} \tag{3.15}$$

As is well-known from the literature, see, e.g., Bickel (1982), a special situation occurs when $\bar{X}^{(n)} = o_{\mathbf{P}}(1)$. In that case there is no efficiency loss when considering the sign-and-rank statistic $\underset{\sim}{\Delta}{}^{(n)}_{\theta,f}$ as compared to the parametrically efficient statistic $\Delta^{(n)}_{\theta,f}$ and the model is adaptive. That is, from a statistical point of view, the parametric model, with f known, and the semiparametric model, with f unknown, are equally difficult. In that case, $\underset{\sim}{\Delta}{}^{(n)}_{\theta,f}$ can also be based on the ranks of the residuals alone and these statistics are also parametrically efficient, compare Hallin and Werker (2003). □

Example 3.2 (continued). An asymptotic representation result similar to (3.13) also holds for autoregressive central sequences. For each $r^{(n)}_{f:i}$ defined in (3.11), let

$$\begin{aligned} \underset{\sim}{r}{}^{(n)}_{f:i} &:= \mathrm{E}_{\theta,f}\left[r^{(n)}_{f;i}\,\Big|N_+(\theta);R^{(n)}_1(\theta),\ldots,R^{(n)}_n(\theta);Y^{(n)}_0\right] \qquad (3.16)\\ &= \frac{1}{n-i}\sum_{t=i+1}^{n}\mathrm{E}_f\left[\left(\frac{-f'}{f}\circ F^{-1}\right)(F(\varepsilon_t))F^{-1}(F(\varepsilon_{t-i}))\,\Big|N_+;R^{(n)}_t,R^{(n)}_{t-i};Y^{(n)}_0\right]\\ &= \frac{1}{n-i}\sum_{t=i+1}^{n} b^{(n)}\left(N_+;R^{(n)}_t,R^{(n)}_{t-i}\right) \end{aligned}$$

with

$$b^{(n)}\left(n_+;r_0,r_1\right) := \mathrm{E}\left[\left(\frac{-f'}{f}\circ F^{-1}\right)(U^{(n)}_t)F^{-1}(U^{(n)}_{t-1})\,\Big|N_+=n_+;R^{(n)}_t=r_0,R^{(n)}_{t-1}=r_1\right].$$

Along the same lines as in Theorem 3.1, we obtain

$$\begin{aligned} \underset{\sim}{r}{}^{(n)}_{f:i} &= r^{(n)}_{f:i} - \frac{1}{(n-i)m_\varepsilon I_f^{1/2}}\sum_{t=i+1}^{n}\mathrm{E}\left[\left(\frac{-f'}{f}\circ F^{-1}\right)(U^{(n)}_0)F^{-1}(U^{(n)}_1)\,\Big|U^{(n)}_0=F(\varepsilon^{(n)}_t(\theta))\right]\\ &\quad-\frac{1}{(n-i)m_\varepsilon I_f^{1/2}}\sum_{t=i+1}^{n}\mathrm{E}\left[\left(\frac{-f'}{f}\circ F^{-1}\right)(U^{(n)}_0)F^{-1}(U^{(n)}_1)\,\Big|U^{(n)}_0\le 1/2\right]\mathrm{sign}(\varepsilon^{(n)}_t(\theta))\\ &\quad+o_{\mathbf{P}}(n^{-1/2}) \qquad (3.17)\\ &= r^{(n)}_{f:i} - \frac{1}{(n-i)m_\varepsilon I_f^{1/2}}\left\{\sum_{t=i+1}^{n}\frac{-f'}{f}(\varepsilon^{(n)}_t(\theta)) - 2f(0)(2N_+-n)\right\}\mu_\varepsilon + o_{\mathbf{P}}(n^{-1/2}). \end{aligned}$$

Similarly, defining

$$\underset{\sim}{\Delta}{}^{(n)}_{k(n),\theta,f} := \mathrm{E}_{\theta,f}\left[\Delta^{(n)}_{k(n),\theta,f}\,\Big|N_+(\theta);R^{(n)}_1(\theta),\ldots,R^{(n)}_n(\theta);Y^{(n)}_0\right] = m_\varepsilon I_f^{1/2}\sum_{i=1}^{k(n)}\theta^{i-1}\frac{n-i}{\sqrt{n}}\underset{\sim}{r}{}^{(n)}_{f:i},$$

we obtain

$$\underset{\sim}{\Delta}^{(n)}_{k(n),\theta,f} = \Delta^{(n)}_{k(n),\theta,f} - \frac{\mu_\varepsilon}{\sqrt{n}} \sum_{i=1}^{k(n)} \theta^{i-1} \left\{ \sum_{t=i+1}^{n} \frac{-f'}{f}(\varepsilon_t^{(n)}(\theta)) - 2f(0)(2N_+ - n) \right\} + o_{\mathbf{P}}(1).$$

Indeed, decompose $\underset{\sim}{\Delta}^{(n)}_{k(n),\theta,f}$ into $\underset{\sim}{\Delta}^{(n)}_{\kappa,\theta,f} + m_\varepsilon I_f^{1/2} \sum_{i=\kappa+1}^{k(n)} \theta^{i-1} \frac{n-i}{\sqrt{n}} \underset{\sim}{r}^{(n)}_{f:i}$: for any $\epsilon > 0$, there exists a $\kappa = \kappa(\epsilon)$ such that, for any $n \geq k^{-1}(\kappa)$,

$$\mathrm{E}\left[\left(\Delta^{(n)}_{k(n),\theta,f} - \Delta^{(n)}_{\kappa,\theta,f} \right)^2 \right] = \sum_{i=\kappa+1}^{k(n)} \theta^{2(i-1)} \frac{(n-i)^2}{n} \left(r^{(n)}_{f:i} \right)^2 < \epsilon^2.$$

On the other hand, Theorem 3.1 applies to $\underset{\sim}{\Delta}^{(n)}_{\kappa,\theta,f}$ (equivalently, (3.17) applies to any finite linear combination of $\underset{\sim}{r}^{(n)}_{i,f}$'s). The desired asymptotic representation follows. □

Theorem 3.1 gives the asymptotic behavior of the sign-and-rank statistic (3.12) under the assumption that the parametric score function is based on the same density f as the actual distribution of the innovations. The key argument in favor of the use of sign-and-rank statistics, is that they are robust to misspecification of the innovation density. Moreover, this robustness does not come at the cost of efficiency loss, as we will see that the sign-and-rank statistics we consider in this section attain the semiparametric efficiency bound (see Section 4). We now first discuss the behavior of the sign-and-rank statistic (3.12) in case the true innovation density is $g \neq f$. As the proof is completely similar to that of Theorem 3.1, it is omitted. Note, however, that the distribution-freeness of the signs and ranks is crucial.

Theorem 3.2 *Consider the sign-and-rank statistic* $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f}$ *as defined in (3.12) in the experiment* $\mathcal{E}_y^{(n)}(f)$ *satisfying Condition (ULAN). Then, for any* $g \in \mathcal{F}_0$, *under* $\mathbf{P}^{(n)}_{\boldsymbol{\theta},g}$ *and as* $n \to \infty$,

$$\begin{aligned} \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f} \;=\; & \frac{1}{n-k} \sum_{t=k+1}^{n} \Big\{ \mathbf{C}_t^{(n)}(Y_0^{(n)}) \boldsymbol{\varphi}_f^{(k)} \left(G\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right), \ldots, G\left(\varepsilon_{t-k}^{(n)}(\boldsymbol{\theta})\right) \right) \\ & \qquad - \bar{\mathbf{C}}^{(n)} \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1, \ldots, U_k) \Big| U_0 = G\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right) \right] \\ & \qquad + \bar{\mathbf{C}}^{(n)} \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1, \ldots, U_k) \Big| U_0 > 1/2 \right] \mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right) \Big\} \\ & + o_{\mathrm{P}}(n^{-1/2}). \end{aligned} \tag{3.18}$$

4 Semiparametric efficiency of sign-and-rank statistics

We now show that the sign-and-rank statistics (3.12) based on the *parametrically* efficient score functions from the LAN structure of parametric subexperiments attain the *semiparametric* efficiency bound for inference about the finite-dimensional parameter $\boldsymbol{\theta}$ of interest.

The key idea can already be seen from the results in Example 3.1. The difference between the parametric central sequence $\Delta_{\theta,f}^{(n)}$ in Example 3.1 and the sign-and-rank statistic $\underset{\sim}{\Delta}_{\theta,f}^{(n)}$ based on the parametric score $\varphi_f^{(0)}$ is given by $\bar{X}^{(n)}\left(\frac{-f'}{f}(\varepsilon) - 2f(0)\mathrm{sign}(\varepsilon)\right)$. In semiparametric parlance, this function is the projection of the parametric score function $-f'/f$ onto the tangent space for the median regression model. The first part of this projection, $\bar{X}^{(n)}\frac{-f'}{f}(\varepsilon)$ was obtained in Bickel (1982), as discussed by Newey (1990). This projection would be relevant if the distribution of ε would be completely unspecified. The second part $-2\bar{X}^{(n)}f(0)\mathrm{sign}(\varepsilon)$ occurs because of the median restriction imposed on the distribution of ε. This explains how the sign-and-rank statistic attains the semiparametric lower bound.

The same analysis extends to the much more general case of median-restricted models under a weak condition on the smoothness of residuals in the model as a function of the parameter of interest $\boldsymbol{\theta}$. Let us first introduce this condition.

Define the p-valued function

$$\boldsymbol{\psi}_f(u) := \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1, \ldots, U_k)\middle| U_0 = u\right]. \tag{4.1}$$

For ease of notation, define

$$\mathbf{V}_{\boldsymbol{\psi}} := \mathrm{E}[\boldsymbol{\psi}_f(U)\boldsymbol{\psi}_f(U)^T] \quad \text{and} \quad \boldsymbol{\mu}_{\boldsymbol{\psi}}^+ := \mathrm{E}[\boldsymbol{\psi}_f(U)|U > 1/2].$$

Note that, in view of (3.7), we have $\mathrm{E}[\boldsymbol{\psi}_f(U)] = \mathbf{0}$, so that $\mathbf{V}_{\boldsymbol{\psi}}$ is the $p \times p$ covariance matrix of $\boldsymbol{\psi}_f$, and $\boldsymbol{\mu}_{\boldsymbol{\psi}}^+ = -\boldsymbol{\mu}_{\boldsymbol{\psi}}^- := -\mathrm{E}[\boldsymbol{\psi}_f(U)|U \leq 1/2] = 2\int_{1/2}^1 \boldsymbol{\psi}_f(u)\mathrm{d}u$. Now, consider the residual statistic

$$\mathbf{T}^{(n)}(\boldsymbol{\theta}) := \frac{1}{\sqrt{n}}\sum_{t=1}^n \begin{pmatrix} \boldsymbol{\psi}_f\left(F\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right)\right) \\ \mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right) \end{pmatrix}.$$

Under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$, $\mathbf{T}^{(n)}(\boldsymbol{\theta})$ is clearly asymptotically (as $n \to \infty$) normal with mean zero and covariance matrix

$$\begin{bmatrix} \mathbf{V}_{\boldsymbol{\psi}} & \boldsymbol{\mu}^{+}_{\boldsymbol{\psi}} \\ \boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}} & 1 \end{bmatrix}.$$

Assume that this convergence also holds jointly with the central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta},f}$, i.e.,

$$\begin{pmatrix} \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta},f} \\ \mathbf{T}^{(n)}(\boldsymbol{\theta}) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{I}_f(\boldsymbol{\theta}) & \bar{\mathbf{C}}\mathbf{V}_{\boldsymbol{\psi}} & \bar{\mathbf{C}}\boldsymbol{\mu}^{+}_{\boldsymbol{\psi}} \\ \mathbf{V}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T & \mathbf{V}_{\boldsymbol{\psi}} & \boldsymbol{\mu}^{+}_{\boldsymbol{\psi}} \\ \boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T & \boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}} & 1 \end{bmatrix}\right)$$

under $\mathbb{P}^{(n)}_{\boldsymbol{\theta},f}$, as $n \to \infty$. From Le Cam's third Lemma, it follows immediately that, under $\mathbb{P}^{(n)}_{\boldsymbol{\theta}_n,f}$ with $\boldsymbol{\theta}_n = \boldsymbol{\theta} + \mathbf{h}_n/\sqrt{n} + o(n^{-1/2})$ and $\mathbf{h}_n \to \mathbf{h}$ as $n \to \infty$, $\mathbf{T}^{(n)}(\boldsymbol{\theta})$ is asymptotically normal with mean

$$\begin{pmatrix} \mathbf{V}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T\mathbf{h} \\ \boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T\mathbf{h} \end{pmatrix},$$

and unchanged covariance matrix. Since $\mathcal{L}_{\boldsymbol{\theta}_n,f}\left(\left(\varepsilon^{(n)}_t(\boldsymbol{\theta}_n)\right)_{t=1}^n\right) = \mathcal{L}_{\boldsymbol{\theta},f}\left(\left(\varepsilon^{(n)}_t(\boldsymbol{\theta})\right)_{t=1}^n\right)$, we find that $\mathbf{T}^{(n)}(\boldsymbol{\theta}_n) + [\mathbf{V}_{\boldsymbol{\psi}}, \boldsymbol{\mu}^{+}_{\boldsymbol{\psi}}]^T\bar{\mathbf{C}}^T\mathbf{h}$ and $\mathbf{T}^{(n)}(\boldsymbol{\theta})$ are "close in distribution", under $\mathbb{P}^{(n)}_{\boldsymbol{\theta},f}$, as $n \to \infty$. The condition (Condition (S)) that we impose, and check in the examples of Section 5, is that they are in fact "close in probability":

Condition (S). Under $\mathbb{P}^{(n)}_{\boldsymbol{\theta},f}$ and as $n \to \infty$, we have

$$\mathbf{T}^{(n)}(\boldsymbol{\theta}_n) - \mathbf{T}^{(n)}(\boldsymbol{\theta}) + \begin{bmatrix} \mathbf{V}_{\boldsymbol{\psi}} \\ \boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}} \end{bmatrix} \bar{\mathbf{C}}^T\mathbf{h} = o_{\mathbf{P}}(1).$$

□

Condition (S) is close to the smoothness condition studied in Bickel et al (1993), Proposition 2.1.2. In the literature it is often referred to as an *asymptotic linearity* property.

Under the above Conditions (ULAN) and (S), we can show that the sign-and-rank statistic $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f}$ attains the semiparametric lower bound in $\mathcal{E}^{(n)}_y$. The proof of this result follows from classical arguments. Statistical inference in a submodel of $\mathcal{E}^{(n)}_y$ is by definition easier than in the complete model $\mathcal{E}^{(n)}_y$, in the sense that the (semiparametric) lower bound is smaller. Thus, to prove that a certain statistic attains the semiparametric lower bound in the larger model $\mathcal{E}_y^{(n)}$, it suffices to prove that it attains the lower bound induced by some (parametric) submodel. We will follow this line of reasoning below as well.

Theorem 4.1 *In the semiparametric model $\mathcal{E}_y^{(n)}$ and under the Conditions (ULAN) and (S), the asymptotic variance of the distribution-free sign-and-rank statistic*

$$\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\theta},f}^{(n)} := \frac{1}{\sqrt{n-k}} \sum_{t=k+1}^{n} \mathbf{C}_t^{(n)}(Y_0^{(n)})\mathbf{a}_k^{(n)}(N_+^{(n)}(\boldsymbol{\theta}); R_t^{(n)}(\boldsymbol{\theta}), \ldots, R_{t-k}^{(n)}(\boldsymbol{\theta})), \tag{4.2}$$

constructed from the parametric score function $\boldsymbol{\varphi}_f^{(k)}$, i.e., with scores

$$\mathbf{a}_k^{(n)}(n_+; r_0, \ldots, r_k) := \mathrm{E}\left\{\boldsymbol{\varphi}_f^{(k)}\left(U_t^{(n)}, \ldots, U_{t-k}^{(n)}\right) \middle| N_+^{(n)} = n_+; R_t^{(n)} = r_0, \ldots, R_{t-k}^{(n)} = r_k\right\}, \tag{4.3}$$

equals, under $\mathbf{P}_{\boldsymbol{\theta},f}^{(n)}$ and as $n \to \infty$, the semiparametric lower bound. More precisely, there exists a regular (LAN) parametric submodel in which the statistic $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\theta},f}^{(n)}$ is (parametrically) efficient, that is, $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\theta},f}^{(n)}$ is the central sequence in this parametric model.

Example 3.1 (continued). In our running regression example, Theorem 4.1 can be applied immediately to the sign-and-rank statistic (3.15). Since Condition (ULAN) is satisfied, it only remains to verify condition (S). We consider the first and second components of the statistic $\mathbf{T}^{(n)}(\theta)$ separately. Smoothness (asymptotic linearity) of the first component is immediate from the ULAN property of the pure location model (i.e., the model of the present example with $X_t^{(n)} = 1$). Smoothness of the second component is a well-known result in the literature (see, e.g., Van Eeden, 1972).

It follows that this statistic attains the semiparametric lower bound. Note that such a statistic could be obtained via the more traditional tangent space arguments as well, but the point here is that the projections needed in such calculations are automatically carried out by using signs and ranks of the innovations. □

Example 3.2 (continued). Here again, condition (ULAN) is satisfied, while condition (S) follows from the ULAN property of autoregressive models with a constant term (a particular case of the model considered, e.g., by Swensen (1985)), along with the smoothness of the sign-based component of $\mathbf{T}^{(n)}(\theta)$. $\square$

Several remarks are appropriate at this point. First of all, the sign-and-rank statistic (4.2) is based on *exact* scores, i.e., the score function $\mathbf{a}_k^{(n)}$ is the exact expectation of the parametric score function $\boldsymbol{\varphi}_f^{(k)}$, as defined in (4.3). Note that the argument of the expectation in (4.3) depends on f, but that calculating the expectation given the values of ranks and signs can be done without knowledge about the actual density, due to the distribution-freeness of the ranks and signs in the model we consider. A simple simulation algorithm thus could be used to calculate $\mathbf{a}_k^{(n)}$ up to arbitrary precision, given $\boldsymbol{\varphi}_f^{(k)}$. Alternatively, one also may use the so-called *approximate scores* that generally result from substituting $F^{-1}(R_t^{(n)}/(n+1))$ for the residuals $\varepsilon_t^{(n)}(\boldsymbol{\theta})$. For the sign-and-rank statistics, these approximate scores take the form

$$
\begin{aligned}
&\mathbf{a}_{k;\mathrm{approx}}^{(n)}(n_+; r_0, \ldots, r_k) \\
&\quad= \boldsymbol{\varphi}_f^{(k)}\Bigg(I\{r_0 \le n_-\}\frac{r_0}{2(n_-+1)} + I\{r_0 > n_-\}\left[\frac{1}{2} + \frac{r_0 - n_-}{2(n_+ + 1)}\right], \ldots, \\
&\qquad\qquad I\{r_k \le n_-\}\frac{r_k}{2(n_-+1)} + I\{r_k > n_-\}\left[\frac{1}{2} + \frac{r_k - n_-}{2(n_+ + 1)}\right]\Bigg),
\end{aligned}
$$

where $n_- = n - n_+$. In most cases, if appropriately centered, these approximate score functions yield the same asymptotic behavior for the sign-and-rank statistic $\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\theta},f}^{(n)}$ as the exact ones $\mathbf{a}_k^{(n)}$, while their numerical evaluation does not require any additional computation once the f-quantile function has been evaluated. This of course saves computing time. However, the additional regularity conditions and the proofs related with these approximate scores are delicate and of little importance to the main concern of the present paper. The interested reader is referred to Hallin et al. (2003) for a detailed treatment of this problem.

Example 3.1 (continued). In the regression example, the semiparametrically efficient sign-and-rank statistic, under approximate score form, easily follows from the remarks above, yielding (for the sake of simplicity, we avoid introducing specific notation for such approximate score versions)

$$
\underset{\sim}{\Delta}_{\theta,f}^{(n)} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_t^{(n)}\left(\frac{-f'}{f}\left(\underset{\sim}{\varepsilon}_{f;t}^{(n)}\right) - z_f\right), \tag{4.4}
$$

where

$$\underset{\sim}{\varepsilon}^{(n)}_{f;t} := F^{-1}\left(I\{R_t \le N_-\}\frac{R_t^{(n)}}{2(N_- + 1)} + I\{R_t^{(n)} > N_-\}\left[\frac{1}{2} + \frac{R_t^{(n)} - N_-}{2(N_+ + 1)}\right]\right) \tag{4.5}$$

(under $\mathbb{P}^{(n)}_{\boldsymbol{\theta},f}$, a *rank-and-sign reconstruction* of $\varepsilon_t^{(n)}$). The centering $z_f = \mathrm{E}(-f'/f)(\underset{\sim}{\varepsilon}^{(n)}_{f;t})$ is needed to ensure that the efficient sign-and-rank statistic is exactly centered. The form (4.4) is more pleasant than (3.15) based on exact scores, as it does not need a simulation routine to calculate the conditional expectation in (3.15). The first-order asymptotic properties of both (4.4) and (3.15) are equal. □

Example 3.2 (continued). Similarly, the approximate score version of $\underset{\sim}{\Delta}^{(n)}_{\theta,f}$ here is obtained from substituting in $\underset{\sim}{\Delta}^{(n)}_{k(n),\theta,f}$ the approximate score versions

$$\underset{\sim}{r}^{(n)}_{f;i} := \frac{1}{(n-i)m_\varepsilon I_f^{1/2}} \sum_{t=i+1}^{n} \left(\frac{-f'}{f}(\underset{\sim}{\varepsilon}^{(n)}_{f;t})\underset{\sim}{\varepsilon}^{(n)}_{f;t-i} - z_f^{(i)} \right), \tag{4.6}$$

of $\underset{\sim}{r}^{(n)}_{f;i}$ for the exact ones. The centering $z_f^{(i)}$ is given by $z_f = \mathrm{E}(-f'/f)(\underset{\sim}{\varepsilon}^{(n)}_{f;t})\underset{\sim}{\varepsilon}^{(n)}_{f;t-i}$, which can again be calculated without simulation. □

A second remark is that the distribution-free sign-and-rank statistic (4.2) reaches the semiparametric efficiency bound associated with density f. This means that semiparametric efficiency is achieved at correctly specified f (i.e., when the actual innovation density g coincides with f). The sign-and-rank statistic (4.2) generally does not attain this bound under incorrectly specified innovation density (under $g \neq f$). It might be possible, under adequate regularity conditions, to have the sign-and-rank statistic adapt to the unknown innovation density by pre-estimating this density, much along the same lines as this is done for rank statistics in Hallin and Werker (2003). The essential difference between the rank-only case and our sign-and-rank case is that in the former the density estimate can be based on the order statistics of the residuals, which are independent of their ranks. In the sign-and-rank case, the order statistics are, however, not independent of the signs that appear in the statistic. Thus, although a result on pre-estimating the density in sign-and-rank statistics would be important, the details probably will be far from trivial and we leave this for possible discussion elsewhere.

5 Further examples

In this final section, we provide several examples to illustrate the scope of our results. Our running examples have shown how to handle simple regression and autoregression models with independent zero median innovations. Our results however apply to more sophisticated models. Some of them are investigated in this section.

Example 5.1 ARMA models

The ARMA(p, q) model is a natural generalization of the AR(1) case considered in Example 3.2. Let the observations $Y_1^{(n)}, \ldots, Y_n^{(n)}$ be generated from the model

$$A(L)Y_t := Y_t - \sum_{i=1}^{p} a_i Y_{t-i} = \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j} =: B(L)\varepsilon_t, \quad t = 1, \ldots, n, \tag{5.1}$$

with starting values $\mathbf{Y}_0^{(n)} = \left(Y_0^{(n)}, Y_{-1}^{(n)}, \ldots, Y_{-p+1}^{(n)}, \varepsilon_0^{(n)}, \varepsilon_{-1}^{(n)}, \ldots, \varepsilon_{-q+1}^{(n)}\right)'$ and i.i.d. innovations $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ with density $f \in \mathcal{F}_0$. Writing $\boldsymbol{\theta} := (a_1, \ldots, a_p, b_1, \ldots, b_q)'$ for the parameter of interest, assume that $a_p \neq 0 \neq b_q$, that $\boldsymbol{\theta}$ is such that the roots of $A(z) = 0$ and $B(z) = 0,\ z \in \mathbb{C}$ are distinct, and that they all lie outside the unit disc; denote by $\boldsymbol{\Theta}$ the set of all such parameter values. As for the innovation density f, the same assumptions are made as in Example 3.2. The model under these assumption satisfies Condition ULAN at f; see, e.g., Hallin and Puri (1985) or Drost et al. (1997).

The explicit form of the central sequence requires some further notation. This notation is cumbersome, but the essential idea of the present paper goes on as before. Letting $C(L) := A(L)B(L) = \sum_{i=1}^{p+q} c_i L^i$, define $g_u = g_u(\boldsymbol{\theta})$, $h_u = h_u(\boldsymbol{\theta})$, and $G_u = G_u(\boldsymbol{\theta})$ by means of

$$(A(L))^{-1} := \sum_{i=1}^{\infty} g_i L^i, \quad (B(L))^{-1} := \sum_{i=1}^{\infty} h_i L^i, \quad (C(L))^{-1} := \sum_{i=1}^{\infty} G_i L^i.$$

It follows from the assumptions on the characteristic roots of $A(z)$ and $B(z)$ that g_i, h_i, and G_i all are $O(\lambda^i)$, as $i \to \infty$, for some $\lambda \in (0,1)$. Define

$$\mathbf{M} = \mathbf{M}(\boldsymbol{\theta}) := \begin{pmatrix} 1 & 0 & & \dots & & 0 & 1 & 0 & & \dots & & 0 & \\ g_1 & 1 & 0 & & \dots & 0 & h_1 & 1 & 0 & & \dots & 0 & \\ \vdots & & & \ddots & & \vdots & & & & \ddots & & \vdots & \\ & & & & & 0 & \vdots & & & & & & \\ g_{p-1} & & \dots & & g_1 & 1 & & & & & & & \\ g_p & & \dots & & g_2 & g_1 & h_{q-1} & & & \dots & h_1 & 1 & \\ \vdots & & & & \vdots & \vdots & h_q & & & \dots & h_2 & h_1 & \\ & & & & & & & & \vdots & & & & \vdots \\ g_{p+q-1} & & \dots & & g_{q+1} & g_q & h_{p+q-1} & & & \dots & h_{p+1} & h_p & \end{pmatrix}$$

and

$$\mathbf{C} = \mathbf{C}(\boldsymbol{\theta}) := \begin{pmatrix} 1 & c_1 & \dots & & c_{p+q-1} \\ 0 & 1 & & & c_{p+q-2} \\ \vdots & & \ddots & & \vdots \\ 0 & & \dots & & c_1 \\ 0 & & \dots & 0 & 1 \end{pmatrix}$$

From Proposition 4.1 in Hallin and Puri (1985), the central sequence can be written, with the same notation as in Example 3.2, as

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta},f} = \left(nm_\varepsilon^2 I_f\right)^{1/2} \mathbf{M}'(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})\left(\sum_{i=1}^{k(n)}(n-i)^{1/2}G_{i-1}r^{(n)}_{f;i}, \dots, \sum_{i=1}^{k(n)}(n-i)^{1/2}G_{i-p-q}r^{(n)}_{f;i}\right)'.$$

Here also, $k(n) \uparrow \infty$ is an increasing sequence that can be chosen arbitrarily in view of the exponential decrease of G_i as $i \to \infty$.

The corresponding efficient sign-and rank central sequence $\underset{\sim}{\Delta}{}^{(n)}_{\theta,f}$ is obtained from replacing the $\underset{\sim}{r}{}^{(n)}_{f;i}$'s either with their conditional expectations (3.16) or by their approximate score sign-and rank counterparts (4.6). □

It is not very difficult to combine the regression results with dynamic models for the regression errors, as the following example shows.

Example 5.2 Dynamic regression models

Consider the regression model with moving average errors of order one

$$Y_t^{(n)} = \boldsymbol{\beta}^T \mathbf{X}_t^{(n)} + \varepsilon_t^{(n)} - \alpha \varepsilon_{t-1}^{(n)},\ t = 1, \ldots, n,$$

where, for simplicity, $Y_0^{(n)} := \varepsilon_0^{(n)}$ is assumed to be observed. The innovations $\varepsilon_t^{(n)}$ are assumed to be i.i.d. with density $f \in \mathcal{F}_0$. We do not know of a paper that discusses a Local Asymptotic Normality result for this model directly, but it is easily verified that Condition (ULAN) is satisfied for the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \alpha)^T$, using, e.g., the results in Drost et al. (1997), under the same assumptions on ε and $\mathbf{X}$ as in the regression example, provided that $|\alpha| < 1$ and ε has finite variance. In that case, the central sequence for $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\Delta}_{\boldsymbol{\theta},f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{bmatrix} \mathbf{X}_t^{(n)} & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -\frac{f'}{f}(\varepsilon_t^{(n)}) \\ -\frac{f'}{f}(\varepsilon_t^{(n)})\varepsilon_{t-1}^{(n)} \end{bmatrix}, \tag{5.2}$$

and Fisher information

$$\mathbf{I}_f(\boldsymbol{\theta}) := I_f \begin{bmatrix} \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t^{(n)} \left(\mathbf{X}_t^{(n)}\right)^T & \mu_\varepsilon \left(\bar{\mathbf{X}}_0\right)^T \\ \mu_\varepsilon \bar{\mathbf{X}}_0 & \mathrm{E}\left(\varepsilon_t^{(n)}\right)^2 \end{bmatrix},$$

with $\mu_\varepsilon := \mathrm{E}\varepsilon_t^{(n)}$, $\bar{\mathbf{X}}_0 := \lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbf{X}_t^{(n)}$, and, as before, $I_f := \int (f'/f)^2 \,\mathrm{d}F$. Theorem 3.1 can be applied directly with $k = 2$ and

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1)\middle| U_0\right] &= \begin{bmatrix} -\frac{f'}{f}(F^{-1}(U_0)) \\ -\frac{f'}{f}(F^{-1}(U_0))\mu_\varepsilon \end{bmatrix}, \\
\mathrm{Var}\left\{\mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1)\middle| U_0\right]\right\} &= I_f \begin{bmatrix} 1 & \mu_\varepsilon \\ \mu_\varepsilon & \mu_\varepsilon^2 \end{bmatrix}, \\
\mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1)\middle| U_0 > 1/2\right] &= 2f(0) \begin{bmatrix} 1 \\ \mu_\varepsilon \end{bmatrix}.
\end{aligned}$$

Concluding, the semiparametrically efficient sign-and-rank statistic is obtained, under approximate score form, upon replacing $\varepsilon_t^{(n)}$ in (5.2) by $\underset{\sim}{\varepsilon}_{f;t}^{(n)}$ defined in (4.5), yielding

$$\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\theta},f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \begin{bmatrix} \mathbf{X}_t^{(n)} & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -\frac{f'}{f}(\underset{\sim}{\varepsilon}_{f;t}^{(n)}) - z_f \\ -\frac{f'}{f}(\underset{\sim}{\varepsilon}_{f;t}^{(n)})\underset{\sim}{\varepsilon}_{f;t-1}^{(n)} - z_f^{(1)} \end{bmatrix}$$

The asymptotic variance, which equals the semiparametric lower bound, is given by (3.14), i.e.,

$$\mathbf{I}_f(\boldsymbol{\theta}) - \Big(I_f - 4f(0)^2\Big) \begin{bmatrix} \bar{\mathbf{X}}_0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \mu_\varepsilon \\ \mu_\varepsilon & \mu_\varepsilon^2 \end{bmatrix} \begin{bmatrix} \bar{\mathbf{X}}_0 & 0 \\ 0 & 1 \end{bmatrix}^T.$$

In order to conclude that the sign-and-rank statistic based on the parametric score function for $\boldsymbol{\theta}$, i.e., based on the central sequence (5.2), provides for semiparametrically efficient inference, we need to check for the smoothness Condition (S). Again, as far as the first component of $\mathbf{T}^{(n)}(\boldsymbol{\theta})$ is concerned, smoothness is a direct consequence of ULAN. Concerning the second component of $\mathbf{T}^{(n)}(\boldsymbol{\theta})$, which contains the signs of the innovations, smoothness is easily verified by combining the corresponding results for the regression model with i.i.d. innovations (our running example) with those for the ARMA models as discussed in the previous example. □

The examples discussed so far, all are essentially variations on location models. Our results are equally applicable to scale models, with the provision that a condition on the median of the innovations generally does not allow for the identification of "unconditional" scale parameters in the model. We start with a model with fully specified heteroskedasticity.

Example 5.3 Median regression with known conditional heteroskedasticity
Consider a model for the conditional median specified as

$$Y_t^{(n)} = \boldsymbol{\beta}^T \mathbf{X}_t^{(n)} + \sigma(\mathbf{X}_t^{(n)})\varepsilon_t^{(n)},\ t = 1, \dots, n, \tag{5.3}$$

for some known function $\sigma(\cdot)$. Under standard regularity conditions as those mentioned above, the central sequence for $\boldsymbol{\theta} = \boldsymbol{\beta}$ is directly seen to be

$$\boldsymbol{\Delta}_{\boldsymbol{\theta},f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} -\frac{f'}{f}(\varepsilon_t^{(n)}) \frac{\mathbf{X}_t^{(n)}}{\sigma(\mathbf{X}_t^{(n)})}. \tag{5.4}$$

This implies that all results of the standard regression model can be applied, upon weighting the observations by $\sigma(\mathbf{X}_t^{(n)})$. This classical approach to heteroskedasticity extends to our results as well. The semiparametrically efficient sign-and-rank statistic is given by

$$\underset{\sim}{\boldsymbol{\Delta}}_{\boldsymbol{\theta},f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(-\frac{f'}{f}(\underset{\sim}{\varepsilon}_{f;t}^{(n)}) - z_f \right) \frac{\mathbf{X}_t^{(n)}}{\sigma(\mathbf{X}_t^{(n)})},$$

where $\underset{\sim}{\varepsilon}{}^{(n)}_{f;t}$ is as in (4.5). □

The next example discusses the case where the heteroskedasticity is unknown, but of a given parametric form. As explained in the introduction, our results concern semiparametric efficiency with respect to unknown innovation densities. Efficiency in models with semiparametric specification of regression functions or volatility functions cannot be obtained using signs and ranks only, since they are not distribution-free with respect to these functional parameters. Parametric functional forms do, however, fall under the scope of sign-and-rank statistics.

Example 5.4 Median regression with parametric conditional heteroskedasticity Consider a model for the conditional median specified as

$$Y_t^{(n)} = \boldsymbol{\beta}^T \mathbf{X}_t^{(n)} + \sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})\varepsilon_t^{(n)},\ t = 1, \dots, n, \tag{5.5}$$

for some known function $\sigma(\cdot;\cdot)$. The central sequence for $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ is known (compare, e.g., Drost et al., 1997) to be

$$\boldsymbol{\Delta}_{\boldsymbol{\theta},f}^{(n)} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \frac{\mathbf{X}_t^{(n)}}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})} & 0 \\ 0 & \frac{\sigma'(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})} \end{bmatrix} \begin{bmatrix} -\frac{f'}{f}(\varepsilon_t^{(n)}) \\ -\left(1 + \varepsilon_t^{(n)} \frac{f'}{f}(\varepsilon_t^{(n)})\right) \end{bmatrix}, \tag{5.6}$$

where $\sigma'(\cdot;\cdot)$ denotes the derivative of $\sigma(\cdot;\cdot)$ with respect to the second argument. The Fisher information follows, again, from the martingale central limit theorem, assuming that a law-of-large-numbers can be applied to the averages of $\frac{\mathbf{X}_t^{(n)}}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}$ and $\frac{\sigma'(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}$. Theorem 3.1 is applicable, with

$$\mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0)\middle| U_0\right] = \boldsymbol{\varphi}_f^{(k)}(U_0), \quad \text{and} \quad \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0)\middle| U_0 > 1/2\right] = \begin{bmatrix} 2f(0) \\ 0 \end{bmatrix},$$

since $\int f'/f(\varepsilon)\mathrm{d}F(\varepsilon) = 0$ and $\int_0^\infty (1 + \varepsilon f'/f(\varepsilon))\,\mathrm{d}F(\varepsilon) = 0$. This shows that the sign-and-rank statistic based on (5.6) is asymptotically equivalent to

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \frac{\mathbf{X}_t^{(n)}}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})} & 0 \\ 0 & \frac{\sigma'(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})} - \overline{\frac{\sigma'(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}} \end{bmatrix} \begin{bmatrix} -\frac{f'}{f}(\varepsilon_t^{(n)}) \\ -\left(1 + \varepsilon_t^{(n)} \frac{f'}{f}(\varepsilon_t^{(n)})\right) \end{bmatrix}$$
$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{bmatrix} \overline{\frac{\mathbf{X}_t^{(n)}}{\sigma(\mathbf{X}_t^{(n)};\boldsymbol{\alpha})}} \left(\frac{f'}{f}(\varepsilon_t^{(n)}) + \mathrm{sign}(\varepsilon_t^{(n)}) 2f(0)\right) \\ 0 \end{bmatrix},$$

where an overline again indicates a time-average. Note that the sign-and-rank statistic does not contain any information for components of $\boldsymbol{\alpha}$ for which $\sigma'(\cdot;\boldsymbol{\alpha})$ is constant. This is due to the fact that such "unconditional" variance parameters are not identified in the semiparametric model. Semiparametric efficiency of the sign-and-rank statistic at correctly specified innovation density f is obtained by verifying Condition (S) for the present model. Again, this follows from the ULAN condition concerning the first component of $\mathbf{T}^{(n)}(\boldsymbol{\theta})$ and for the second component from results on the behavior of sign statistics in scale models as in Hájek and Šidák (1967). Once more, the semiparametrically efficient sign-and-rank statistic based on approximate scores is obtained by replacing $\varepsilon_t^{(n)}$ by $\underset{\sim}{\varepsilon}{}_{f;t}^{(n)}$ in (5.6). □

6 Appendix

Proof of Proposition 2.4: For ease of notation, we consider the univariate case only, i.e., $p = 1$, and one-dimensional regression constants $c_t^{(n)}$ with mean $\bar{c}^{(n)}$. Now, observe

$$\mathrm{E}_f^{(n)}\left[S_k^{(n)} \middle| N_+^{(n)} \right] = \mathrm{E}_f^{(n)}\left[S_k^{(n)} \middle| \boldsymbol{\varepsilon}_{(\cdot)}^{(n)} \right],$$

since, also conditionally on $N_+^{(n)}$, the ranks $(R_1^{(n)},\ldots,R_n^{(n)})$ and the order statistics $\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}$ are independently distributed. Consequently, in order to establish (2.8), it suffices to prove that $\mathrm{Var}_f^{(n)}\left[S_k^{(n)} - T_{\varphi_k;f;k}^{(n)} \middle| \boldsymbol{\varepsilon}_{(\cdot)}^{(n)} \right] = o_{\mathrm{P}}(1/n)$. Let $U_t^{(n)} := F(\varepsilon_t^{(n)})$ and denote by $\mathbf{U}_{(\cdot)}^{(n)}$ the corresponding vector of order statistics. Observe that the ranks $R_1^{(n)},\ldots,R_n^{(n)}$ of $\varepsilon_1^{(n)},\ldots,\varepsilon_n^{(n)}$ are also those of $U_1^{(n)},\ldots,U_n^{(n)}$. Finally, denote

$$\begin{aligned}
\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}(i_0,\ldots,i_k) &:= a_k^{(n)}\left(N_+^{(n)}; i_0,\ldots,i_k\right) - \varphi_k\left(U_{(i_0)}^{(n)},\ldots,U_{(i_k)}^{(n)}\right),\\
\bar{\alpha}^{(n)} &:= \mathrm{E}_f^{(n)}\left[\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}\left(R_t^{(n)},\ldots,R_{t-k}^{(n)}\right) \middle| \mathbf{U}_{(\cdot)}^{(n)} \right]\\
&= \mathrm{E}_f^{(n)}\left[\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}\left(R_t^{(n)},\ldots,R_{t-k}^{(n)}\right) \middle| \boldsymbol{\varepsilon}_{(\cdot)}^{(n)} \right].
\end{aligned}$$

With this notation, one immediately notices that

$$\mathrm{E}_f^{(n)}\left[T_{\varphi_k;f;k}^{(n)} \middle| \boldsymbol{\varepsilon}_{(\cdot)}^{(n)} \right] = \bar{c}^{(n)}\mathrm{E}_f^{(n)}\left[\varphi_k\left(F(\varepsilon_1^{(n)}),\ldots,F(\varepsilon_{k+1}^{(n)})\right) \middle| \boldsymbol{\varepsilon}_{(\cdot)}^{(n)} \right].$$

Now,

$$\begin{aligned}
&\mathrm{Var}_f^{(n)}\left[S_k^{(n)} - T_{\varphi_k;f;k}^{(n)}\middle|\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right] \\
&= \mathrm{Var}_f^{(n)}\left[\frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)}\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}\left(R_t^{(n)},\ldots,R_{t-k}^{(n)}\right)\middle|\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right] \\
&= \mathrm{E}_f^{(n)}\left[\left(\frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)}\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}\left(R_t^{(n)},\ldots,R_{t-k}^{(n)}\right) - \bar{c}^{(n)}\bar{\alpha}^{(n)}\right)^2\middle|\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right] \\
&= \mathrm{E}_f^{(n)}\left[\left(\frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)}\left(\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}\left(R_t^{(n)},\ldots,R_{t-k}^{(n)}\right) - \bar{\alpha}^{(n)}\right)\right)^2\middle|\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right] \\
&\leq \left[\frac{1}{n-k}\sum_{t=k+1}^{n}\left(c_t^{(n)}\right)^2\right]\times \\
&\qquad \mathrm{E}_f^{(n)}\left[\left(\frac{1}{n-k}\sum_{t=k+1}^{n}\left(\alpha_{\mathbf{U}_{(\cdot)};k}^{(n)}\left(R_t^{(n)},\ldots,R_{t-k}^{(n)}\right) - \bar{\alpha}^{(n)}\right)\right)^2\middle|\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right].
\end{aligned}$$

The first factor is $O(1)$ by assumption and the second factor can be seen to be $o_{\mathbf{P}}(1/n)$ following exactly the lines of the proof of Lemma 4.1 in Hallin et al. (2003). This establishes (2.8).

In order to prove (2.9), observe

$$\begin{aligned}
&T_{\varphi_k;f;k}^{(n)} - \mathrm{E}_f^{(n)}\left[T_{\varphi_k;f;k}^{(n)}\middle|\boldsymbol{\varepsilon}_{(\cdot)}^{(n)}\right] \\
&= \frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)}\varphi_k\left(U_t^{(n)},\ldots,U_{t-k}^{(n)}\right) - \bar{c}^{(n)}\mathrm{E}_f^{(n)}\left[\varphi_k\left(U_1^{(n)},\ldots,U_{k+1}^{(n)}\right)\middle|\mathbf{U}_{(\cdot)}^{(n)}\right] \\
&= \frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)}\varphi_k\left(U_t^{(n)},\ldots,U_{t-k}^{(n)}\right) \\
&\quad -\frac{\bar{c}^{(n)}}{n-k}\sum_{t=k+1}^{n}\sum_{l=0}^{k}\int_{u_0=0}^{1}\cdots\int_{u_k=0}^{1}\varphi_k\left(u_0,\ldots,u_{l-1},U_t^{(n)},u_{l+1},\ldots,u_k\right)\mathrm{d}u_0\cdots\mathrm{d}u_k \\
&\quad +k\bar{c}^{(n)}\int_{u_0=0}^{1}\cdots\int_{u_k=0}^{1}\varphi_k(u_0,\ldots,u_k)\mathrm{d}u_0\cdots\mathrm{d}u_k + o_{\mathbf{P}}(n^{-1/2}) \\
&= \frac{1}{n-k}\sum_{t=k+1}^{n}\left(c_t^{(n)} - \bar{c}^{(n)}\right)\varphi_k\left(U_t^{(n)},\ldots,U_{t-k}^{(n)}\right) \\
&\quad +\frac{\bar{c}^{(n)}}{n-k}\sum_{t=k+1}^{n}\varphi_k^*\left(U_t^{(n)},\ldots,U_{t-k}^{(n)}\right) + o_{\mathbf{P}}(n^{-1/2}),
\end{aligned}$$

where the second equality is due to standard U-statistic results and the fact that φ_k is square-integrable (see, e.g., Serfling (1980), Section 5.3). □

Proof of Proposition 2.5: Again, we give the proof for the univariate case $p = 1$ with regression constants $c_t^{(n)}$. First of all, observe

$$
\begin{aligned}
& \mathrm{E}^{(n)}\left[S_k^{(n)} \middle| \mathbf{N}^{(n)}\right] \\
&= \mathrm{E}^{(n)}\left[\frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)} \mathrm{E}^{(n)}\left[\varphi_k\left(F(\varepsilon_t^{(n)}),\ldots,F(\varepsilon_{t-k}^{(n)})\right)\middle| N_+; R_t^{(n)},\ldots,R_{t-k}^{(n)}\right]\middle| \mathbf{N}^{(n)}\right] \\
&= \frac{1}{n-k}\sum_{t=k+1}^{n} c_t^{(n)} \mathrm{E}^{(n)}\left[\varphi_k\left(F(\varepsilon_t^{(n)}),\ldots,F(\varepsilon_{t-k}^{(n)})\right)\middle| \mathbf{N}^{(n)}\right] \\
&= \bar{c}^{(n)} \mathrm{E}^{(n)}\left[\bar{\varphi}_k\left(s_t^{(n)},\ldots,s_{t-k}^{(n)}\right)\middle| \mathbf{N}^{(n)}\right] + o_{\mathbf{P}}(n^{-1/2}) \\
&= \bar{c}^{(n)} \sum_{(s_0,\ldots,s_k)\in\{-1,1\}^{k+1}} \bar{\varphi}_k\left(s_0,\ldots,s_k\right)\left(\frac{N_+^{(n)}}{n}\right)^{\#\{i:\ s_i=1\}}\left(1-\frac{N_+^{(n)}}{n}\right)^{k+1-\#\{i:\ s_i=1\}} \\
&\quad + o_{\mathbf{P}}(n^{-1/2}),
\end{aligned}
$$

where, in the last equality, we have used the binomial approximation of the hypergeometric distribution. Now, since

$$
\left.\frac{\mathrm{d}}{\mathrm{d}p} p^j\left(1-p\right)^{k+1-j}\right|_{p=1/2} = (2j-(k+1))\left(\frac{1}{2}\right)^k,
$$

we obtain

$$
\begin{aligned}
& \mathrm{E}^{(n)}\left[S_k^{(n)} \middle| \mathbf{N}^{(n)}\right] - \mathrm{E}^{(n)}\left[S_k^{(n)}\right] \\
&= \bar{c}^{(n)}\left(\mathrm{E}^{(n)}\left[\bar{\varphi}_k\left(s_t^{(n)},\ldots,s_{t-k}^{(n)}\right)\middle| \mathbf{N}^{(n)}\right] - \mathrm{E}^{(n)}\left[\bar{\varphi}_k\left(s_t^{(n)},\ldots,s_{t-k}^{(n)}\right)\right]\right) \\
&= \bar{c}^{(n)} \sum_{(s_0,\ldots,s_k)\in\{-1,1\}^{k+1}} \bar{\varphi}_k\left(s_0,\ldots,s_k\right) \times \\
&\qquad\left(\left(\frac{N_+^{(n)}}{n}\right)^{\#\{i:\ s_i=1\}}\left(1-\frac{N_+^{(n)}}{n}\right)^{\#\{i:\ s_i=-1\}} - \left(\frac{1}{2}\right)^{k+1}\right) \\
&\quad + o_{\mathbf{P}}(n^{-1/2}) \\
&= \bar{c}^{(n)} \sum_{(s_0,\ldots,s_k)\in\{-1,1\}^{k+1}} \bar{\varphi}_k\left(s_0,\ldots,s_k\right)\frac{2\#\{i:\ s_i=1\}-(k+1)}{2^k}\left(\frac{N_+^{(n)}}{n}-\frac{1}{2}\right) \\
&\quad + o_{\mathbf{P}}(n^{-1/2}) \\
&= \bar{c}^{(n)}\bar{\varphi}_k\left(\frac{N_+^{(n)}}{n}-\frac{1}{2}\right) + o_{\mathbf{P}}(n^{-1/2}).
\end{aligned}
$$

This proves (2.11); (2.12) follows immediately. □

Proof of Theorem 3.1: Since $Y_0^{(n)}$ is exogenous, we may apply Proposition 2.5 conditionally on $Y_0^{(n)}$. The fact that $\boldsymbol{\varphi}_f^{(k)}$ is a score-generating function for $\mathbf{a}_k^{(n)}$ follows immediately from Proposition 3.1 in Hallin et al. (2003). Now, with the notation (2.7), and letting $\boldsymbol{\varphi}_k := \boldsymbol{\varphi}_f^{(k)}$, we obtain, from the martingale difference condition (3.7),

$$\boldsymbol{\varphi}_k^*(u_0, \ldots, u_k) := \boldsymbol{\varphi}_f^{(k)}(u_0, \ldots, u_k) - \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0, U_1, \ldots, U_k) \middle| U_0 = u_0\right].$$

It remains to determine the value of the constant vector $\bar{\boldsymbol{\varphi}}_k$. With the notation of Proposition 2.5, define the signs $s_l = \mathrm{sign}(U_l - 1/2)$, $l = 0, \ldots, k$. Consequently, conditionally on $s_1, \ldots, s_k$, we have that $\bar{\boldsymbol{\varphi}}_k(s_0, s_1, \ldots, s_k)$ has a distribution with a two-point support: $\bar{\boldsymbol{\varphi}}_k(-1, s_1, \ldots, s_k)$ and $\bar{\boldsymbol{\varphi}}_k(+1, s_1, \ldots, s_k)$ (both with probability 1/2). Again due to the martingale difference condition (3.7), we have $\bar{\boldsymbol{\varphi}}_k(-1, s_1, \ldots, s_k) = -\bar{\boldsymbol{\varphi}}_k(+1, s_1, \ldots, s_k)$. Now, still conditionally on $s_1, \ldots, s_k$, the number of positive signs $\#\{l = 0, \ldots, k : \; s_l = 1\}$ also takes a two-point distribution but with values $\#\{l = 1, \ldots, k : \; s_l = 1\}$ and $\#\{l = 1, \ldots, k : \; s_l = 1\} + 1$. Consequently, taking these results together,

$$\begin{aligned}
&\mathrm{E}\left[\bar{\boldsymbol{\varphi}}_k(s_0, s_1, \ldots, s_k)\left(\#\{l = 0, \ldots, k : \; s_l = 1\} - \frac{k+1}{2}\right) \middle| s_1, \ldots, s_k\right] \\
&\quad = \frac{1}{2}\left[\bar{\boldsymbol{\varphi}}_k(+1, s_1, \ldots, s_k)\left(\#\{l = 1, \ldots, k : \; s_l = 1\} + 1 - \frac{k+1}{2}\right)\right. \\
&\quad\quad \left. + \bar{\boldsymbol{\varphi}}_k(-1, s_1, \ldots, s_k)\left(\#\{l = 1, \ldots, k : \; s_l = 1\} - \frac{k+1}{2}\right)\right] \\
&\quad = \frac{1}{2}\bar{\boldsymbol{\varphi}}_k(+1, s_1, \ldots, s_k).
\end{aligned}$$

The constant $\bar{\boldsymbol{\varphi}}_k$, as defined in Proposition 2.5, readily follows:

$$\begin{aligned}
\bar{\boldsymbol{\varphi}}_k &= 4\mathrm{E}\left[\bar{\boldsymbol{\varphi}}_k(s_0, s_1, \ldots, s_k)\left(\#\{l = 0, \ldots, k : \; s_l = 1\} - \frac{k+1}{2}\right)\right] \\
&= 2\mathrm{E}\left[\bar{\boldsymbol{\varphi}}_k(s_0, s_1, \ldots, s_k) | \, s_0 = 1\right] = 2\mathrm{E}\left[\boldsymbol{\varphi}_k(U_0, U_1, \ldots, U_k) | \, U_0 > 1/2\right] \\
&= -2\mathrm{E}\left[\boldsymbol{\varphi}_k(U_0, U_1, \ldots, U_k) | \, U_0 \leq 1/2\right].
\end{aligned}$$

The result (3.13) then is a direct consequence of Proposition 2.5, using the fact that $(N_+^{(n)}/n - 1/2) = (2n)^{-1} \sum_{t=1}^n \mathrm{sign}(\varepsilon_t^{(n)})$.

The limiting distribution (3.14) follows immediately upon applying the martingale central limit theorem to

$$\frac{1}{\sqrt{n-k}}\sum_{t=k+1}^{n}\left[\begin{array}{c}\mathbf{C}_t^{(n)}(Y_0^{(n)})\boldsymbol{\varphi}_f^{(k)}\left(F\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right),\ldots,F\left(\varepsilon_{t-k}^{(n)}(\boldsymbol{\theta})\right)\right)\\ \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0,U_1,\ldots,U_k)\Big|\,U_0=F\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right)\right]\\ \mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta})\right)\end{array}\right].$$

Defining $\mathbf{V}_{\boldsymbol{\psi}} := \mathrm{Var}\left\{\mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0,\ldots,U_k)\Big|\,U_0\right]\right\}$ and $\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+} := \mathrm{E}\left[\boldsymbol{\varphi}_f^{(k)}(U_0,\ldots,U_k)\Big|\,U_0>1/2\right]$ this limiting distribution is $\mathcal{N}(\mathbf{0},\mathbf{W})$ with

$$\mathbf{W} := \left[\begin{array}{ccc}\mathbf{I}_f(\boldsymbol{\theta}) & \bar{\mathbf{C}}\mathbf{V}_{\boldsymbol{\psi}} & \bar{\mathbf{C}}\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\\ \mathbf{V}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T & \bar{\mathbf{C}}\mathbf{V}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T & \bar{\mathbf{C}}\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\\ \left(\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\right)^T\bar{\mathbf{C}}^T & \left(\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\right)^T\bar{\mathbf{C}}^T & 1\end{array}\right].$$

Section 4 contains more detailed calculations that can be used to verify this result. □

The proof of Theorem 4.1 is based on the following lemma. Note that this lemma is not restricted to the specification (3.6) for the central sequence. Nor does it use the uniformity in the Condition (ULAN) (LAN would be enough).

Lemma 6.1 *Assume that the Conditions (ULAN) and (S) hold at the submodel $\mathcal{E}_y^{(n)}(f_0)$. Define the parametric family of densities*

$$f_{\boldsymbol{\eta}}(z) := f_0(z)\exp\left(-a(\boldsymbol{\eta})\boldsymbol{\eta}^T\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\mathrm{sign}(z)+\boldsymbol{\eta}^T\boldsymbol{\psi}_{f_0}(F_0(z))+b(\boldsymbol{\eta})\right),\quad \boldsymbol{\eta}\in\mathbb{R}^p, \tag{6.7}$$

where $a:\mathbb{R}^p\to\mathbb{R}$ and $b:\mathbb{R}^p\to\mathbb{R}$ are such that $\int_{-\infty}^0 f_{\boldsymbol{\eta}}(z)dz=\int_0^\infty f_{\boldsymbol{\eta}}(z)dz=1/2$; clearly, for all $\boldsymbol{\eta}$, $f_{\boldsymbol{\eta}}\in\mathcal{F}_0$. Then the sequence of experiments

$$\bar{\mathcal{E}}_y^{(n)}(f_0) := \left(\mathbb{R}^n,\mathcal{B}^n,\bar{\mathcal{P}}_y^{(n)} := \left\{\mathbf{P}_{\boldsymbol{\theta},\boldsymbol{\eta}}^{(n)} := \mathbf{P}_{f_{\boldsymbol{\eta}}}^n\,\mathcal{T}_{\boldsymbol{\theta}}^{(n)\leftarrow}\;:\;\boldsymbol{\theta}\in\Theta,\ \boldsymbol{\eta}\in\mathbb{R}^p\right\}\right)$$

is LAN at $(\boldsymbol{\theta},\boldsymbol{\eta})=(\boldsymbol{\theta}_0,\mathbf{0})$, with central sequence

$$\left(\begin{array}{c}\boldsymbol{\Delta}_{\boldsymbol{\theta}_0,f_0}^{(n)}\\ \frac{1}{\sqrt{n}}\sum_{t=1}^n\boldsymbol{\psi}_{f_0}(F_0(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)))+\mathrm{E}[\boldsymbol{\psi}_{f_0}(U_0)|U_0\le 1/2]\mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)\right)\end{array}\right), \tag{6.8}$$

and Fisher information

$$\left[\begin{array}{cc}\mathbf{I}_{f_0}(\boldsymbol{\theta}_0) & \bar{\mathbf{C}}\left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+T}\right)\\ \left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+T}\right)\bar{\mathbf{C}}^T & \mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+T}\end{array}\right]. \tag{6.9}$$

Proof of Lemma 6.1: Clearly, for each $\boldsymbol{\eta} \in \mathbb{R}^p$ we have $f_{\boldsymbol{\eta}}(z) \geq 0$. Hence, for each $n \in \mathbb{N}$, $\bar{\mathcal{E}}_y^{(n)}(f_0)$ is a (parametric) subexperiment of $\mathcal{E}_y^{(n)}$. In order to establish the LAN result, we need a few auxiliary calculations. First of all, observe that we can take $a(\mathbf{0}) = 1$ and have $b(\mathbf{0}) = 0$. Moreover, differentiation yields

$$\frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\eta}}(z) \bigg|_{\boldsymbol{\eta}=\mathbf{0}} = f_0(z) \left\{ -\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+} \mathrm{sign}(z) + \boldsymbol{\psi}_{f_0}(F_0(z)) + b'(\mathbf{0}) \right\}.$$

Integrating this expression over the half-line $(-\infty, 0]$ and using $\int_{-\infty}^{0} f_{\boldsymbol{\eta}}(z) \mathrm{d}z = 1/2$ yields

$$\mathbf{0} = \boldsymbol{\mu}_{\boldsymbol{\psi}}^{+}/2 + \mathrm{E}\left[\boldsymbol{\psi}_{f_0}(U) | U \leq 1/2\right] + b'(\mathbf{0}),$$

which implies $b'(\mathbf{0}) = \mathbf{0}$. Similarly, differentiating twice with respect to $\boldsymbol{\eta}$ gives

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\eta}^2} f_{\boldsymbol{\eta}}(z) \bigg|_{\boldsymbol{\eta}=\mathbf{0}} &= f_0(z) \left\{ \left[-\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+} \mathrm{sign}(z) + \boldsymbol{\psi}_{f_0}(F_0(z)) \right] \left[-\boldsymbol{\mu}_{\boldsymbol{\psi}}^{+} \mathrm{sign}(z) + \boldsymbol{\psi}_{f_0}(F_0(z)) \right]^T \right. \\ &\quad \left. - \left(a'(\mathbf{0}) \boldsymbol{\mu}_{\boldsymbol{\psi}}^{+T} + \boldsymbol{\mu}_{\boldsymbol{\psi}} a'(\mathbf{0})^T \right) \mathrm{sign}(z) + b''(\mathbf{0}) \right\}. \end{aligned}$$

which, by integrating over the real line, implies

$$b''(\mathbf{0}) = -\left(\mathbf{V}_{\boldsymbol{\psi}} - \boldsymbol{\mu}_{\boldsymbol{\psi}}^{+} \boldsymbol{\mu}_{\boldsymbol{\psi}}^{+T} \right).$$

To prove that the LAN condition holds for the subexperiment $\bar{\mathcal{E}}_y^{(n)}(f_0)$ at the point $(\boldsymbol{\theta}, \boldsymbol{\eta}) = (\boldsymbol{\theta}_0, \mathbf{0}) \in \boldsymbol{\Theta} \times \mathbb{R}^p$, consider a sequence $(\mathbf{h}_n)_{n=1}^{\infty}$ in $\mathbb{R}^k$ with $\mathbf{h}_n \to \mathbf{h}$ as $n \to \infty$, and a sequence $(\mathbf{g}_n)_{n=1}^{\infty}$ in $\mathbb{R}^p$ with $\mathbf{g}_n \to \mathbf{g}$ as $n \to \infty$. Observe that, as a result of the Condition (LAN), the sequences of probability measures $\mathbb{P}_{\boldsymbol{\theta}_0 + n^{-1/2}\mathbf{h}_n, f_{\mathbf{0}}}^{(n)}$ and $\mathbb{P}_{\boldsymbol{\theta}_0, f_{\mathbf{0}}}^{(n)}$ are contiguous.

Now, in view of the LAN condition on the parametric model, the definition of $f_{\boldsymbol{\eta}}$, Condition (S), the contiguity mentioned above, and the result $b'(\mathbf{0}) = 0$ and $b''(\mathbf{0}) =$

$-\left(\mathbf{V}_{\boldsymbol{\psi}} - \boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right)$ above, we obtain, under $\mathbf{P}^{(n)}_{\boldsymbol{\theta}_0, f_0}$ and as $n \to \infty$,

$$
\begin{aligned}
&\log \frac{\mathrm{d}\mathbf{P}^{(n)}_{f_{\mathbf{g}_n/\sqrt{n}}}\left(\mathcal{T}^{(n)\leftarrow}_{\boldsymbol{\theta}_0+\mathbf{h}_n/\sqrt{n}}(\mathbf{Y}^{(n)})\right)}{\mathrm{d}\mathbf{P}^{(n)}_{f_\mathbf{0}}\left(\mathcal{T}^{(n)\leftarrow}_{\boldsymbol{\theta}_0}(\mathbf{Y}^{(n)})\right)} \\
&= \log \frac{\mathrm{d}\mathbf{P}^{(n)}_{f_{\mathbf{g}_n/\sqrt{n}}}\left(\mathcal{T}^{(n)\leftarrow}_{\boldsymbol{\theta}_0+\mathbf{h}_n/\sqrt{n}}(\mathbf{Y}^{(n)})\right)}{\mathrm{d}\mathbf{P}^{(n)}_{f_\mathbf{0}}\left(\mathcal{T}^{(n)\leftarrow}_{\boldsymbol{\theta}_0+\mathbf{h}_n/\sqrt{n}}(\mathbf{Y}^{(n)})\right)} + \log \frac{\mathrm{d}\mathbf{P}^{(n)}_{f_\mathbf{0}}\left(\mathcal{T}^{(n)\leftarrow}_{\boldsymbol{\theta}_0+\mathbf{h}_n/\sqrt{n}}(\mathbf{Y}^{(n)})\right)}{\mathrm{d}\mathbf{P}^{(n)}_{f_\mathbf{0}}\left(\mathcal{T}^{(n)\leftarrow}_{\boldsymbol{\theta}_0}(\mathbf{Y}^{(n)})\right)} \\
&= \sum_{t=1}^{n} \log \frac{f_{\mathbf{g}_n/\sqrt{n}}}{f_\mathbf{0}}\left(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0 + \mathbf{h}_n/\sqrt{n})\right) + \mathbf{h}_n^T \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0, f_\mathbf{0}} - \frac{1}{2}\mathbf{h}_n^T \mathbf{I}_{f_\mathbf{0}}(\boldsymbol{\theta}_0)\mathbf{h}_n + o_\mathbf{P}(1) \\
&= -a\left(\frac{\mathbf{g}_n}{\sqrt{n}}\right)\frac{\mathbf{g}_n^T}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{\mu}^+_{\boldsymbol{\psi}}\mathrm{sign}(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0 + \mathbf{h}_n/\sqrt{n})) \\
&\qquad + \frac{\mathbf{g}_n^T}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{\psi}_{f_\mathbf{0}}(F_\mathbf{0}(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0 + \mathbf{h}_n/\sqrt{n}))) + nb\left(\frac{\mathbf{g}_n}{\sqrt{n}}\right) \\
&\qquad + \mathbf{h}_n^T \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0, f_\mathbf{0}} - \frac{1}{2}\mathbf{h}_n^T \mathbf{I}_{f_\mathbf{0}}(\boldsymbol{\theta}_0)\mathbf{h}_n + o_\mathbf{P}(1) \\
&= -\frac{\mathbf{g}_n^T}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{\mu}^+_{\boldsymbol{\psi}}\mathrm{sign}(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0)) + \mathbf{g}_n^T\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T\mathbf{h}_n \\
&\qquad + \frac{\mathbf{g}_n^T}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{\psi}_{f_\mathbf{0}}(F_\mathbf{0}(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0))) - \mathbf{g}_n^T V_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T\mathbf{h}_n \\
&\qquad + \frac{1}{2}\mathbf{g}_n^T b''(\mathbf{0})\mathbf{g}_n + \mathbf{h}_n^T \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0, f_0} - \frac{1}{2}\mathbf{h}_n^T \mathbf{I}_{f_0}(\boldsymbol{\theta}_0)\mathbf{h}_n + o_\mathbf{P}(1) \\
&= \left[\mathbf{h}_n^T, \mathbf{g}_n^T\right]\begin{pmatrix} \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0, f_\mathbf{0}} \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\psi}_{f_\mathbf{0}}\left(F_\mathbf{0}(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0))\right) + \mathrm{E}[\boldsymbol{\psi}_{f_\mathbf{0}}(U)|U \le 1/2]\mathrm{sign}\left(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0)\right) \end{pmatrix} \\
&\qquad - \frac{1}{2}\begin{bmatrix} \mathbf{h}_n \\ \mathbf{g}_n \end{bmatrix}^T \boldsymbol{\Sigma} \begin{bmatrix} \mathbf{h}_n \\ \mathbf{g}_n \end{bmatrix} + o_\mathbf{P}(1),
\end{aligned}
$$

where

$$
\boldsymbol{\Sigma} := \begin{bmatrix} \mathbf{I}_{f_\mathbf{0}}(\boldsymbol{\theta}_0) & -\bar{\mathbf{C}}\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}} + \bar{\mathbf{C}}\mathbf{V}_{\boldsymbol{\psi}} \\ -\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T + \mathbf{V}_{\boldsymbol{\psi}}\bar{\mathbf{C}}^T & -b''(\mathbf{0}) \end{bmatrix}.
$$

Note that the quadratic term in the development indeed equals the asymptotic variance of the linear term, as

$$
\mathrm{Cov}\{\boldsymbol{\psi}_{f_0}(U), \mathrm{sign}(U - 1/2)\} = \boldsymbol{\mu}^+_{\boldsymbol{\psi}},
$$

which in turn implies

$$
\mathrm{Var}\left(\boldsymbol{\psi}_{f_0}\left(F_0(\varepsilon)\right) - \boldsymbol{\mu}^+_{\boldsymbol{\psi}}\mathrm{sign}(\varepsilon)\right) = \mathbf{V}_{\boldsymbol{\psi}} - \boldsymbol{\mu}_{\boldsymbol{\psi}}\boldsymbol{\mu}^T_{\boldsymbol{\psi}} - \boldsymbol{\mu}_{\boldsymbol{\psi}}\boldsymbol{\mu}^T_{\boldsymbol{\psi}} + \boldsymbol{\mu}_{\boldsymbol{\psi}}\boldsymbol{\mu}^T_{\boldsymbol{\psi}} = \mathbf{V}_{\boldsymbol{\psi}} - \boldsymbol{\mu}_{\boldsymbol{\psi}}\boldsymbol{\mu}^T_{\boldsymbol{\psi}}.
$$

Along the same lines, we find that the asymptotic covariance between the central sequence $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0,f_0}$ and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\psi}_{f_0}\left(F_0(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0))\right)-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)\right)$$

equals $-\bar{\mathbf{C}}\left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right)$. This completes the proof of the lemma. □

We may now prove the semiparametric efficiency of the sign-and-rank statistic $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f}$ in (3.12).

Proof of Theorem 4.1: Consider the parametric submodel as constructed in the previous lemma. From the convergence of local experiments to Gaussian shifts (which follows from the ULAN condition), we know that locally and asymptotically optimal inference for $\boldsymbol{\theta}$ in this model, should be based on the p-dimensional upper block of

$$\begin{bmatrix} \mathbf{I}_{f_0}(\boldsymbol{\theta}_0) & \bar{\mathbf{C}}\left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right) \\ \left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right)\bar{\mathbf{C}}^T & \mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\end{bmatrix}^{-1}\begin{pmatrix}\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0,f_0} \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\psi}_{f_0}(F_0(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)))-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)\right)\end{pmatrix}$$

(i.e., the components corresponding to $\boldsymbol{\theta}$). Using the classical formula for partitioned inverses (see, e.g., Magnus and Neudecker, 1988, page 11), this p-dimensional upper block is

$$\left(\mathbf{I}_{f_0}(\boldsymbol{\theta}_0)-\bar{\mathbf{C}}\left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right)\bar{\mathbf{C}}^T\right)^{-1}\left\{\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}_0,f_0}-\bar{\mathbf{C}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{\psi}_{f_0}(F_0(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)))-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\mathrm{sign}\left(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)\right)\right\}.$$

From Theorem 3.1 we see that this equals

$$\left(\mathbf{I}_{f_0}(\boldsymbol{\theta}_0)-\bar{\mathbf{C}}\left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right)\bar{\mathbf{C}}^T\right)^{-1}\ \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f}+o_{\mathrm{P}}(1),$$

with $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f}$ defined in (3.12). Observe also that, under $\mathbb{P}^{(n)}_{\boldsymbol{\theta}_0,f_0}$,

$$\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\boldsymbol{\theta},f}\xrightarrow{\mathcal{L}}\mathcal{N}(\mathbf{0},\mathbf{I}_{f_0}(\boldsymbol{\theta}_0)-\bar{\mathbf{C}}\left(\mathbf{V}_{\boldsymbol{\psi}}-\boldsymbol{\mu}^+_{\boldsymbol{\psi}}\boldsymbol{\mu}^{+T}_{\boldsymbol{\psi}}\right)\bar{\mathbf{C}}^T),$$

which completes the proof. □

References

[1] Bhattacharya, P.K. and A.K. Gangopadhyay (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics* **18**, 1400–1415.

[2] Bickel, P.J. (1982). On adaptive estimation. *Annals of Statistics* **10**, 647–671.

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

[4] Drost, F.C., C.A.J. Klaassen, and B.J.M. Werker (1997). Adaptive estimation in time-series models. *Annals of Statistics* **25**, 786–818.

[5] Hájek, J. and Z. Šidák (1967). *Theory of Rank Tests*. Academic Press, New York.

[6] Hallin, M., and M.L. Puri (1994). Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.* **50**, 175-237. .

[7] Hallin, M., C. Vermandele, and B.J.M. Werker (2003). Serial and non-Serial Sign-and-Rank Statistics: Asymptotic Representation and Asymptotic Normality. Working paper, ISRO, Université libre de Bruxelles.

[8] Hallin, M. and B.J.M. Werker (2003). Semiparametric Efficiency, Distribution-Freeness, and Invariance. *Bernoulli* **9**, 137–165.

[9] Han, A.K. (1987). Non-parametric analysis of a generalized regression model. *Journal of Econometrics* **35**, 303–316.

[10] Horowitz, J.L. and V.G. Spokoiny (2002). An adaptive, rate-optimal test of linearity for median regression models. *Journal of the American Statistical Association* **97**, 822–835.

[11] Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* **11**, 818–887.

[12] Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* **46**, 33–50.

[13] Kreiss, J.P. (1987). On adaptive estimation in stationary ARMA processes. *Annals of Statistics* **15**, 112–133.

[14] Le Cam, L. and G.L. Yang (1990). *Asymptotics in Statistics; Some Basic Concepts.* Springer-Verlag, New York.

[15] Lehmann, E.L. (1986). *Testing Statistical Hypotheses.* 2nd edition, John Wiley, New York.

[16] Magnus, J.R. and H. Neudecker (1988). *Matrix Differential Calculus.* Wiley, New York.

[17] Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5**, 99–135.

[18] Sherman, R.P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123–137.

[19] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

[20] Stute, W. (1986). Conditional empirical processes. *Annals of Statistics* **14**, 638-647.

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

[22] van Eeden, C. (1972). An analogue, for signed-rank statistics, for Jurečková's asymptotic linearity theorem for rank statistics. *Annals of Mathematical Statistics* **43**, 791–802.

[23] Zhou, Y. and H. Liang (2000). Asymptotic normality for L_1 norm kernel estimator of conditional median under alpha-mixing dependence. *Journal of Multivariate Statistics* **73**, 136–154.

[24] Zhou, Y. and H. Liang (2003). Asymptotic properties for L_1 norm kernel estimator of conditional median under dependence. *Journal of Nonparametric Statistics* **15**, 205–219.