Bernoulli **9**(1), 2003, 137–165

Semi-parametric efficiency, distribution-freeness and invariance

MARC HALLIN[1] and BAS J.M. WERKER[2]

[1]*I.S.R.O., Université Libre de Bruxelles, CP 210, B-1050, Bruxelles, Belgium. E-mail: mhallin@ulb.ac.be*

[2]*Econometrics Group and Finance Group, CentER, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands. E-mail: Werker@TilburgUniversity.nl*

Semi-parametric models typically involve a finite-dimensional parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, along with an infinite-dimensional nuisance parameter f. Quite often, the submodels corresponding to a fixed value of $\boldsymbol{\theta}$ possess a group structure that induces a maximal invariant σ-field $\mathcal{B}(\boldsymbol{\theta})$. In classical examples, where f denotes the density of some independent and identically distributed innovations, $\mathcal{B}(\boldsymbol{\theta})$ is the σ-field generated by the ranks of the residuals associated with the parameter value $\boldsymbol{\theta}$. It is shown that semi-parametrically efficient distribution-free inference procedures can generally be constructed from parametrically optimal ones by conditioning on $\mathcal{B}(\boldsymbol{\theta})$; this implies, for instance, that semi-parametric efficiency (at given $\boldsymbol{\theta}$ and f) can be attained by means of rank-based methods. The same procedures, when combined with a consistent estimation of the underlying nuisance density f, yield conditionally distribution-free semi-parametrically efficient inference methods, for example, semi-parametrically efficient permutation tests. Remarkably, this is achieved without any explicit tangent space or efficient score computations, and without any sample-splitting device. By means of several examples, including both i.i.d. and time-series models, we show how these results apply in models for which rank-based inference or permutation tests have so far seldom been considered.

Keywords: adaptiveness; distribution-freeness; local asymptotic normality; ranks

1. Introduction

Semi-parametric models typically involve a finite-dimensional parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$ of interest, along with an infinite-dimensional nuisance parameter f – often the unspecified density of some white noise underlying the data generating process. Classical examples include the one-, two-, and K-sample location and scale models, linear models with independent autoregressive moving average (ARMA) or conditional heteroscedastic (ARCH) type errors, linear as well as nonlinear time series models. The fact that f, in the semi-parametric model, remains unspecified induces, in general, a loss of efficiency with respect to the parametric situation. This loss is essentially formal, though, since parametric models seldom can be trusted to provide a correct description of the situation at hand: semi-parametric efficiency, as a rule, is the best realistic objective statisticians should contemplate. While recognizing the practical importance of a semi-parametric approach to statistical problems, much of the literature and a great deal of everyday practice still avoids addressing the specific features of semi-parametric models, contenting itself with the

1350–7265

fact that traditional parametric approaches remain valid in the semi-parametric setting. This is the case for the so-called *pseudo*- or *quasi*-likelihood methods, where the parametric procedures associated with some predetermined f_0 (mostly Gaussian or exponential likelihood procedures) are imported, without much change, into the semi-parametric context. Such procedures, if valid under unspecified f, reach parametric efficiency at Gaussian f only.

Efficiency, here and in what follows, is to be understood *á la* Le Cam, in a local and asymptotic sense, under the local asymptotic normality (LAN) structure (more precise definitions are given below). LAN induces bounds on the asymptotic performance of statistical procedures in parametric as well as semi-parametric models. The following somewhat heavy, terminology will be needed. We say that a method is *somewhere parametrically efficient* (parametrically efficient at f_0) if it is efficient in the parametric model induced by some f_0. A *parametrically efficient* method (a method that is parametrically efficient at any f) is clearly uniformly best. Such a method is usually called adaptive and generally does not exist. A similar terminology is used with respect to semi-parametric efficiency. Methods that are semi-parametrically efficient at some f_0 are called *somewhere semi-parametrically efficient*, and a method that is semi-parametrically efficient at any f is simply called *semi-parametrically efficient*. Finally, when the parametric and semi-parametric lower bounds coincide at some f_0, the model is called *somewhere adaptive* (adaptive at f_0). If the two bounds coincide for all f, the model is called *adaptive*.

Approaches that take into account the semi-parametric nature of the models at hand mainly belong to one of two types. Basically, the two possible attitudes are either to estimate the nuisance f in some way, or to take care of it by means of some adequate statistical principle. These two attitudes, combined with the LAN paradigm, are the starting points of two seemingly quite different strands of the statistical literature: efficient and adaptive inference on the one hand, where f is estimated from the sample; and permutation tests and rank-based inference on the other, where the influence of f is eliminated via either a conditioning or an invariance argument.

A comprehensive account of efficient and adaptive inference can be found in the monograph by Bickel *et al*. (1993) for semi-parametric models with independent observations. Semi-parametric time series models have been studied in a series of papers – Kreiss (1987; 1990), Jeganathan (1995; 1997), Drost *et al*. (1997), Koul and Schick (1997), Schick and Wefelmeyer (2000), and Müller and Wefelmeyer (2001) to name but a few. The basic idea in this literature is to estimate the underlying f, and to consider the so-called (locally and asymptotically, at fixed f and $\boldsymbol{\theta}$) 'least favourable parametric submodel' of the full semi-parametric model. The Le Cam theory allows *parametrically efficient* inference procedures to be constructed for such submodels. If, for the original semi-parametric model, inference procedures can be constructed which are (locally and asymptotically, either at some f or over some class $\mathcal{C}$ of f values) equivalent to these *submodel-parametrically efficient* ones, then they can be considered semi-parametrically efficient (either at some f or over some class $\mathcal{C}$ of f values). This approach – let us call it the *tangent space* approach – thus involves reducing, locally and asymptotically, the semi-parametric problem to a simpler parametric one, through the least favourable submodel argument. In general, the resulting computations are non-trivial.

The second strand of literature addresses the same problem of constructing inference procedures in semi-parametric models, and also involves reducing the original semi-parametric problem to a parametric one. The reduction, however, is based on an entirely different argument and is related to classical unbiasedness or invariance principles. The submodels corresponding to a fixed value of $\boldsymbol{\theta}$ do quite often either possess a group structure or allow for a complete sufficient statistic. The traditional invariance argument leads to the consideration of statistical procedures which are measurable with respect to the corresponding maximal invariant σ-field $\mathcal{B}(\boldsymbol{\theta})$. If, for instance, the semi-parametric nature of the model is due to the unspecified density f of some innovation process, then this approach typically leads to rank-based inference, where the ranks are those of the innovations. In a hypothesis testing context, the unbiasedness principle and Neyman structure argument lead to permutation tests.

Somewhere parametrically efficient rank-based procedures do not always exist. However, we show in Section 3 that somewhere *semi*-parametrically efficient procedures can always be found within the class of rank-based procedures. While the classical literature on rank tests so far has essentially been limited to linear models with independent observations (see Puri and Sen 1985; or Hájek *et al*. 1999), we establish this result in a much broader context, including most familiar time series models. Rank-based methods inherit all the desirable features resulting from their invariance properties, which entail, for example, distribution-freeness of test statistics. They are robust, and may even outperform, uniformly in f and $\boldsymbol{\theta}$, the more classical methods, based on *ad hoc*, mostly Gaussian, quasi-likelihoods (Chernoff and Savage 1958; Hallin 1994). Relevant papers in this direction are Hallin *et al*. (1985) and Hallin and Puri (1991; 1994). Similar properties also hold for permutation tests, of which rank tests are only a particular case.

A related approach to the same problem – albeit in a completely opposite direction – can be found in van der Vaart (1988), where sufficiency in the nonparametric fixed-$\boldsymbol{\theta}$ submodels is exploited rather than invariance (for independent observations).

The objective of this paper is to bring together the two approaches to semi-parametric inference just mentioned – the tangent space approach and the invariance approach. Earlier results in this direction have been obtained by, for example, Hájek (1962) (see Hájek *et al*. 1999, Section 8.5, for a discussion) and Beran (1974). These results, however, concentrate on specific cases, and are restricted to adaptive models (symmetric location, regression, etc.) with independent observations. As mentioned above, we establish that the usual way to construct rank-based inference procedures generally yields somewhere semi-parametrically efficient procedures, and this extends to a very general class of possibly nonadaptive time series models. Furthermore, we show that the benefits of both approaches can be cumulated, yielding semi-parametrically efficient and conditionally distribution-free inference procedures.

The basic results we obtain are described in Section 2 in a general and rather abstract context. These general results are specialized to (signed) rank-based inference in Section 3; in order to illustrate the general ideas, we employ the very simple example of a moving-average model of order one. However, we stress that our results are not limited to rank-based methods, and are valid whatever the underlying invariance or distribution-freeness structure of the fixed-$\boldsymbol{\theta}$-submodels, as long as the fixed-f-submodels satisfy the LAN property and some additional regularity conditions. In Section 3.4 we discuss the construction of semi-parametrically efficient conditionally distribution-free inference procedures, such as semi-parametrically efficient permutation tests. Finally, in Section 4, we particularize further to some specific examples, such as location, regression and autoregression models, and a simple ARCH model. We stress that the results are applicable to many further situations, such as ARMA models, threshold autoregressive models, other ARCH-type models and random coefficient autoregressive models, for which rank-based inference has so far seldom been considered.

2. General results

2.1. Local asymptotic normality

Consider a sequence of semi-parametric models

$$\mathcal{E}^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\varphi} : \boldsymbol{\theta} \in \Theta, \varphi \in \Phi\}), \qquad n \in \mathbb{N},$$

with Θ an open subset of $\mathbb{R}^k$ and Φ an arbitrary set. We will study this sequence of models in the neighbourhood of arbitrary, but for the rest of this section fixed, parameter values $(\boldsymbol{\theta}_0, \varphi_0) \in \Theta \times \Phi$. Write $\mathrm{P}^{(n)}_0$ for $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,\varphi_0}$, and denote expectations under $\mathrm{P}^{(n)}_0$ by $\mathrm{E}^{(n)}_0$. We assume throughout this paper that the sequence of parametric models obtained by fixing the nuisance parameter at $\varphi = \varphi_0$, that is,

$$\mathcal{E}^{(n)}_{\varphi_0} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}_{\varphi_0} = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\varphi_0} : \boldsymbol{\theta} \in \Theta\}), \qquad n \in \mathbb{N},$$

is locally asymptotically normal at $\boldsymbol{\theta}_0$, with central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)$ and Fisher information $\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0)$. To be precise, we have the following assumption.

Assumption A. *For any bounded sequence* $(\boldsymbol{\tau}_n)$ *in* $\mathbb{R}^k$, *we have*

$$\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}_n/\sqrt{n},\varphi_0}}{\mathrm{dP}^{(n)}_0} = \exp\left(\boldsymbol{\tau}_n^{\mathrm{T}}\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) - \frac{1}{2}\boldsymbol{\tau}_n^{\mathrm{T}}\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0)\boldsymbol{\tau}_n + r_n\right), \tag{2.1}$$

where $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0))$ *and* $r_n \xrightarrow{\mathrm{P}} 0$, *under* $\mathrm{P}^{(n)}_{\mathbf{0}}$, as $n \to \infty$.

2.2. Least favourable submodels

Though much of the semi-parametric terminology has been developed in the context of independent observations, it transposes, in a more or less obvious way, into a context of serially dependent observations; whenever possible, we will try to avoid formal redefinitions.

Semi-parametrically efficient inference about the parameter $\boldsymbol{\theta}$ in the sequence of semi-parametric models $(\mathcal{E}^{(n)})$ must be based on the so-called *efficient influence function*, which is obtained as the influence function (for inference about $\boldsymbol{\theta}$) of a parametric submodel $\mathcal{E}^{(n)}_{q_{\mathrm{lf}}}$ (of the form (2.2) described below) that includes the so-called 'least favourable direction'

$\mathbf{H}^{(n)}_{q_{\rm lf}}(\boldsymbol{\theta}, \varphi)$. For independently and identically distributed (i.i.d.) semi-parametric models, this most difficult direction is found by projecting the score for $\boldsymbol{\theta}$ onto the tangent space generated by the nuisance parameter (see, for example, Bickel *et al*. 1993). However, in our general set-up, we do not necessarily have a product structure for the experiments and, therefore, the definition of a tangent space concept is more delicate.

As usual, semi-parametric efficiency considerations are based on a local (and asymptotic) analysis of suitable parametric submodels of the full semi-parametric model $\mathcal{E}^{(n)}$. Let $\mathcal{Q}$ denote the set of all maps $q : (-1, 1)^k \to \Phi$ such that: (i) $q(\mathbf{0}) = \varphi_0$; and (ii) the sequence of parametric experiments $(\mathcal{E}^{(n)}_q, n \in \mathbb{N})$ defined by

$$\mathcal{E}^{(n)}_q(\varphi_0) := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}_q = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}, q(\eta)} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\eta} \in (-1, 1)^k\}) \tag{2.2}$$

is locally asymptotically normal (with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$) at $(\boldsymbol{\theta}_0, \mathbf{0})$. We write $(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\rm T}, \mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)^{\rm T})^{\rm T}$ for the corresponding central sequence, and

$$\mathbf{I}_q(\boldsymbol{\theta}_0, \varphi_0) := \begin{pmatrix} \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) & \mathbf{C}^{\rm T}_q(\boldsymbol{\theta}_0, \varphi_0) \\ \mathbf{C}_q(\boldsymbol{\theta}_0, \varphi_0) & \mathbf{I}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0) \end{pmatrix} \tag{2.3}$$

for the Fisher information matrix at $(\boldsymbol{\theta}_0, \mathbf{0})$.

Asymptotically efficient inference on $\boldsymbol{\theta}$ in this parametric submodel should be based on the so-called *influence function*, which is defined as the $\boldsymbol{\theta}$-component of

$$\mathbf{I}^{-1}_q(\boldsymbol{\theta}_0, \varphi_0)(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\rm T}, \mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)^{\rm T})^{\rm T},$$

taking (after elementary algebra) the form

$$(\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{C}^{\rm T}_q(\boldsymbol{\theta}_0, \varphi_0)\mathbf{I}^{-1}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0)\mathbf{C}_q(\boldsymbol{\theta}_0, \varphi_0))^{-1}\boldsymbol{\Delta}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0), \tag{2.4}$$

where

$$\boldsymbol{\Delta}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) := \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{C}^{\rm T}_q(\boldsymbol{\theta}_0, \varphi_0)\mathbf{I}^{-1}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0)\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0). \tag{2.5}$$

Intuitively, $\boldsymbol{\Delta}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$ is the residual of the regression of the $\boldsymbol{\theta}$-part $\boldsymbol{\Delta}^{(n)}$ of the central sequence with respect to the $\boldsymbol{\eta}$-part $\mathbf{H}^{(n)}_q$, using the asymptotic covariance matrix (2.3). And the classical theory of Gaussian inference tells us that, in the Gaussian shift experiment

$$\mathcal{G}_q(\boldsymbol{\theta}_0, \varphi_0) := (\mathbb{R}^{2k}, \mathcal{B}^{2k}, \{N(\mathbf{I}_q(\boldsymbol{\theta}_0, \varphi_0)(\boldsymbol{\tau}^{\rm T}, \boldsymbol{\rho}^{\rm T})^{\rm T}, \mathbf{I}_q(\boldsymbol{\theta}_0, \varphi_0)) : (\tau, \boldsymbol{\rho}) \in \mathbb{R}^{2k}\}),$$

optimal inference on τ, when $\boldsymbol{\rho}$ is unspecified, should be based on such a residual. In view of the convergence, in the Le Cam distance, of the local experiments

$$\mathcal{E}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}_q = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0+n^{-1/2}\boldsymbol{\tau}, q(n^{-1/2}\boldsymbol{\rho})} : (\boldsymbol{\tau}, \boldsymbol{\rho}) \in \mathbb{R}^{2k}\}), \qquad n \in \mathbb{N},$$

to $\mathcal{G}_q(\boldsymbol{\theta}_0, \varphi_0)$, the same holds, in a local and asymptotic sense, for inference on $\boldsymbol{\theta}$ in $\mathcal{E}^{(n)}_q(\varphi_0)$ when $\boldsymbol{\eta}$ remains unspecified. The residual Fisher information in the experiment characterized by $\boldsymbol{\Delta}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$ is $\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{C}^{\rm T}_q(\boldsymbol{\theta}_0, \varphi_0)\mathbf{I}^{-1}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0)\mathbf{C}_q(\boldsymbol{\theta}_0, \varphi_0)$. The loss of $\boldsymbol{\theta}$-information due to the non-specification of $\boldsymbol{\eta}$ in $\mathcal{E}^{(n)}_q(\varphi_0)$ is thus $\mathbf{C}^{\rm T}_q(\boldsymbol{\theta}_0, \varphi_0)\mathbf{I}^{-1}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0)\mathbf{C}_q(\boldsymbol{\theta}_0, \varphi_0)$.

Now, the L_2 projection (2.5) takes care, in a local and asymptotically optimal way, of the influence on $\boldsymbol{\Delta}^{(n)}$ of local perturbations of φ_0 in the parametric subexperiment characterized by q. The question is: does there exist a $q_{\rm lf} \in \mathcal{Q}$ such that (a version of) the corresponding

$\mathbf{\Delta}^{(n)}_{q_{\rm lf}}$ is asymptotically orthogonal to (some version of) *any* $\mathbf{H}_q(\boldsymbol{\theta}_0, \varphi_0)$, $q \in \mathcal{Q}$? If such a $q_{\rm lf}$ exists, the corresponding loss of information clearly will be maximal. Thus, $\mathbf{H}_{q_{\rm lf}}$ quite naturally can be called a *least favourable* direction, and the corresponding $\mathbf{\Delta}^{(n)}_{q_{\rm lf}}(\boldsymbol{\theta}_0, \varphi_0)$ a *semi-parametrically efficient* (at $(\boldsymbol{\theta}_0, \varphi_0)$) central sequence.

2.3. Distribution-free sub-σ-algebras

Turning to the fixed-$\boldsymbol{\theta}$ subexperiments, denote by

$$\mathcal{E}^{(n)}_{\boldsymbol{\theta}_0} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)}_{\boldsymbol{\theta}_0} = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,\varphi} : \varphi \in \Phi\}), \qquad n \in \mathbb{N},$$

the sequence of nonparametric experiments corresponding to a fixed value $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. The present paper focuses on the situation where these experiments $(\mathcal{E}^{(n)}_{\boldsymbol{\theta}_0}, n \in \mathbb{N})$ allow for some distribution-freeness or invariance structure. More precisely, assume the existence, for all $\boldsymbol{\theta}_0$, of a sequence $(\mathcal{B}^{(n)}(\boldsymbol{\theta}_0), n \in \mathbb{N})$ of σ-fields, with $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0) \subset \mathcal{A}^{(n)}$, such that the restriction of $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,\varphi}$ to $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$, which we denote by $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,\varphi|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)}$, does not depend on $\varphi \in \Phi$. This distribution-freeness of $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$ generally follows from some invariance property, under which $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$ is generated by the orbits of some group acting on $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)})$ – see Section 3 for examples concerned with ranks. Note that the sequences $\mathcal{B}^{(n)}(\boldsymbol{\theta})$ in general strongly depend on the parameter $\boldsymbol{\theta}$ but not on the nuisance φ.

The following proposition constitutes an essential ingredient of this paper. It states that, whenever $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,\varphi|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)}$ does not depend on $\varphi \in \Phi$, the conditional (upon $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$) expectation of the $\boldsymbol{\eta}$-part $\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$ of the central sequence associated with experiment (2.2) converges to zero in the L_1 norm, under $\mathrm{P}^{(n)}_{\mathbf{0}}$, as $n \to \infty$.

Proposition 2.1. *Fix $q \in \mathcal{Q}$, and consider the sequence of experiments $(\mathcal{E}^{(n)}_q)$ as defined in* (2.2)*. Let $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0) \subset \mathcal{A}^{(n)}$ be a sequence of σ-fields such that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,q(\boldsymbol{\eta})|\mathcal{B}(\boldsymbol{\theta}_0)}$ does not depend on $\boldsymbol{\eta}$. Moreover, assume that the $\boldsymbol{\eta}$-part $\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$ of the central sequence corresponding to $\mathcal{E}^{(n)}_q$ is uniformly integrable under $\mathrm{P}^{(n)}_{\mathbf{0}}$. Then*

$$\mathrm{E}^{(n)}_{\mathbf{0}}\{\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)\} = o_{L_1}(1),$$

under $\mathrm{P}^{(n)}_{\mathbf{0}}$, as $n \to \infty$.

Note that the assumption on the uniform integrability of $\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$ does not induce any restriction on $\mathcal{E}^{(n)}$. Indeed, uniformly integrable versions of the central sequences of locally asymptotically normal experiments always exist. This interesting property of locally asymptotically normal models appears not to be so well known, and we could not find any explicit reference to it in the literature. Therefore, we turn it into a formal statement and prove the following lemma.

Lemma 2.2. *Denote by $\mathbf{\Delta}^{(n)}_{\boldsymbol{\theta}}$ an arbitrary central sequence in some locally asymptotically normal sequence of experiments $\mathcal{E}^{(n)}$, with parameter $\boldsymbol{\theta}$ and probability distributions $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$. Then there exists a sequence $\mathbf{\Delta}^{(n)\circ}_{\boldsymbol{\theta}} = (\mathbf{\Delta}^{(n)\circ}_{\boldsymbol{\theta};i})$, $n \in \mathbb{N}$ such that:*

(i) *for any* $p \in (-1, \infty)$, *the sequences* $|\boldsymbol{\Delta}^{(n)\circ}_{\boldsymbol{\theta};i}|^p$, $n \in \mathbb{N}$, $i = 1, \ldots, k$, *are uniformly integrable;*

(ii) $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \boldsymbol{\Delta}^{(n)\circ}_{\boldsymbol{\theta}} = o_{\mathrm{P}}(1)$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ *as* $n \to \infty$, *so that, for any* $p \in (-1, \infty)$, $\boldsymbol{\Delta}^{(n)\circ}_{\boldsymbol{\theta}}$ *constitutes a uniformly pth order integrable version of the central sequence.*

Proof. Denote by $(I_{ij}(\boldsymbol{\theta}))$ the information matrix for $\mathcal{E}^{(n)}$, by Φ the standard normal distribution function, and by $F^{(n)}_{\boldsymbol{\theta};i}$ the marginal distribution function, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$, of the ith component $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};i}$ of $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}$. With no loss of generality, we may assume that $F^{(n)}_{\boldsymbol{\theta};i}$ is continuous (if necessary, $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};i}$ can be smoothed in an asymptotically negligible way). Put $\boldsymbol{\Delta}^{(n)\circ}_{\boldsymbol{\theta};i} := I^{1/2}_{ii}(\boldsymbol{\theta})\Phi^{-1}(F^{(n)}_{\boldsymbol{\theta};i}(\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};i}))$: clearly, the distribution of $\boldsymbol{\Delta}^{(n)\circ}_{\boldsymbol{\theta}}$ does not depend on n, and coincides with the asymptotic normal distribution of $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};i}$. Due to the convergence in distribution of $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};i}$ and the continuous mapping theorem (since Φ^{-1} is continuous), $\boldsymbol{\Delta}^{(n)\circ}_{\boldsymbol{\theta};i} - \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta};i}$ is $o_{\mathrm{P}}(1)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$. The lemma then follows from the fact that Gaussian random variables have finite absolute moments of order p for any $p \in (-1, \infty)$. □

In order to prove Proposition 2.1, we need the following simple lemma.

Lemma 2.3. *Consider the sequence of experiments* $(\mathcal{E}^{(n)}_q)$ *as in* (2.2). *Write*

$$\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0, q(\rho/\sqrt{n})}}{\mathrm{dP}^{(n)}_0} = \exp\left(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) - \frac{1}{2}\boldsymbol{\rho}^{\mathrm{T}}\mathbf{I}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0)\boldsymbol{\rho} + r^{(n)}_q(\boldsymbol{\rho})\right),$$

where, under $\mathrm{P}^{(n)}_0$, $\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$ *is uniformly integrable,* $\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0))$, *and* $r^{(n)}_q(\boldsymbol{\rho}) \xrightarrow{\mathrm{P}} 0$. *Then*

$$\lim_{\boldsymbol{\rho}\to\mathbf{0}} \limsup_{n\to\infty} \mathrm{E}^{(n)}_0[\|\boldsymbol{\rho}\|^{-1}|\exp(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) + r^{(n)}_q(\boldsymbol{\rho})) - 1 - \boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)|] = 0.$$

Proof. The contiguity of the sequences $(\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0, q(\rho/\sqrt{n})})$ and $(\mathrm{P}^{(n)}_{\mathbf{0}})$ (which is an immediate consequence of LAN), together with the uniform integrability of $(\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0))$, implies that

$$\exp(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) + r^{(n)}_q(\boldsymbol{\rho})) - 1 - \boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)$$

is uniformly integrable under $\mathrm{P}^{(n)}_0$. Hence, as $n \to \infty$,

$$\begin{aligned} &\mathrm{E}^{(n)}_0[\|\boldsymbol{\rho}\|^{-1}|\exp(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0) + r^{(n)}_q(\boldsymbol{\rho})) - 1 - \boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}^{(n)}_q(\boldsymbol{\theta}_0, \varphi_0)|] \\ &\qquad \to \mathrm{E}[\|\boldsymbol{\rho}\|^{-1}|\exp(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}) - 1 - \boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}|], \end{aligned}$$

where $\mathbf{H} \sim N(\mathbf{0}, \mathbf{I}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0))$. Letting $\boldsymbol{\rho} \to \mathbf{0}$ yields the desired result. □

Proof of Proposition 2.2. Invariance of $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0, q(\boldsymbol{\eta})|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)}$ with respect to $\boldsymbol{\eta}$ implies that, for all $\boldsymbol{\rho} \in (-1, 1)^k$,

$$\mathrm{E}^{(n)}_0\left\{\frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0, q(\rho/\sqrt{n})}}{\mathrm{dP}^{(n)}_0}\middle|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)\right\} = 1.$$

Therefore, we may write

$$\|\boldsymbol{\rho}\|^{-1}\mathrm{E}_0^{(n)}\{\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}_q^{(n)}(\boldsymbol{\theta}_0, \varphi_0)|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)\}$$

$$= \|\boldsymbol{\rho}\|^{-1}\mathrm{E}_0^{(n)}\Big\{\exp\Big(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}_q^{(n)}(\boldsymbol{\theta}_0, \varphi_0) - \frac{1}{2}\boldsymbol{\rho}^{\mathrm{T}}\mathbf{I}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0)\boldsymbol{\rho} + r_q^{(n)}(\boldsymbol{\rho})\Big) - 1$$

$$- \boldsymbol{\rho}^{\mathrm{T}}\mathbf{H}_q^{(n)}(\boldsymbol{\theta}_0, \varphi_0)|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)\Big\}.$$

Letting $n \to \infty$ and then $\boldsymbol{\rho} \to \mathbf{0}$, Lemma 2.3 gives the desired result. □

2.4. Least favourable submodels

Proposition 2.1 implies that, for any $q \in \mathcal{Q}$, the $\boldsymbol{\eta}$-part $\mathbf{H}_q$ if the central sequence is asymptotically orthogonal to $\mathcal{B}^{(n)}(\boldsymbol{\theta})$. This suggests that

$$\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi) := \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi) - \mathrm{E}_{\boldsymbol{\theta},\varphi}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\} \tag{2.6}$$

might be a reasonable candidate as a least favourable direction,

$$\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi) := \mathrm{E}_{\boldsymbol{\theta},\varphi}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\} \tag{2.7}$$

being the corresponding semi-parametrically efficient (at φ) central sequence. This suggested *semi-parametrically efficient central sequence* thus results from reducing the information available in the original experiment by conditioning on $\mathcal{B}^{(n)}(\boldsymbol{\theta})$. There is no guarantee that such conditioning preserves LAN nor, for that matter, any other properties of the original experiment. However, see Le Cam and Yang (1988) for some positive results in this direction.

We therefore establish a sufficient condition for $\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi)$ to be least favourable. This condition is extremely mild, as it simply requires $\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi)$ to correspond to some mapping $q \in \mathcal{Q}$ (for a point estimation version of the same condition in the i.i.d. context, see Bickel *et al*. 1993, Section 3.1; or Klaassen 1987, Theorem 3.1):

(LF1) $\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi_0)$ defined in (2.6) can be obtained as the $\boldsymbol{\eta}$-part of the central sequence $(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi_0)^{\mathrm{T}}, \mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi_0)^{\mathrm{T}})^{\mathrm{T}}$ of some parametric submodel $\mathcal{E}_{\mathrm{lf}}^{(n)}(\varphi_0)$ of the form (2.2).

Proposition 2.4. *Assume that condition* (LF1) *holds. Then* $\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi_0)$ *is least favourable, and* $\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi_0)$ *is a semi-parametrically efficient (at* φ_0*) central sequence in the sense of Section* 2.2.

Proof. For arbitrary $q \in \mathcal{Q}$, let $\mathbf{H}_q^{(n)}(\boldsymbol{\theta}, \varphi_0)$ be a uniformly integrable version of the $\boldsymbol{\eta}$-part of the corresponding central sequence (such a version exists in view of Lemma 2.2). It follows from Proposition 2.1 that a further version of the same central sequence, $\mathbf{H}_{q\perp}^{(n)}(\boldsymbol{\theta}, \varphi_0)$, say, is orthogonal to $\mathcal{B}^{(n)}(\boldsymbol{\theta})$, hence to $\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi)$ defined in (2.7) for all n. Since this holds for any

q, and since, under condition (LF1), $\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi)$ is obtained from some parametric submodel $\mathcal{E}_{\mathrm{lf}}^{(n)}(\varphi_0)$ of the form (2.2), the result follows. □

Proposition 2.4 deals with a fixed φ value, hence with semi-parametric efficiency at given φ. If efficiency is to be attained over some class $\mathcal{C}$ of values of φ, an additional requirement is the following:

(LF2) For all $\varphi \in \mathcal{C} \subseteq \Phi$, there exists a version of the influence function (2.4) for $\boldsymbol{\theta}$ in $\mathcal{E}_{\mathrm{lf}}^{(n)}(\varphi)$ that does not depend on φ.

In general, condition (LF2) will be satisfied when a version $\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta})$ that does not depend on $\varphi \in \mathcal{C}$ of the efficient central sequence $\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi) := \mathrm{E}_{\boldsymbol{\theta},\varphi}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\}$ can be obtained, along with a consistent (for any $\varphi \in \mathcal{C} \subseteq \Phi$) estimator of its covariance matrix.

Under condition (LF2), semi-parametrically efficient inference over $\mathcal{C}$ clearly can be based on $\boldsymbol{\Delta}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta})$.

Assuming that condition (LF1) holds, the information matrix in $\mathcal{E}_{\mathrm{lf}}^{(n)}$ has the form

$$\begin{pmatrix} \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) & \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0) \\ \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0) & \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0) \end{pmatrix}, \tag{2.8}$$

where $\mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0)$ is the variance of the limiting distribution of $\mathrm{E}_{\boldsymbol{\theta},\varphi}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\}$. Up to $o_{\mathrm{P}}(1)$ terms, the efficient influence function is thus the $\boldsymbol{\theta}$-part of

$$\begin{pmatrix} \mathbf{I}(\boldsymbol{\theta}, \varphi) & \mathbf{I}(\boldsymbol{\theta}, \varphi) - \mathbf{I}^*(\boldsymbol{\theta}, \varphi) \\ \mathbf{I}(\boldsymbol{\theta}, \varphi) - \mathbf{I}^*(\boldsymbol{\theta}, \varphi) & \mathbf{I}(\boldsymbol{\theta}, \varphi) - \mathbf{I}^*(\boldsymbol{\theta}, \varphi) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi) \\ \mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, \varphi) \end{pmatrix} \tag{2.9}$$

which, still under (LF1), admits the asymptotic representation.

$$\mathbf{I}^*(\boldsymbol{\theta}, \varphi)^{-1}\mathrm{E}_{\boldsymbol{\theta},\varphi}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, \varphi)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\} \ + \ o_{\mathrm{P}}(1) \tag{2.10}$$

as $n \to \infty$ under $P_{\theta,\varphi}^{(n)}$, for any $(\theta\varphi) \in \Theta \times \boldsymbol{\Phi}$.
If the information matrix in (2.9) is not invertible (this happens, for instance, in adaptive models, where $\mathbf{I}(\boldsymbol{\theta}, \varphi) = \mathbf{I}^*(\boldsymbol{\theta}, \varphi)$), its inverse should be replaced by any generalized inverse.

In practice, conditions (LF1) and (LF2) cannot easily be checked for. We will not discuss the conditions under which condition (LF2) can be satisfied in the general set-up of this section. We only do so, in Section 3.4, for the specific case of rank-based inference. However, we do give a precise statement of a set of assumptions that are sufficient for condition (LF1) to hold and that can be easily checked for in a given model. As before, fix $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ and $\varphi_0 \in \boldsymbol{\Phi}$, and consider the following set of assumptions.

Assumption B.

(i) *Let* $\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)$ *defined in* (2.6) *be such that* $(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}}, \mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}})^{\mathrm{T}}$ *is asymptotically normal* (*automatically, with mean zero, and covariance matrix* (2.8)) *under* $\mathrm{P}_{\mathbf{0}}^{(n)}$, *as* $n \to \infty$.

(ii) *There exists a function* $q_{\mathrm{lf}} : (-1, 1)^k \to \Phi$ *such that, for any sequence* $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + O(1/\sqrt{n})$, *the sequence of experiments*

$$(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, q_{\mathrm{lf}}(\boldsymbol{\eta})} : \boldsymbol{\eta} \in (-1, 1)^k\})$$

is locally asymptotically normal at $\boldsymbol{\eta} = \mathbf{0}$ *with central sequence* $\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_n, \varphi_0)$ *and Fisher information matrix* $\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0)$.

(iii) *The sequence* $\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}, \varphi_0)$ *satisfies a local asymptotic linearity property, in the sense that, for any bounded sequence* $\boldsymbol{\tau}_n$ *in* $\mathbb{R}^k$*, we have*

$$\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0 + \boldsymbol{\tau}_n/\sqrt{n}, \varphi_0) = [\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0)]\boldsymbol{\tau}_n + o_{\mathrm{P}}(1)$$

under $\mathrm{P}^{(n)}_{\mathbf{0}}$*, as* $n \to \infty$.

Under Assumptions A and B, we have a LAN property (jointly) in $(\boldsymbol{\theta}, \boldsymbol{\eta})$ at $(\boldsymbol{\theta}_0, \mathbf{0})$ for the parametric subexperiments $\mathcal{E}^{(n)}_{q_{\mathrm{lf}}}(\varphi_0)$ of $\mathcal{E}^{(n)}$ given by (2.2). Formally, we prove that q_{lf} belongs to $\mathcal{Q}$. Note that Assumption B(ii) does not directly assert that $q_{\mathrm{lf}} \in \mathcal{Q}$, because it requires the quadratic expansion of the LAN condition only with respect to $\boldsymbol{\eta}$ and not jointly in $(\boldsymbol{\theta}, \boldsymbol{\eta})$.

Proposition 2.5. *Suppose that Assumptions A and B are satisfied. Then the model*

$$(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}, q_{\mathrm{lf}}(\boldsymbol{\eta})} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\eta} \in (-1, 1)^k\}),$$

is locally asymptotically normal at $(\boldsymbol{\theta}_0, \mathbf{0})$ *with central sequence* $(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}}, \mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}})^{\mathrm{T}}$ *and Fisher information matrix* (2.8)*, and condition* (LF1) *is satisfied.*

Proof. Let $\boldsymbol{\tau}_n$ and $\boldsymbol{\rho}_n$ be bounded sequences in $\mathbb{R}^k$. Observe, using Assumption B(ii), the contiguity of $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}_n/\sqrt{n},\varphi_0}$ and $\mathrm{P}^{(n)}_{\mathbf{0}}$ which follows from Assumption A via Le Cam's first lemma, and Assumption B(iii), that

$$\begin{aligned}
\log \frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}_n/\sqrt{n}, q(\boldsymbol{\rho}_n/\sqrt{n})}}{\mathrm{dP}^{(n)}_{\mathbf{0}}} &= \log \frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}_n/\sqrt{n}, q(\boldsymbol{\rho}_n/\sqrt{n})}}{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}_n/\sqrt{n},\varphi_0}} + \log \frac{\mathrm{dP}^{(n)}_{\boldsymbol{\theta}_0+\boldsymbol{\tau}_n/\sqrt{n},\varphi_0}}{\mathrm{dP}^{(n)}_{\mathbf{0}}} \\
&= \boldsymbol{\rho}_n^{\mathrm{T}}\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0 + \boldsymbol{\tau}_n/\sqrt{n}, \varphi_0) - \frac{1}{2}\boldsymbol{\rho}_n^{\mathrm{T}}[\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0)]\boldsymbol{\rho}_n \\
&\quad + \boldsymbol{\tau}_n^{\mathrm{T}}\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) - \frac{1}{2}\boldsymbol{\tau}_n^{\mathrm{T}}\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0)\boldsymbol{\tau}_n + o_{\pi}(1) \\
&= \boldsymbol{\rho}_n^{\mathrm{T}}\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0, \varphi_0) + \boldsymbol{\tau}_n^{\mathrm{T}}\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) - \boldsymbol{\rho}_n^{\mathrm{T}}[\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0)]\boldsymbol{\tau}_n \\
&\quad - \frac{1}{2}\boldsymbol{\rho}_n^{\mathrm{T}}[\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) - \mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0)]\boldsymbol{\rho}_n - \frac{1}{2}\boldsymbol{\tau}_n^{\mathrm{T}}\mathbf{I}(\boldsymbol{\theta}_0, \varphi_0)\boldsymbol{\tau}_n + o_{\mathrm{P}}(1).
\end{aligned}$$

Assumption B(i) asserts the asymptotic normality of the central sequence, and this completes the proof. □

2.5. Adaptiveness

An important property of some semi-parametric models is adaptiveness. Adaptiveness occurs if inference problems are no more difficult (locally and asymptotically) in the semi-parametric model than in the underlying parametric model. In the above notation, adaptiveness occurs at $(\boldsymbol{\theta}_0, \varphi_0)$ if $\mathbf{I}^*(\boldsymbol{\theta}_0, \varphi_0) = \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0)$ – hence if, under $\mathrm{P}_{\mathbf{0}}^{(n)}$,

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) = \mathrm{E}_{\mathbf{0}}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)\} + o_{\mathrm{P}}(1), \tag{2.11}$$

or, for that matter, if there exists any $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$-measurable version of the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)$. These statements are made rigorous in the following proposition.

Proposition 2.6. *Fix $q \in \mathcal{Q}$, that is, let $(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \{\mathrm{P}_{\boldsymbol{\theta},q(\boldsymbol{\eta})}^{(n)} : (\boldsymbol{\theta}, \boldsymbol{\eta}) \in \boldsymbol{\Theta} \times (-1, 1)^k\})$ denote an arbitrary submodel of the semi-parametric model that satisfies the LAN property at $(\boldsymbol{\theta}_0, \mathbf{0})$, with central sequence $(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}}, \mathbf{H}_q^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}})^{\mathrm{T}}$.*

If a $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$-measurable version $\boldsymbol{\Delta}_{\mathcal{B}}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)$ of the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)$ exists, then, under $\mathrm{P}_{\mathbf{0}}^{(n)}$ and as $n \to \infty$, we have, for any $q \in \mathcal{Q}$,

$$\begin{pmatrix} \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0) \\ \mathbf{H}_q^{(n)}(\boldsymbol{\theta}_0, \varphi_0) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \begin{pmatrix} \mathbf{I}(\boldsymbol{\theta}_0, \varphi_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\mathbf{H}_q}(\boldsymbol{\theta}_0, \varphi_0) \end{pmatrix}\right),$$

so that the model $\mathcal{E}^{(n)}$ is adaptive at $(\boldsymbol{\theta}_0, \varphi_0)$.

Proof. The fact that $\mathrm{P}_{\boldsymbol{\theta}_0,q(\boldsymbol{\eta})}^{(n)}|\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$ does not depend on $\boldsymbol{\eta} \in (-1, 1)^k$ implies that

$$\mathcal{L}_{\boldsymbol{\theta}_0,q(\boldsymbol{\eta})}(\boldsymbol{\Delta}_{\mathcal{B}}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)) = \mathcal{L}_{\mathbf{0}}(\boldsymbol{\Delta}_{\mathcal{B}}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)),$$

so that the limiting distribution of $(\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}}, \mathbf{H}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)^{\mathrm{T}})^{\mathrm{T}}$ is the same under $\boldsymbol{\eta} = \mathbf{0}$ as under local alternatives of the form $(\boldsymbol{\theta}_0, \boldsymbol{\eta}_n := \mathbf{0} + O(1/\sqrt{n}))$. Le Cam's third lemma completes the proof. □

Remark 2.1. Under condition (LF1), Proposition 2.6 can be reinforced: adaptiveness holds (at $(\boldsymbol{\theta}_0, \varphi_0)$) if and only if (2.11) holds, that is, if and only if a $\mathcal{B}^{(n)}(\boldsymbol{\theta}_0)$-measurable version of the central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, \varphi_0)$ exists.

3. Semi-parametrically efficient rank-based inference

3.1. White noise, invariance and ranks

In this section, we specialize the general results of Section 2 to models where the randomness is due to the presence of some underlying white noise – a sequence of i.i.d. random variables $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ with unspecified probability density *f*. To be precise, we consider a sequence of semi-parametric models

$$\mathcal{E}^{(n)} = (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} = \{\mathrm{P}_{\boldsymbol{\theta},f}^{(n)} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, f \in \mathcal{F}\}), \tag{3.1}$$

with $\boldsymbol{\Theta}$ an open subset of $\mathbb{R}^k$ and $\mathcal{F}$ a set of densities. We suppose again that the parametric model obtained by fixing $f \in \mathcal{F}$ is locally asymptotically normal. For the sake of simplicity, we restrict ourselves to the case where $\mathcal{F}$ is a subset of the class $\mathcal{F}_0$ of all non-vanishing densities over the real line. Other classes of densities can be considered with more or less obvious changes, such as the class of non-vanishing densities which are symmetric with respect to the origin (yielding signed ranks instead of ranks) and the class of non-vanishing densities with median zero (yielding signs *and* ranks).

The role of this underlying white noise is described by the following pair of assumptions.

Assumption I.

(i) *For all $f \in \mathcal{F}$, the parametric model $\mathcal{E}_f^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_f^{(n)} = \{\mathrm{P}_{\boldsymbol{\theta},f}^{(n)} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\})$ is locally asymptotically normal in $\boldsymbol{\theta}$ at all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, with central sequence $\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)$ of the form*

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f) := \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-q}^{(n)}(\boldsymbol{\theta})), \tag{3.2}$$

where the function $\mathbf{J}_f : \mathbb{R}^{p+1} \to \mathbb{R}^k$ is allowed to depend on f and $\boldsymbol{\theta}$ (for the sake of simplicity, however, we avoid the notation $\mathbf{J}_{\boldsymbol{\theta},f}$); the residuals $\varepsilon_t^{(n)}(\boldsymbol{\theta})$, $t = 1, \ldots, n$, *are an invertible function of the observations such that $\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})$ under $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$ are i.i.d. with density f. The Fisher information matrix corresponding to this LAN condition is denoted as $\mathbf{I}_f(\boldsymbol{\theta})$.*

(ii) *The class $\mathcal{F}$ is a subset of the class $\mathcal{F}_0$ of all densities f (over the real line) such that $f(x) > 0$, $x \in \mathbb{R}$.*

As an illustration, we consider a simple MA(1) process as an example.

Example 3.1. Denote by $Y_1^{(n)}, \ldots, Y_n^{(n)}$ a finite realization of the MA(1) process characterized by

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \qquad t \in \mathbb{N},$$

where $\theta \in (-1, 1)$, $\{\varepsilon_t, t \geqslant 1\}$ is a process of independent random variables with common density *f*, and ε_0 is a fixed starting value that, for convenience, we assume is observed. We need the assumption that *f* is absolutely continuous, with finite variance σ_f^2 and finite Fisher information for location $\mathcal{I}(f) := \int (f'/f)^2 f \, \mathrm{d}x < \infty$. Under this assumption, the results of, for example, Kreiss (1987) imply that this MA(1) model is locally asymptotically normal with (univariate) central sequence

$$\Delta^{(n)}(\theta, f) = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \frac{-f'}{f}(\varepsilon_t^{(n)}(\theta))\varepsilon_{t-1}^{(n)}(\theta),$$

where the residuals $\varepsilon_t^{(n)}(\theta)$, $1 \leqslant t \leqslant n$, are computed from the recursion $\varepsilon_t^{(n)}(\theta) := Y_t^{(n)} - \theta\varepsilon_{t-1}^{(n)}$, with initial value $\varepsilon_0^{(n)}(\theta) = \varepsilon_0$. This central sequence is clearly of the form (3.2).

Under Assumption I(ii), for fixed $\boldsymbol{\theta}$ and *n*, the nonparametric model

$$\mathcal{E}_{\boldsymbol{\theta}}^{(n)} := (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}_{\boldsymbol{\theta}}^{(n)} = \{\mathrm{P}_{\boldsymbol{\theta},t}^{(n)} f : f \in \mathcal{F}\})$$

is generated by the group $(\mathcal{G}\mathrm{z}^{(n)}, \circ) = (\{\mathcal{G}_h^{(n)}, h \in \mathcal{H}\}, \circ)$ of continuous order-preserving transformations acting on $\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})$. More precisely, these transformations $\mathcal{G}_h^{(n)}$ are such that

$$\mathcal{G}_h^{(n)}(\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})) := (h(\varepsilon_1^{(n)}(\boldsymbol{\theta})), \ldots, h(\varepsilon_n^{(n)}(\boldsymbol{\theta}))), \tag{3.3}$$

where h belongs to the set $\mathcal{H}$ of all functions $h : \mathbb{R} \to \mathbb{R}$ that are continuous, strictly increasing, and satisfy $\lim_{x\to\pm\infty} h(x) = \pm\infty$. The corresponding maximal invariant σ-algebra is the σ-algebra $\sigma(R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta}))$ generated by the ranks $R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta})$ of the residuals $\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})$. This σ-algebra is distribution-free under $\mathcal{E}_{\boldsymbol{\theta}}^{(n)}$ and will play the role, in this section, of $\mathcal{B}^{(n)}(\boldsymbol{\theta})$.

In order for the results of Section 2 to be applicable, we need conditions on the function $\mathbf{J}_f$ ensuring that either condition (LF1) alone or conditions (LF1) and (LF2) together hold. In Section 3.2, we check that condition (LF1) is satisfied, with a rank-based efficient (at given f) influence function. Section 3.3 similarly deals with condition (LF2), and Section 3.4 draws some conclusions.

3.2. Ranks and the least favourable direction: checking for condition (LF1)

The notation $\varepsilon_0, \ldots, \varepsilon_p$ here is used for an arbitrary $(p+1)$-tuple of i.i.d. random variables with density f, independent of the residuals $\varepsilon_t^{(n)}$. Expectations with respect to $\varepsilon_0, \ldots, \varepsilon_p$ will be denoted by E_f which indicates that they depend neither on n nor on $\boldsymbol{\theta}$. Consider the following set of conditions on the function $\mathbf{J}_f$.

Assumption J.

(i) *The function $\mathbf{J}_f$ is such that $0 < \mathrm{E}_f\{\|\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p)\|^2\} < \infty$, and, $(\varepsilon_1, \ldots, \varepsilon_p)$-a.e.,*

$$\mathrm{E}_f\{\mathbf{J}_f(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_p)|\varepsilon_1, \ldots, \varepsilon_p\} = \mathbf{0}. \tag{3.4}$$

(ii) *The function $\mathbf{J}_f$ is componentwise monotone increasing with respect to all its arguments, or a linear combination of such functions.*

(iii) *For all sequences $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + O(1/\sqrt{n})$, we have, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,f}$, as $n \to \infty$,*

$$\frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} [\mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}_n), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}_n)) - \mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}_0))]$$

$$= -\mathbf{I}_f(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) + o_{\mathrm{P}}(1), \tag{3.5}$$

where the matrix-valued function $\boldsymbol{\theta} \mapsto \mathbf{I}_f(\boldsymbol{\theta}) := \mathrm{E}_f\{\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p)(\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p))^{\mathrm{T}}\}$ is continuous in $\boldsymbol{\theta}$ for all f (recall that $\mathbf{J}_f$ may depend on $\boldsymbol{\theta}$). Moreover, letting

$$\mathbf{J}_f^*(\varepsilon_t^{(n)}, \ldots, \varepsilon_{t-p}^{(n)}) := \mathbf{J}_f(\varepsilon_t^{(n)}, \ldots, \varepsilon_{t-p}^{(n)}) - \mathrm{E}_f\{\mathbf{J}_f(\varepsilon_t^{(n)}, \varepsilon_1, \ldots, \varepsilon_p)|\varepsilon_t^{(n)}\}, \tag{3.6}$$

we have, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_0,f}$,

$$\frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}[\mathbf{J}_f^*(\varepsilon_t^{(n)}(\boldsymbol{\theta}_n),\,\ldots,\,\varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}_n))-\mathbf{J}_f^*(\varepsilon_t^{(n)}(\boldsymbol{\theta}_0),\,\ldots,\,\varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}_0))]$$

$$= -\mathbf{I}_f^*(\boldsymbol{\theta}_0)\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}_0)+o_{\mathrm{P}}(1), \quad (3.7)$$

where $\boldsymbol{\theta}\mapsto\mathbf{I}_f^*(\boldsymbol{\theta}):=\mathrm{E}_f\{\mathbf{J}_f^*(\varepsilon_0,\,\ldots,\,\varepsilon_p)(\mathbf{J}_f^*(\varepsilon_0,\,\ldots,\,\varepsilon_p))^{\mathrm{T}}\}$ *is also continuous in* $\boldsymbol{\theta}$ *for all* f*.*

The assumption that $\mathbf{J}_f$ depends only on a finite number p of lagged innovations is made for notational convenience. This assumption rules out the autoregressive model. However, such models can easily be incorporated at the cost of technical details that are discussed in detail in Remark 3.2. Moreover, we do not explicitly allow here for the possibility that $\mathbf{J}_f$ depends on exogenous variables. This would rule out, for example, regression models with random design. Also this situation can easily be incorporated (see Remark 3.3 for the technicalities). The existence, for $\mathbf{J}_f$, of a moment of order 2 is needed to establish Proposition 3.1. Condition (3.4) allows us to conclude that the parametric central sequence $\boldsymbol{\Delta}_{\boldsymbol{\theta},f}^{(n)}$ is asymptotically normal,

$$\frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}\mathbf{J}_f(\varepsilon_t^{(n)},\,\ldots,\,\varepsilon_{t-p}^{(n)})\xrightarrow{\mathcal{L}}N(\mathbf{0},\,\mathbf{I}_f(\boldsymbol{\theta})),$$

via, for example, the martingale central limit theorem. Assumption J(ii) will allow us to replace the score functions $\mathbf{J}_f(\varepsilon_t^{(n)},\,\ldots,\,\varepsilon_{t-p}^{(n)})$ with rank-based versions. More precisely, the same argument as in Lemma 3.1 of Hallin and Puri (1991) shows the existence of functions $\mathbf{a}_f^{(n)}:\{1,\,\ldots,\,n\}^{p+1}\to\mathbb{R}^k$, such that

$$\lim_{n\to\infty}\mathrm{E}_f[\|\mathbf{J}_f(\varepsilon_1^{(n)},\,\ldots,\,\varepsilon_{p+1}^{(n)})-\mathbf{a}_f^{(n)}(R_1^{(n)},\,\ldots,\,R_{p+1}^{(n)})\|^2]=0, \quad (3.8)$$

as $n\to\infty$, where $R_1^{(n)},\,\ldots,\,R_n^{(n)}$ denote the ranks of $\varepsilon_1^{(n)},\,\ldots,\,\varepsilon_n^{(n)}$. Classical choices for $\mathbf{a}_f^{(n)}$ (with F the distribution function corresponding to f) are

$$\mathbf{a}_f^{(n)}(R_t^{(n)},\,\ldots,\,R_{t-p}^{(n)}):=\mathrm{E}_f\{\mathbf{J}_f(\varepsilon_t^{(n)},\,\ldots,\,\varepsilon_{t-p}^{(n)})|R_1^{(n)},\,\ldots,\,R_n^{(n)}\}$$

(exact scores), and

$$\mathbf{a}_f^{(n)}(R_t^{(n)},\,\ldots,\,R_{t-p}^{(n)}):=\mathbf{J}_f\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1}\right),\,\ldots,\,F^{-1}\left(\frac{R_{t-p}^{(n)}}{n+1}\right)\right)$$

(approximate scores). Assumption J(iii) is a smoothness condition on $\mathbf{J}_f$ and $\mathbf{J}_f^*$ that is required to verify Assumption B(iii). Moreover, we will use it in Section 3.4 to handle the effects of preliminary estimation of $\boldsymbol{\theta}$. Recall that the LAN condition together with (3.5) is equivalent to the LAN condition uniformly over $1/\sqrt{n}$-neighbourhoods of $\boldsymbol{\theta}$, referred to as the ULAN condition.

Example 3.1 (continued). For the MA(1) model (3.1), the function J_f is given by

$$J_f(\varepsilon_0, \varepsilon_1) := \frac{-f'}{f}(\varepsilon_0)\varepsilon_1,$$

which obviously satisfies Assumption J(i). Assumption J(ii) imposes an extra regularity condition on the density f that is satisfied for most well-known densities; recall that a monotone increasing log-derivative $-f'/f$ characterizes the class of strongly unimodal densities *f*. Writing $\mu_f := \int zf(z)\mathrm{d}z$, we immediately obtain

$$J^*_f(\varepsilon_0, \varepsilon_1) = \frac{-f'}{f}(\varepsilon_0)(\varepsilon_1 - \mu_f).$$

In order to verify Assumption J(iii), observe that

$$\varepsilon_t^{(n)}(\theta_n) = Y_t^{(n)} - \theta_n\varepsilon_{t-1}^{(n)}(\theta_n) = \varepsilon_t^{(n)}(\theta_0) + \theta_0\varepsilon_{t-1}^{(n)}(\theta_0) - \theta_n\varepsilon_{t-1}^{(n)}(\theta_n).$$

Assumption J(iii) now follows from standard arguments; see, for example Drost *et al*. (1997, Theorem 2.1).

The results of Section 2 suggest that the least favourable direction is, in the present case, generated by the sequence

$$\begin{aligned}\mathbf{H}_{\mathrm{lf}}^{(n)}(\boldsymbol{\theta}, f) &= \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f) - \mathrm{E}_{\boldsymbol{\theta},f}^{(n)}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\}\\ &= \frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}\mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}))\\ &\quad - \mathrm{E}_{\boldsymbol{\theta},f}^{(n)}\left\{\frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}\mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}))\middle| R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta})\right\}\\ &= \frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n}[\mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}))\\ &\quad - \mathrm{E}_{\boldsymbol{\theta},f}^{(n)}\{\mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}))|R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta})\}].\end{aligned} \tag{3.9}$$

Under Assumptions I and J, this can indeed be proved by checking that condition (LF1) is satisfied.

Before doing this, we need a corollary of a result in Hallin *et al*. (1985). In the rest of this paper, we simplify notation by writing $\varepsilon_t^{(n)}$ for $\varepsilon_t^{(n)}(\boldsymbol{\theta}_0)$ when it is safe to do so. Recall that $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. with density f under $\mathrm{P}_{\boldsymbol{\theta}_0,f}^{(n)}$. The following result is proved in Section 4.1 of Hallin *et al*. (1985).

Proposition 3.1. *Suppose that Assumptions J(i) and J(ii) are satisfied. Let* $\mathbf{a}_f^{(n)}$ *be any function satisfying* (3.8). *Writing*

$$\mathbf{S}_f^{(n)}(\boldsymbol{\theta}_0) := \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{a}_f^{(n)}(R_t^{(n)}, R_{t-1}^{(n)}, \ldots, R_{t-p}^{(n)}), \tag{3.10}$$

$$\mathbf{m}_f^{(n)} := [n(n-1)\cdots(n-p)]^{-1} \sum_{1 \leqslant i_0 \neq \ldots \neq i_p \leqslant n} \cdots \sum \mathbf{a}_f^{(n)}(i_0, \ldots, i_p), \tag{3.11}$$

define

$$\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta}_0) := \sqrt{n-p}(\mathbf{S}_f^{(n)} - \mathbf{m}_f^{(n)}). \tag{3.12}$$

Then,

$$\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \mathbf{J}_f(\varepsilon_t^{(n)}, \ldots, \varepsilon_{t-p}^{(n)}) - \frac{\sqrt{n-p}}{n(n-1)\cdots(n-p)} \sum_{1 \leqslant i_0 \neq \ldots \neq i_p \leqslant n} \cdots \sum \mathbf{J}_f(\varepsilon_{i_0}^{(n)}, \ldots, \varepsilon_{i_p}^{(n)}) + o_{L_2}(1), \tag{3.13}$$

under $\mathrm{P}_{\boldsymbol{\theta}_0, f}^{(n)}$, *as* $n \to \infty$.

Expression (3.13), and more particularly the U-statistic term with kernel $\mathbf{J}_f$, looks somewhat awkward. We therefore provide the following corollary.

Corollary 3.2. *Under the same conditions as in Proposition* 3.1, *we also have*

$$\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \mathbf{J}_f^*(\varepsilon_t^{(n)}, \ldots, \varepsilon_{t-p}^{(n)}) + o_{\mathrm{P}}(1),$$

with J_f^* *defined in* (3.6).

Proof. Note that by Proposition 3.1 we have

$$\sqrt{n-p}\left(\mathbf{S}_f^{(n)} - \mathbf{m}_f^{(n)} - \frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{J}_f^*(\varepsilon_t^{(n)}, \ldots, \varepsilon_{t-p}^{(n)})\right)$$

$$= \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \mathrm{E}_f\{\mathbf{J}_f(\varepsilon_t^{(n)}, \varepsilon_1, \ldots, \varepsilon_p) | \varepsilon_t^{(n)}\} - \frac{\sqrt{n-p}}{n(n-1)\cdots(n-p)} \sum_{1 \leqslant i_0 \neq \ldots \neq i_p \leqslant n} \cdots \sum \mathbf{J}_f(\varepsilon_{i_0}^{(n)}, \ldots, \varepsilon_{i_p}^{(n)}) + o_{\mathrm{P}}(1)$$

$$= -\frac{1}{\sqrt{n-p}} \frac{1}{n(n-1)\cdots(n-p+1)} \sum_{1 \leqslant i_0 \neq \ldots \neq i_p \leqslant n} \cdots \sum \mathbf{J}_f^*(\varepsilon_{i_0}^{(n)}, \varepsilon_{i_1}^{(n)}, \ldots, \varepsilon_{i_p}^{(n)}) + o_{\mathrm{P}}(1).$$

By classical results on U-statistics (see, for example, Serfling 1980, Section 5.3.4) and the fact that $\mathrm{E}_f\{\mathbf{J}^*_f(\varepsilon_0, \ldots, \varepsilon_p)|\varepsilon_s\} = \mathbf{0}$, for $s = 0, 1, \ldots, p$, the claim is proved. □

The following theorem shows that, under the assumptions made, condition (LF1) is indeed satisfied.

Proposition 3.3. *Consider the semi-parametric model* (3.1). *Assume that Assumptions I and J are satisfied. Then, there exists a mapping* $q : (-1, 1)^k \to \mathcal{F}$ *such that the parametric model*

$$\mathcal{E}^{(n)}_q = (\mathcal{X}^{(n)}, \mathcal{A}^{(n)}, \mathcal{P}^{(n)} = \{\mathrm{P}^{(n)}_{\boldsymbol{\theta}, q(\boldsymbol{\eta})} \colon \boldsymbol{\theta} \in \boldsymbol{\Theta}, \boldsymbol{\eta} \in (-1, 1)^k\}),$$

is locally asymptotically normal at $(\boldsymbol{\theta}_0, \mathbf{0})$ *with central sequence*

$$\frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \begin{pmatrix} J_f(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0), \ldots, \varepsilon^{(n)}_{t-p}(\boldsymbol{\theta}_0)) \\ J^*_f(\varepsilon^{(n)}_t(\boldsymbol{\theta}_0), \ldots, \varepsilon^{(n)}_{t-p}(\boldsymbol{\theta}_0)) \end{pmatrix} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \begin{bmatrix} \mathbf{I}_f(\boldsymbol{\theta}_0) & \mathbf{I}_f(\boldsymbol{\theta}_0) - \mathbf{I}^*_f(\boldsymbol{\theta}_0) \\ \mathbf{I}_f(\boldsymbol{\theta}_0) - \mathbf{I}^*_f(\boldsymbol{\theta}_0) & \mathbf{I}_f(\boldsymbol{\theta}_0) - \mathbf{I}^*_f(\boldsymbol{\theta}_0) \end{bmatrix}\right), \tag{3.14}$$

where $\mathbf{I}^*_f(\boldsymbol{\theta}_0)$ *is the variance of the limiting distribution of* $\mathrm{E}^{(n)}_{\boldsymbol{\theta},f}\{\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f)|\mathcal{B}^{(n)}(\boldsymbol{\theta})\}$.

Hence, $\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}, f)$ *satisfies condition* (LF1); *the corresponding influence function (for* $\boldsymbol{\theta}$*) is* $\mathbf{I}^*_f(\boldsymbol{\theta}_0)^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}_0)$. *It follows that semi-parametrically efficient inference (at* $\boldsymbol{\theta}$ *and f) can be based on the efficient central sequence*

$$\boldsymbol{\Delta}^{*(n)}_{\boldsymbol{\theta},f} := \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \mathbf{J}^*_f(\varepsilon^{(n)}_t, \ldots, \varepsilon^{(n)}_{t-p}), \tag{3.15}$$

of which $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta})$ *defined in* (3.12) *provides a rank-based version.*

Proof. We need to verify the conditions of Proposition 2.5. Assumption A is just the equivalent of Assumption I. In order to verify Assumption B, we first rewrite the candidate least favourable direction $\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0, f)$. Let $\mathbf{S}^{(n)}_{f;\mathrm{exact}}$ be defined by (3.10), using the exact score functions given above. Observe that the corresponding centring constants $\mathbf{m}^{(n)}_{f;\mathrm{exact}}$ in (3.11) are zero. Hence, from (3.9) and Corollary 3.3,

$$\begin{aligned}\mathbf{H}^{(n)}_{\mathrm{lf}}(\boldsymbol{\theta}_0, f) &= \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, f) - \sqrt{n-p}\,\mathbf{S}^{(n)}_{f;\mathrm{exact}} \\ &= \boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}_0, f) - \sqrt{n-p}(\mathbf{S}^{(n)}_{f;\mathrm{exact}} - \mathbf{m}^{(n)}_{f;\mathrm{exact}}) \\ &= \frac{1}{\sqrt{n-p}} \mathrm{E}_f\{\mathbf{J}_f(\varepsilon^{(n)}_t, \varepsilon_1, \ldots, \varepsilon_p)|\varepsilon^{(n)}_t\} + o_{\mathrm{P}}(1).\end{aligned}$$

In view of Corollary 3.2, the same result actually holds for any score functions $\mathbf{a}_f$ satisfying (3.8). Assumption B(i) now follows easily from the martingale central limit theorem. Moreover, Assumption B(iii) is an immediate consequence of Assumption J(iii) (in particular, of subtracting (3.5) and (3.7)). To prove that Assumption B(ii) is satisfied, we use the general construction of least favourable parametric submodels for i.i.d. models as in Example 3.2.1 of Bickel *et al*. (1993). More precisely, we define

$$f_{\boldsymbol{\eta}}(x) := \frac{f(x)\psi(\boldsymbol{\eta}^{\mathrm{T}}\mathrm{E}_f\{\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p)|\varepsilon_0 = x\})}{\int \psi(\boldsymbol{\eta}^{\mathrm{T}}\mathrm{E}_f\{\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p)|\varepsilon_0 = z\})f(z)\mathrm{d}z}, \tag{3.16}$$

where ψ is a positive, bounded, three times continuously differentiable function with bounded derivatives and $\psi(0) = \psi'(0) = 1$. For example, let $\psi(z) = 2/(1 + \exp(-2z))$. Intuitively, the $\boldsymbol{\eta}$-score now takes the form

$$\mathrm{grad}_{\boldsymbol{\eta}} \log f_{\boldsymbol{\eta}}(\varepsilon_0)|_{\boldsymbol{\eta}=\mathbf{0}} = \mathrm{E}_f\{\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p)|\varepsilon_0\}, \tag{3.17}$$

and the LAN condition in $\boldsymbol{\eta}$ required in Assumption B(ii) follows as in Example 3.2.1 of Bickel *et al*. (1993). The form of the resulting influence function directly follows from (2.10). □

Proposition 3.3 shows that $\underset{\sim}{\boldsymbol{\Delta}}{}_f^{(n)}(\boldsymbol{\theta}_0)$ constitutes a rank-based version of the semi-parametrically efficient (at f) central sequence. Using only the ranks of the innovations, one can thus construct somewhere efficient inference procedures in the semi-parametric model where f is considered a nuisance parameter. This result, moreover, is obtained without explicitly going through tangent space calculations, and without computing the efficient score function (which, as we will see, turns out to be $\mathbf{J}_f^*$). The intuition for this can, however, easily be explained using tangent space arguments. The efficient score is obtained by taking the residual of the projection of the parametric score on the tangent space generated by f (see Bickel *et al*. 1993, Chapter 3). In the set-up of this subsection, f is completely unrestricted, so that the tangent space consists of all square-integrable functions of $\varepsilon_t^{(n)}$, which have expectation zero and vanish when f vanishes. The projection of $\mathbf{J}_f(\varepsilon_t^{(n)}, \ldots, \varepsilon_{t-p}^{(n)})$ onto this space obviously is $\mathrm{E}_f\{\mathbf{J}_f(\varepsilon_t^{(n)}, \varepsilon_1, \ldots, \varepsilon_p)|\varepsilon_t^{(n)}\}$, which therefore generates the least favourable submodel. This substantiates our claim that $\mathbf{J}_f^*$ indeed is the efficient score function.

Example 3.1 (continued). For the MA(1) model, the efficient score function is $J_f^*(\varepsilon_0, \varepsilon_1) = (-f'/f)(\varepsilon_0)(\varepsilon_1 - \mu_f)$. In case $\mu_f = 0$, the efficient score function J_f^* and the parametric score function J_f coincide, and the model is adaptive – a well-known result (see Kreiss 1987; or Drost *et al*. 1997, Example 4.2).

We conclude this section with some remarks.

Remark 3.1. The semi-parametric model is often not defined for *all* densities *f*, but is restricted, for example, to those that are absolutely continuous, with finite Fisher information for location as in Example 3.1. In this case, (3.16) need not define a true parametric *sub*model, as f_η does not necessarily satisfy the same regularity conditions. However, the class of densities that do satisfy the regularity conditions is generally 'dense' in the class of all densities (this can actually be taken as a definition of what is meant by 'regularity condition'). In that case, it is always possible to choose a family of densities $f_{\boldsymbol{\eta}}$ that satisfy the imposed regularity conditions, while generating the same $\boldsymbol{\eta}$-score as in (3.17). Formally, it can be shown that the tangent spaces generated by such dense (in the L_2 sense) subsets coincide with those of the complete model; see, for example, Bickel *et al*. (1993, Chapter 4), for similar problems.

Remark 3.2. The restriction, in Assumption J, to score functions involving only a finite number p of residuals formally excludes some important models such as the AR ones. This is not a serious restriction, however. To show this, let us consider the simple AR(1) model $Y_t - \theta Y_{t-1} = \varepsilon_t$, with $|\theta| < 1$ and innovation density f. This model, under the same assumptions on f as in Example 3.1, is locally asymptotically normal, with central sequence

$$\Delta^{(n)}(\theta, f) = \frac{1}{\sqrt{n-1}} \sum_{t=2}^{n} \phi_f(\varepsilon_t(\theta)) Y_{t-1},$$

where $\varepsilon_t(\theta) := Y_t - \theta Y_{t-1}$ (assuming, for simplicity, that Y_0 has been observed, and that the likelihoods are conditional upon Y_0) and $\phi_f := -f'/f$. Since, under $\mathrm{P}^{(n)}_{\theta,f}$, $Y_{t-1} = \sum_{i=0}^{t-1} \theta^i \varepsilon_{t-1-i}$, the same central sequence can be re-expressed (up to $o_{\mathrm{P}}(1)$ terms) as

$$\Delta^{(n)}(\theta, f) = \frac{1}{\sqrt{n-1}} \sum_{i=1}^{n-1} \theta^{i-1} \sum_{t=i+1}^{n} \phi_f(\varepsilon_t(\theta)) \varepsilon_{t-i}(\theta) \tag{3.18}$$

$$= \sum_{i=1}^{q} \theta^{i-1} \frac{1}{\sqrt{n-i}} \sum_{t=i+1}^{n} \phi_f(\varepsilon_t(\theta)) \varepsilon_{t-i}(\theta) + R_q^{(n)} = \Delta^{(n)}_{(q)}(\theta, f) + R_q^{(n)},$$

say, where

$$\theta^{-q} R_q^{(n)} := \sum_{i=1}^{n-q-1} \theta^{i-1} \frac{1}{\sqrt{n-i}} \sum_{t=i+1}^{n} \phi_f(\varepsilon_t(\theta)) \varepsilon_{t-i}(\theta),$$

hence

$$\mathrm{var}_\theta(R_q^{(n)}) = \theta^{2q} \sigma_f^2 \mathcal{I}(f) \sum_{i=1}^{n-q-1} \theta^{2(i-1)} \leqslant \theta^{2q} \sigma_f^2 \mathcal{I}(f) \frac{1}{1-\theta^2}. \tag{3.19}$$

Letting $q = q(n) \uparrow \infty$, (this convergence to infinity can be arbitrarily slow, as long as $q(n) \leqslant n-1$), we thus have that $R^{(n)}_{q(n)} = O_{\mathrm{P}}(|\theta|^{q(n)}) = o_{\mathrm{P}}(1)$ as $n \to \infty$. It follows that $\Delta^{(n)}_{(q(n))}(\theta, f)$ is a central sequence. Now, for any value of q, $\Delta^{(n)}_{(q)}(\theta, f)$ is a linear combination of terms satisfying Assumption J. A straightforward adaptation of the proofs shows that all the results of this section still hold in such cases. The higher-order AR(p) case follows along the same lines, with the coefficient θ^{2q} in (3.19) replaced with Λ^{2q}, where Λ^{-1} denotes the modulus of the characteristic root which lies closest to the unit circle.

Remark 3.3. The results of this section remain valid if the model contains exogenous variables, that is, observable random variables whose distribution does not depend on the parameters $\boldsymbol{\theta}$ or on f, such as the regression model with random design to be discussed in detail in Example 4.2. In this case, the exogenous variables, denoted, for instance, by $\mathbf{X}^{(n)} := (X_1^{(n)}, \ldots, X_n^{(n)})$, can be included in the invariant σ-algebra $\mathcal{B}^{(n)}(\theta)$; using the same notation as above, we would obtain $\mathcal{B}^{(n)}(\boldsymbol{\theta}) = \sigma(R_1^{(n)}(\boldsymbol{\theta}), \ldots, R_n^{(n)}(\boldsymbol{\theta}); \mathbf{X}^{(n)})$. The central sequence will now generally be of the form

$$\boldsymbol{\Delta}^{(n)}(\boldsymbol{\theta}, f) = \frac{1}{\sqrt{n-p}} \sum_{t=p+1}^{n} \mathbf{J}_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_{t-p}^{(n)}(\boldsymbol{\theta}); X_t^{(n)}).$$

Writing $\mathbf{J}_{f.t}(\varepsilon_0, \ldots, \varepsilon_p)$ for $\mathbf{J}_f(\varepsilon_0, \ldots, \varepsilon_p; X_t^{(n)})$, Assumptions J(i) and J(ii) are supposed to hold for $\mathbf{J}_{f,t}$. Assumption J(iii) also is supposed to hold, with (cf. Drost *et al.*, 1997, relation (2.3))

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{J}_{f,t}(\varepsilon_0, \ldots, \varepsilon_p)(\mathbf{J}_{f,t}(\varepsilon_0, \ldots, \varepsilon_p))^{\mathrm{T}} \xrightarrow{\mathrm{P}} \mathbf{I}_f(\boldsymbol{\theta}),$$

and

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \mathbf{J}^*_{f,t}(\varepsilon_0, \ldots, \varepsilon_p)(\mathbf{J}^*_{f,t}(\varepsilon_0, \ldots, \varepsilon_p))^{\mathrm{T}} \xrightarrow{\mathrm{P}} \mathbf{I}^*_f(\boldsymbol{\theta}).$$

Proposition 3.3 now continues to hold, under appropriate assumptions on the asymptotic behaviour of the exogenous variables, and with obvious notational changes.

3.3. Ranks and the efficient influence function: checking for condition (LF2)

From the results of Section 3.2, it thus follows that locally and asymptotically optimal (at $\boldsymbol{\theta}$ and f) inference in the parametric submodel of Proposition 3.3 should be based on the central sequence (3.15), of which the $\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta})$, with either exact or approximate scores $\mathbf{a}_f^{(n)}$, provide rank-based versions. The same inference will be semi-parametrically efficient (still at $\boldsymbol{\theta}$) over some subset $\mathcal{C} \subseteq \mathcal{F}$ provided that not only condition (LF1) but also condition (LF2) can be shown to hold for all $f \in \mathcal{C}$.

In order to satisfy condition (LF2), we should be able to neutralize the dependence on f of the influence function $\mathbf{I}^*_f(\theta)^{-1}\underset{\sim}{\boldsymbol{\Delta}}_f^{(n)}(\boldsymbol{\theta})$ by substituting some adequate estimator $\hat{f}_n$ for f. Define $\mathcal{C}$ as the class of all densities f in $\mathcal{F}$ such that:

(i) $\mathbf{J}_f$ satisfies Assumption J; and
(ii) there exists an estimator $\hat{f}_n$, measurable with respect to the order statistics of the residuals $\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})$, such that

$$\mathrm{E}_f\{\|\mathbf{a}_{\hat{f}_n}^{(n)}(R_1^{(n)}, \ldots, R_{p+1}^{(n)}) - \mathbf{a}_f^{(n)}(R_1^{(n)}, \ldots, R_{p+1}^{(n)})\|^2 | \hat{f}_n\} = o_{\mathrm{P}}(1) \qquad (3.20)$$

as $n \to \infty$, where the rank scores $\mathbf{a}_f^{(n)}$ are defined in (3.8).

The class $\mathcal{C}$ thus contains all densities $f \in \mathcal{F}$ such that Assumption J holds, and the rank scores $\mathbf{a}_f^{(n)}$ can be estimated consistently. A possible estimator, for an important special case that arises in all the examples considered Section 4, is given in (1.5.7) of Hájek and Šidák

(1967, Chapter VII). While this estimator may not be the best choice from a practical point of view, it shows that the assumption that $f \in \mathcal{C}$ is not overly restrictive.

Finally, denote by $\hat{\mathbf{I}}^{(n)*}_{\hat{f}_n}(\boldsymbol{\theta}) := \text{var}(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\hat{f}_n}(\boldsymbol{\theta})|\hat{f}_n)$ the covariance matrix of $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\hat{f}_n}(\boldsymbol{\theta})$ conditional on $\hat{f}_n$. Since $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\hat{f}_n}(\boldsymbol{\theta})$ is conditionally distribution-free, $\hat{\mathbf{I}}^{(n)*}_{\hat{f}_n}(\boldsymbol{\theta})$ does not depend on *f*, and can be computed from the observations; as we shall see, it consistently estimates $\mathbf{I}^*_f(\boldsymbol{\theta})$. The following result shows that conditions (LF1) and (LF2) are satisfied for all $f \in \mathcal{C}$.

Proposition 3.4. *For all* $f \in \mathcal{C}$ *and* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\hat{\mathbf{I}}^{(n)*}_{\hat{f}_n}(\boldsymbol{\theta})^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\hat{f}_n}(\boldsymbol{\theta}) = \mathbf{I}^*_f(\boldsymbol{\theta})^{-1}\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta}) + o_{\mathrm{P}}(1) \qquad \text{under } \mathrm{P}^{(n)}_{\boldsymbol{\theta},f}.$$

Conditions (LF1) *and* (LF2) *are thus satisfied, and* $\boldsymbol{\Delta}^*_{\boldsymbol{\theta},f_n}$*, as well as* $\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\hat{f}_n}(\boldsymbol{\theta})$*, are versions of the efficient (at f and* $\boldsymbol{\theta}$*) central sequence for* $\mathcal{E}^{(n)}$.

Proof. Let $f \in \mathcal{C}$. It follows from Lemma 4 of Hallin *et al*. (1985) that, for any measurable function $d : \mathbb{R}^{p+1} \to \mathbb{R}$ such that $\mathrm{E}\{d(R^{(n)}_1, \ldots, R^{(n)}_{p+1})^2\} < \infty$ and any $|s-t| > p$, we have

$$|\text{cov}(d(R^{(n)}_s, \ldots, R^{(n)}_{s+p}), d(R^{(n)}_t, \ldots, R^{(n)}_{t+p}))| \leqslant \frac{K}{n}\mathrm{E}\{d(R^{(n)}_1, \ldots, R^{(n)}_{p+1})^2\}, \qquad n \in \mathbb{N}.$$

This implies that, for any sequence $d^{(n)} := a^{(n)}_{\hat{f}_n} - a^{(n)}_f$ such that $\hat{f}^{(n)}$ satisfies (3.20), we have

$$\begin{aligned}
&\text{var}\left(\frac{1}{\sqrt{n-p}}\sum_{t=p+1}^{n} d^{(n)}(R^{(n)}_t, \ldots, R^{(n)}_{t-p})|\hat{f}_n\right)\\
&\quad= \frac{1}{n+p}\sum_{t=p+1}^{n}\text{var}(d^{(n)}(R^{(n)}_t, \ldots, R^{(n)}_{t-p})|\hat{f}_n)\\
&\qquad+\sum_{k=1}^{p}\frac{2}{n-p}\sum_{t=p+1}^{n}\text{cov}(d^{(n)}(R^{(n)}_t, \ldots, R^{(n)}_{t-p}), d^{(n)}(R^{(n)}_{t+k}, \ldots, R^{(n)}_{t+k-p})|\hat{f}_n)\\
&\qquad+\frac{2}{n-p}\sum_{t=p+1}^{n}\sum_{s=t+p+1}^{n}\text{cov}(d^{(n)}(R^{(n)}_t, \ldots, R^{(n)}_{t-p}), d^{(n)}(R^{(n)}_s, \ldots, R^{(n)}_{s-p})|\hat{f}_n)\\
&\quad\leqslant (1+2p)\,\text{var}(d^{(n)}(R^{(n)}_t, \ldots, R^{(n)}_{t-p})|\hat{f}_n)\\
&\qquad+\frac{2}{n-p}(n-p)\frac{1}{2}(n-2p)\frac{K}{n}\mathrm{E}\{d^{(n)}(R^{(n)}_1, \ldots, R^{(n)}_{p+1})^2|\hat{f}_n\} = o_{\mathrm{P}}(1).
\end{aligned}$$

This result readily extends to vector-valued functions $\mathbf{d}^{(n)} = \mathbf{a}^{(n)}_{\hat{f}_n} - \mathbf{a}^{(n)}_f : \mathbb{R}^{p+1} \to \mathbb{R}^k$, yielding

$$\text{var}\left(\underset{\sim}{\boldsymbol{\Delta}}^{(n)}_{\hat{f}_n}(\boldsymbol{\theta}) - \underset{\sim}{\boldsymbol{\Delta}}^{(n)}_f(\boldsymbol{\theta})|\hat{f}_n\right) = o_{\mathrm{P}}(1). \tag{3.21}$$

Since $\mathrm{E}[\underset{\sim}{\boldsymbol{\Delta}}_{\hat{f}_n}^{(n)}(\boldsymbol{\theta})] = \mathbf{0}$ $\mathrm{P}_{\boldsymbol{\theta},f}^{(n)}$-almost surely and $\mathrm{E}[\underset{\sim}{\boldsymbol{\Delta}}_{f}^{(n)}(\boldsymbol{\theta})] = \mathbf{0}$, (3.21) and the Chebyshev inequality imply that

$$\underset{\sim}{\boldsymbol{\Delta}}_{\hat{f}_n}^{(n)}(\boldsymbol{\theta}) = \underset{\sim}{\boldsymbol{\Delta}}_{f}^{(n)}(\boldsymbol{\theta}) + o_{\mathrm{P}}(1). \tag{3.22}$$

Similarly, (3.21) also implies that

$$\begin{aligned}\mathrm{var}\left(\underset{\sim}{\boldsymbol{\Delta}}_{\hat{f}_n}^{(n)}(\boldsymbol{\theta})|\hat{f}_n\right) - \mathrm{var}\left(\underset{\sim}{\boldsymbol{\Delta}}_{f}^{(n)}(\boldsymbol{\theta})|\hat{f}_n\right) &= \mathrm{var}\left(\underset{\sim}{\boldsymbol{\Delta}}_{\hat{f}_n}^{(n)}(\boldsymbol{\theta})|\hat{f}_n\right) - \mathrm{var}\left(\underset{\sim}{\boldsymbol{\Delta}}_{f}^{(n)}(\boldsymbol{\theta})\right) \\ &= \hat{\mathbf{I}}_{\hat{f}_n}^{(n)*}(\theta) - \mathbf{I}_f^*(\theta) = o_{\mathrm{P}}(1). \end{aligned} \tag{3.23}$$

The end of the proof directly follows from (3.22) and (3.23). □

3.4. Semi-parametrically efficient, conditionally distribution-free inference, and semi-parametrically efficient permutation tests

Proposition 3.4 establishes the possibility of semi-parametrically efficient inference using the rank-based central sequences constructed from estimated densities. In particular, comparing $\underset{\sim}{\boldsymbol{\Delta}}_{\hat{f}_n}(\boldsymbol{\theta})$-measurable statistics with their conditional (with respect to $\hat{f}_n$) quantiles – i.e., treating them as if they were genuine rank statistics, with deterministically determined scores – yields distribution-free, semi-parametrically efficient sequences of tests; see Choi *et al*. (1996) or Hallin and Werker (1999) for a discussion of locally asymptotically optimal testing in locally asymptotically normal families. Such tests actually are *permutation tests*, exhibiting Neyman α-structure with respect to the (sufficient and complete) order statistic of the residuals $(\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta}))$.

The ideas leading to rank-based inference, as outlined above, remain valid in models with other invariance structures. Suppose, for example, that the density $f \in \mathcal{F}$ of the innovations is known to be symmetric with respect to the origin. The corresponding maximal invariant σ-algebra, playing the role of $\mathcal{B}^{(n)}(\boldsymbol{\theta})$, is now generated by the signs of the innovations and the ranks of their absolute values, that is, the traditional *signed ranks*. The functions $\mathbf{a}_f^{(n)}$ introduced in Assumption J(ii) will then depend on these signed ranks as well. Hence, the density estimator in (3.20) may only depend on the order statistics of the absolute values of the innovations (as is intuitively clear). A signed-rank equivalent of Proposition 3.1 is established in Hallin and Puri (1991). The construction of the least favourable parametric submodel follows along the same lines as above, and Proposition 3.4 remains valid with obvious adaptations. Assuming that the densities f are symmetric with respect to the origin provides the innovations $\varepsilon_t^{(n)}$ with a well-identified location. Another way of achieving this involves restricting f to the class of densities having zero median. The corresponding maximal invariant σ-algebra, playing the role of $\mathcal{B}^{(n)}(\boldsymbol{\theta})$, is now generated by the signs of the innovations and their ranks. Methods based on signs and ranks are the subject of ongoing research (Hallin *et al*. 2002).

4. Examples

In this section, we briefly discuss several examples where the general theory of the previous sections can be used. For pedagogical reasons, we restrict the discussion to the simplest cases, but it should be stressed that much more elaborate models can be treated similarly.

Example 4.1 *One-sample location models*. Consider the model defined by

$$Y_t^{(n)} = \theta + \varepsilon_t^{(n)}, \qquad t = 1, \ldots, n,$$

where $\theta \in \mathbb{R}$ ($p = K = 1$) and $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. with density f. Under the condition that f is absolutely continuous, with $\int (f'(z)/f(z))^2 f(z)\mathrm{d}z < \infty$, this model satisfies the ULAN condition with central sequence (3.2), and $J_f(\varepsilon) = -f'(\varepsilon)/f(\varepsilon)$. Let $R_1^{(n)}, \ldots, R_n^{(n)}$ denote the ranks of the residuals $\varepsilon_1^{(n)}(\theta), \ldots, \varepsilon_n^{(n)}(\theta)$, where $\varepsilon_t^{(n)}(\theta) = Y_t^{(n)} - \theta$. These ranks coincide with the ranks of the observations $Y_1^{(n)}, \ldots, Y_n^{(n)}$ themselves. If the density f is completely unrestricted (but for the regularity imposed above), it follows from Section 3 that an efficient central sequence is

$$\mathrm{E}_\theta^{(n)}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^n J_f(\varepsilon_t^{(n)}) | R_1^{(n)}, \ldots, R_n^{(n)}\right\} = 0. \tag{4.1}$$

This is of course not surprising, since θ in this model is not identified.

In order to identify θ, we need to impose a *location* restriction on f. If we impose the condition that $f \in \mathcal{F}_0$, that is, that f has median zero, the maximal invariant σ-field (at θ) is generated by the ranks $R_1^{(n)}, \ldots, R_n^{(n)}$ and the signs $s_1^{(n)}, \ldots, s_n^{(n)}$, where $s_t^{(n)} = \operatorname{sign}(\varepsilon_t^{(n)}(\theta))$. The efficient central sequence is now

$$\mathrm{E}_\theta^{(n)}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^n J_f(\varepsilon_t^{(n)}) | R_1^{(n)}, \ldots, R_n^{(n)}, s_1^{(n)}, \ldots, s_n^{(n)}\right\} = \frac{2f(0)}{\sqrt{n}}\sum_{t=1}^n s_t^{(n)}, \tag{4.2}$$

since $\mathrm{E}_f\{J_f(\varepsilon)| \operatorname{sign}(\varepsilon)\} = 2f(0)\operatorname{sign}(\varepsilon)$. The density $f(0)$ at the origin of course should be estimated (as it only appears as a constant factor, a simple consistent estimation is sufficient). A vast literature has been devoted to this estimation problem; see Yang (1985), Falk (1986) or Zelterman (1990) for recent references.

It follows from (4.2) that the optimal semi-parametric estimator is the sample median, the optimal semi-parametric test the sign test. This is quite a classical result; see, for example, Lehmann (1986). No further improvement can be expected from estimating f.

If we restrict the model further by imposing the condition that $f \in \mathcal{F}_+$ (f symmetric about zero), then adaptive inference is possible, since, if we denote by $s_1^{(n)} R_{+;1}^{(n)}, \ldots, s_n^{(n)} R_{+;n}^{(n)}$ the signed ranks associated with $\varepsilon_t^{(n)}(\theta)$, we obtain, from Hájek's projection theorem (see, for example, Puri and Sen 1985),

$$\mathrm{E}_{\theta}^{(n)}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} J_f(\varepsilon_t^{(n)})|\, s_1^{(n)}R_1^{(n)}, \ldots, s_n^{(n)}R_n^{(n)}\right\} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} s_t^{(n)} J_f\left(F_1^{-1}\left(\frac{n+1+R_{+;t}^{(n)}}{2(n+1)}\right)\right) + o_{\mathrm{P}}(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} J_f(\varepsilon_t^{(n)}) + o_{\mathrm{P}}(1). \tag{4.3}$$

Signed rank methods here are thus somewhere efficient. Substituting an adequate estimator $\hat{f}_n$ for f (see Section 3.4) yields adaptive (a fortiori semi-parametrically efficient) inference.

Example 4.2 *Regression models*. Let $\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}$ be a sequence of K-dimensional exogenous covariables. Consider the model defined by

$$Y_t^{(n)} = (\mathbf{X}_t^{(n)})^{\mathrm{T}}\boldsymbol{\theta} + \varepsilon_t^{(n)}, \qquad t = 1, \ldots, n,$$

where $\boldsymbol{\theta} \in \mathbb{R}^k$ and $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. with density f. In this case, the maximal invariant (at $\boldsymbol{\theta}$) σ-algebra $\mathcal{B}^{(n)}(\boldsymbol{\theta})$ is generated by $(R_1^{(n)}, \ldots, R_n^{(n)}; \mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)})$, where $R_t^{(n)} = R_t^{(n)}(\boldsymbol{\theta})$ again denotes the rank of the residual $\varepsilon_t^{(n)}(\boldsymbol{\theta}) = Y_t^{(n)} - (\mathbf{X}_t^{(n)})^{\mathrm{T}}\boldsymbol{\theta}$ among $\varepsilon_1^{(n)}(\boldsymbol{\theta}), \ldots, \varepsilon_n^{(n)}(\boldsymbol{\theta})$; see Remark 3.3. The ULAN property is satisfied if the exogenous variables are square-integrable and f satisfies the conditions of Example 4.1. The central sequence then is

$$\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} J_f(\varepsilon_t^{(n)})\mathbf{X}_t^{(n)}. \tag{4.4}$$

Letting $\alpha_f^{(n)}(r) = \mathrm{E}_f^{(n)}\{J_f(\varepsilon_t^{(n)})|R_t^{(n)} = r\}$, the efficient central sequence follows from

$$\mathrm{E}_f^{(n)}\{\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\theta})|\mathcal{B}^{(n)}(\boldsymbol{\theta})\} = \mathrm{E}_f^{(n)}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} J_f(\varepsilon_t^{(n)}(\boldsymbol{\theta}))\mathbf{X}_t^{(n)}|\mathcal{B}^{(n)}(\boldsymbol{\theta})\right\}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \alpha_f^{(n)}(R_t^{(n)})\mathbf{X}_t^{(n)}. \tag{4.5}$$

It can be shown (Bickel 1982; or Drost *et al*. 1997) that a version of the efficient central sequence for this model is

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} J_f\left(\varepsilon_t^{(n)}\right)\left(\mathbf{X}_t^{(n)} - \overline{\mathbf{X}}^{(n)}\right), \tag{4.6}$$

provided that $\overline{\mathbf{X}}^{(n)} = (1/n)\sum_{t=1}^{n}\mathbf{X}_t^{(n)} = o_{\mathrm{P}}(1)$. One may also verify directly that (4.5) and (4.6) are asymptotically equivalent, since

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} J_f(\varepsilon_t^{(n)})(\mathbf{X}_t^{(n)} - \overline{\mathbf{X}}^{(n)}) - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \alpha_f^{(n)}(R_t^{(n)})\mathbf{X}_t^{(n)}$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(J_f(\varepsilon_t^{(n)}) - \alpha_f^{(n)}(R_t^{(n)}))(\mathbf{X}_t^{(n)} - \overline{\mathbf{X}}^{(n)}) - \frac{1}{\sqrt{n}}\overline{\mathbf{X}}^{(n)}\sum_{t=1}^{n} \alpha_f^{(n)}(R_t^{(n)}),$$

where the first term is $o_{\mathrm{P}}(1)$ by Hájek's projection theorem, and the second term is $o_{\mathrm{P}}(1)$ since $\overline{\mathbf{X}}^{(n)}$ is $o_{\mathrm{P}}(1)$, and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\alpha_f^{(n)}(R_t^{(n)}) = \frac{1}{\sqrt{n}}\sum_{r=1}^{n}\alpha_f^{(n)}(r) = \sqrt{n}\mathrm{E}_f[J_f(\varepsilon)] = 0.$$

Note that extremely little has been assumed about the exogeneous covariates.

Instead of the exact scores $\alpha_f^{(n)}(r)$, one may prefer using the approximate scores; (4.5) and (4.6) remain asymptotically equivalent, provided that the centred version

$$\alpha_f^{(n)}(r) = J_f\left(F^{-1}\left(\frac{r}{n+1}\right)\right) - \frac{1}{n}\sum_{s=1}^{n} J_f\left(F^{-1}\left(\frac{s}{n+1}\right)\right)$$

is considered.

If the model contains a constant term, that is, if the first component of $\mathbf{X}_i^{(n)}$, say, equals 1 for all i, then the efficient central sequence for this component equals zero. This is again due to unidentifiability. However, the efficient central sequence for the other components equals the one in the parametric model where the constant term is considered to be a nuisance parameter. In that sense, adaptive estimation of the other components is possible (cf. Bickel 1982). Under a symmetry assumption ($f \in \mathcal{F}_+$), adaptiveness holds for all parameters. Under a weaker zero median assumption ($f \in \mathcal{F}_0$), the constant term is identified and consistently (but not adaptively) estimable, whereas adaptivity holds for all other components. Again, adaptiveness and the underlying invariance structure of the model under study are very closely related.

The following is an example of a time series model fitting into our framework.

Example 4.3 *Autoregressive models*. For simplicity we consider the AR(1) model

$$Y_t^{(n)} = \theta Y_{t-1}^{(n)} + \varepsilon_t^{(n)}, \qquad t = 1, \ldots, n,$$

where $\theta \in$: i $\Theta = (-1, 1)$, with starting value $Y_0^{(n)} = 0$, and i.i.d. innovations $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$, with density f. Assuming that f is absolutely continuous, with finite second-order moments and finite Fisher information for location ($\int (f'(z)/f(z))^2 f(z)\mathrm{d}z < \infty$), it is well known (Swensen 1985; Kreiss 1987) that this model is ULAN, with central sequence $\Delta_f^{(n)}(\theta)$ given in (3.18). The score function J_f thus actually involves infinitely many lagged residuals. However, the results of Section 3 remain valid in view of Remark 3.2. In this case, $\Delta_f^{(n)}(\theta)$ also decomposes into $\Delta_f^{(n)}(\theta) = \sqrt{n}\sum_{i=1}^{n-1}\theta^i r_{f;i}^{(n)}$, where $r_{f;i}^{(n)} := 1/n\sum_{t=i+1}^{n}\phi_f(\varepsilon_t^{(n)}(\theta))\varepsilon_{t-i}^{(n)}(\theta)$ can be interpreted as a measure of serial dependence at lag i among residuals, a form of generalized residual autocorrelation, adapted to the innovation density f (see Hallin and Werker 1999, for details).

If no further assumption is made about $f \in \mathcal{F}$ – more particularly, if f has arbitrary mean μ_f – then the group of order-preserving transformations $\mathcal{G}^{(n)}, \circ$ described in (3.3) is a generating group in the sense of Section 3, with maximal invariant the ranks $(R_1^{(n)}, \ldots, R_n^{(n)})$ of the residuals $\varepsilon_t^{(n)}(\theta)$. Denoting by f_1 the standardized version of f and by F_1 the corresponding distribution function, consider the so-called rank-based residual autocorrelations

$$\underset{\sim}{r}^{(n)}_{f;i} := \left[\frac{1}{n-i} \sum_{t=i+1}^{n} \phi_{f_1}\left(F_1^{-1}\left(\frac{R_t^{(n)}}{n+1} \right) \right) F_1^{-1}\left(\frac{R_{t-i}^{(n)}}{n+1} \right) - m_f^{(n)} \right] (\sigma^{(n)}_{f;i})^{-1} \tag{4.7}$$

introduced in Hallin and Puri (1991) ($m_f^{(n)}$ and $\sigma^{(n)}_{f;i}$ are the exact mean and standard error respectively, so that $\mathrm{E}^{(n)}_{\theta,f}[\underset{\sim}{r}^{(n)}_{f;i}] = 0$ and $\mathrm{E}^{(n)}_{\theta,f}[(n-i)(\underset{\sim}{r}^{(n)}_{f;i})^2] = 1$). It is easily seen that

$$\underset{\sim}{r}^{(n)}_{f;i} = \mathrm{E}^{(n)}_{\theta,f}[r^{(n)}_{f;i} | R_1^{(n)}, \ldots, R_n^{(n)}] + o_{\mathrm{P}}(1/\sqrt{n}),$$

under $\mathrm{P}^{(n)}_{\theta,f}$, so that Proposition 3.1 and Corollary 3.2 imply that

$$\underset{\sim}{r}^{(n)}_{f;i} = \frac{1}{n} \sum_{t=i+1}^{n} \phi_f(\varepsilon_t^{(n)}(\theta))(\varepsilon^{(n)}_{t-i}(\theta) - \mu_f) + o_{\mathrm{P}}(1/\sqrt{n}),$$

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

$$\underset{\sim}{\Delta}^{(n)}_f(\theta) := \sqrt{n} \sum_{i=1}^{n-1} \theta^{i-1} \underset{\sim}{r}^{(n)}_{f;i}$$

is a rank-based efficient (at θ and f) central sequence. Clearly, $\underset{\sim}{\Delta}^{(n)}_f(\theta)$ and $\Delta^{(n)}_f(\theta)$ do not coincide (asymptotically), unless $\mu_f = 0$. This confirms the fact that the semi-parametric model under which f remains totally unspecified within $\mathcal{F}$ is not adaptive (Drost *et al*. 1997), as well as the results of Sections 3.3 and 3.4 that parametrically efficient rank-based inference is impossible in non-adaptive models.

Now, if a location parameter is specified for f, adaptivity and rank-based efficiency are recovered for θ. If, for instance, the usual assumption that $\mu_f = 0$ is made, then $\underset{\sim}{\Delta}^{(n)}_f(\theta)$ and $\Delta^{(n)}_f(\theta)$ are asymptotically equivalent. Similarly, if the median of f is assumed to be zero (i.e., $f \in \mathcal{F}_0$), then the maximal invariant σ-field (at θ) is generated by the ranks $R_1^{(n)}, \ldots, R_n^{(n)}$ and the signs $s_1^{(n)}, \ldots, s_n^{(n)}$, and sign-and-rank versions of the residual autocorrelations (4.7) can be used in the construction of an invariant version of $\Delta^{(n)}_f(\theta)$; see Hallin *et al*. (2002) for details. If f is assumed to be symmetric with respect to the origin ($f \in \mathcal{F}_+$), the same conclusions hold for the signed-rank autocorrelations defined in Hallin and Puri (1991).

Finally, adaptive rank-based residual correlograms, consisting of autocorrelations of the form $\underset{\sim}{r}^{(n)}_{\hat{f}_n;i}$, can be constructed along the lines of Proposition 3.4, leading to conditionally distribution-free, semi-parametrically efficient inference; see Hallin and Werker (1999).

We conclude this section with an example of a simple ARCH model.

Example 4.4 *ARCH models*. Consider the ARCH(1) model given by

$$Y_t^{(n)} = \sqrt{1 + \theta(Y^{(n)}_{t-1})^2}\, \varepsilon_t^{(n)}, \tag{4.8}$$

where $\varepsilon_1^{(n)}, \ldots, \varepsilon_n^{(n)}$ are i.i.d. square-integrable random variables with density f and, in order to guarantee the existence of a stationary solution, $0 < \theta < 1/\int x^2 f(x)\mathrm{d}x$. We assume that

$Y_0^{(n)} = 0$. In this pure scale model, absolute continuity of $xf(x)$ guarantees that the ULAN condition is satisfied if

$$\int ([xf(x)]'/f(x))^2 f(x)\mathrm{d}x < \infty.$$

This follows, for example, from Drost *et al.* (1997, Theorem 2.1), combined with Hájek *et al.* (1999, Theorem 7.2.2.1). The central sequence is given by

$$\Delta_f^{(n)}(\theta) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_f(\varepsilon_t^{(n)}(\theta)) \frac{(Y_{t-1}^{(n)})^2}{1 + \theta(Y_{t-1}^{(n)})^2},$$

where $\psi_f(x) := -[xf(x)]'/f(x)$ and $\varepsilon_t^{(n)}(\theta) := Y_t^{(n)}/\sqrt{1 + \theta(Y_{t-1}^{(n)})^2}$, $t = 1, \ldots, n$. Iterating (4.8) yields $(Y_{t-1}^{(n)})^2 = \sum_{i=1}^{t-1} \theta^{i-1} \prod_{j=1}^{i} (\varepsilon_{t-j}^{(n)})^2$, which shows that the function J_f actually involves infinitely many lagged residuals. This again, much as in the AR model of Example 4.3, does not affect our general results. In the present case, letting

$$e_{t-1}(p) := \frac{(Y_{t-1}^{(n)})^2}{1 + \theta(Y_{t-1}^{(n)})^2} - \frac{\sum_{i=1}^{p} \theta^{i-1} \prod_{j=1}^{i} (\varepsilon_{t-j}^{(n)})^2}{1 + \theta \sum_{i=1}^{p} \theta^{i-1} \prod_{j=1}^{i} (\varepsilon_{t-j}^{(n)})^2},$$

we have that $|e_{t-1}(p)| \leqslant \sum_{i=p+1}^{t-1} \theta^{i-1} \prod_{j=1}^{i} (\varepsilon_{t-j}^{(n)})^2$. Hence, writing $\sigma^2 := \int x^2 f(x)\mathrm{d}x$, we obtain

$$\mathrm{E}_0^{(n)} |e_{t-1}(p)| \leqslant \frac{\sigma^2}{1 - \theta\sigma^2} \left(\theta\sigma^2\right)^{p+1},$$

which, as for the autoregressive model, converges to zero at a geometric rate, uniformly in n.

Somewhere semi-parametrically efficient inference on θ should, consequently, be based on

$$\underset{\sim}{\Delta}_f^{(n)}(\theta) := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\psi_f\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1} \right) \right) - \overline{\psi}_f^{(n)} \right] \frac{\sum_{i=1}^{p} \theta^{i-1} \prod_{j=1}^{i} F^{-1}(R_{t-j}^{(n)}/(n+1))^2}{1 + \theta \sum_{i=1}^{p} \theta^{i-1} \prod_{j=1}^{i} F^{-1}(R_{t-1}^{(n)}/(n+1))^2},$$

where

$$\overline{\psi}_f^{(n)} := \frac{1}{n} \sum_{t=1}^{n} \psi_f\left(F^{-1}\left(\frac{R_t^{(n)}}{n+1} \right) \right),$$

and p is suitably large.

Acknowledgements

The research for this paper was largely carried out while the second author was at the Université Libre de Bruxelles. The research of the first author was supported by a PAI of the Belgian federal government and an Action de Recherche Concertée contract of the Communauté française de Belgique. The authors very gratefully acknowledge the careful reading of the initial version of this paper by two anonymous referees, whose comments helped improve both the presentation and the results. Thanks are due to Anton Schick for his pertinent remarks about the uniform integrability of central sequences.

References

Beran, R. (1974) Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.*, **2**, 63–74.

Bickel, P.J. (1982) On adaptive estimation. *Ann. Statist.*, **10**, 647–671.

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

Chernoff, H. and Savage, I.R. (1958) Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.*, **29**, 972–994.

Choi, S., Hall, W.J. and Schick, A. (1996) Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.*, **24**, 841–861.

Drost, F.C., Klaassen, C.A.J. and Werker, B.J.M. (1997) Adaptive estimation in time-series models. *Ann. Statist.*, **25**, 786–818.

Falk, M. (1986) On the estimation of the quantile density function. *Statist. Probab. Lett.*, **4**, 69–73.

Hájek, J. (1962) Asymptotically most powerful rank-order tests. *Ann. Math. Statist.*, **33**, 1124–1147.

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

Hájek, J., Šidák, Z. and Sen, P.K. (1999) *Theory of Rank Tests*, 2nd edn. New York: Academic Press.

Hallin, M. (1994) On the Pitman-nonadmissibility of correlogram-based methods. *J. Time Ser. Anal.*, **15**, 607–612.

Hallin, M. and Puri, M.L. (1991) Time series analysis via rank-order theory: signed-rank tests for ARMA models. *J. Multivariate Anal.*, **39**, 175–237.

Hallin, M. and Puri, M.L. (1994) Aligned rank tests for linear models with autocorrelated error terms. *J. Multivariate Anal.*, **50**, 1–29.

Hallin, M. and Werker, B.J.M. (1999) Optimal testing for semiparametric AR models: from Lagrange multipliers to autoregression rank scores and adaptive tests. In S. Ghosh (ed.), *Asymptotics, Nonparametrics and Time Series*, pp. 295–350. New York: Marcel Dekker.

Hallin, M., Ingenbleek, J.-F. and Puri, M.L. (1985) Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.*, **13**, 1156–1181.

Hallin, M., Vermandele, C. and Werker, B.J.M. (2002) Sign and rank-based efficient inference for the median autoregressive model. Unpublished manuscript.

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

Jeganathan, P. (1997) On asymptotic inference in linear cointegrated time series systems. *Econometric Theory*, **13**, 692–745.

Klaassen, C.A.J. (1987) Consistent estimation of the influence function of locally asymptotically linear estimates. *Ann. Statist.*, **15**, 1548–1562.

Koul, H.L. and Schick, A. (1996) Adaptive estimation in a random coefficient autoregressive model, *Ann. Statist.*, **24**, 1025–1052.

Koul, H.L. and Schick, A. (1997) Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli*, **3**, 247–277.

Kreiss, J.P. (1987) On adaptive estimation in stationary ARMA processes. *Ann. Statist.*, **15**, 112–133.

Kreiss, J.P. (1990) Local asymptotic normality for autoregression with infinite order. *J. Statist. Plann. Inference*, **26**, 185–219.

Le Cam, L. and Yang, G.L. (1988) On the preservation of local asymptotic normality under information loss. *Ann. Statist.*, **16**, 483–520.

Lehmann, E.L. (1986) *Testing Statistical Hypotheses*, 2nd edn. New York: Wiley.

Müller, U. and Wefelmeyer, W. (2001) Autoregression, estimating functions, and optimality criteria. Preprint, Universität Siegen.

Puri, M.L. and Sen, P.K. (1985) *Nonparametric Methods in General Linear Models*. New York: Wiley.

Schick, A. and Wefelmeyer, W. (2000) Efficient estimation in invertible linear processes. Preprint, Universität Siegen.

Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

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

van der Vaart, A. (1988) Estimating a real parameter in a class of semiparametric models. *Ann. Statist.*, **16**, 1450–1474.

Yang, S.S. (1985) A smooth nonparametric estimation of the quantile function. *J. Amer. Statist. Assoc.*, **80**, 1004–1011.

Zeltermann, D. (1990) Smooth nonparametric estimation of the quantile function. *J. Statist. Plann. Inference*, **26**, 339–352.

Received April 2000 and revised March 2002