Estimations and Optimal Tests in Some Parametric Models

Salim Bouzebda[1] and Tewfik Lounis[1]

[1]Alliance Sorbonne Universités, Université de Technologie de Compiègne,
Laboratoire de Mathématiques Appliquées de Compiègne,
salim.bouzebda@utc.fr, tewfik.lounis@gmail.com

ABSTRACT. In the present paper, we introduce an efficient method for the estimation in the multidimensional case. The key idea is based on a good assessment of the error without using confidence intervals. The consistency of the proposed estimate is established. Consequently, we discuss the estimation in statistical tests corresponding to parametric context, and prove that this kind of estimators ensures the optimality of statistical tests. We partially extend the scope of our study to some processes. In order to examine the performance of our methodology, finite sample results are performed. This work completes and extends in nontrivial way the results obtained by Lounis (2017).
AMS Classifications: primary: 62G30; 62G20; 60F17; secondary: 62F03; 62F12; 60F15.
KEYWORDS. ARCH Models, Contiguity, Le Cam's third lemma, Local asymptotic normality, Modified estimators, Time series models.

1. Introduction and Motivation

Parametric estimation has been the subject of intense investigation for many years and this has led to the development of a large variety of methods. Because of numerous applications and their important role in mathematical statistics, the problem of estimating the parametric models has been the subject of considerable interest during the last decades. For good sources of references to research literature in this area along with statistical Pfanzagl (1994) Lindsey (1996), Bickel *et al.* (1998), Lehmann and Casella (1998), van der Vaart (1998), Lehmann and Romano (2005) and Cheng (2017). Assume the model can be parameterized as $\boldsymbol{\theta} \mapsto \mathbb{P}_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta}$ is a Euclidean parameter. Our main interest is estimation and inference for the parameter of interest $\boldsymbol{\theta}$. Various parametric methods of estimation have been extensively investigated, among others, including the method of moments, Least Square Estimator (LSE), Maximum of Likelihood (ML) and Delta method. Attention was confined to parametric models and much effort has been expended in constructing efficient estimators. Consequently, applications were provided in several mathematical fields, such as in testing problems. Recall that, large variety of tests statistic and their power functions are expressed in terms of unknown parameter. Often, estimation of this parameter induces an error which deteriorate the asymptotic power. In general, most of these tests are not asymptotically distribution free (ADF). The obtaining of an explicit expression of the power functions is often more complicated. To be more precise, computation of their limiting distribution under alternative hypothesis is more difficult. But, when the null and alternative hypothesis are contiguous, thus will be deduced by making use of the Le Cam's third Lemma. In practice, the direct proof of the contiguity is delicate. However, this property follows when the local asymptotic normality (LAN) is established. It is worth to notice that several versions of LAN exist in literature, among others, we cite the versions of Le Cam (1960), Hall and Mathiason (1990), Swensen (1985), Hwang and Basawa (2001, 2003) and the references therein. For excellent resource of books the interested reader may refer to Roussas (1972), Le Cam (1974, 1986), van der Vaart (1998) and Shiryaev and Spokoiny (2000). Notice that with the contribution of Le Cam's LAN methodology, optimality of test statistics was obtained in some research works, we may refer to Linton (1993), for related topics we refer to Drost *et al.* (1997), Robinson (2005), Koul and Ling

(2006) and the references therein. For this same purpose, a modified estimator (ME) was introduced in Lounis (2011), that has some limitations in practical considerations. To be more precise, let us recall the corresponding definitions and notations. Let $Y_1, \dots, Y_n, \dots, Y_N$ be observations from a random variable Y with a probability $\mathbb{P}_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta}$ a parameter in an open subset $\boldsymbol{\Theta}$ of $\mathbb{R}^d$. $\mathbf{V}_n$ denotes a real random function which will be assumed to be defined and differentiable around the parameter $\boldsymbol{\theta}$. Suppose that the following identity will be satisfied

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_n + o_{\mathbb{P}}(1), \tag{1.1}$$

where $\mathcal{D}_n$ is a specified random function and $\widehat{\boldsymbol{\theta}}_n$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}$. Under (1.1), in Lounis (2011), it has been shown the existence of another estimator $\overline{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ such that we have

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right) + o_{\mathbb{P}}(1). \tag{1.2}$$

Here $o_{\mathbb{P}}(1)$ denotes a random variable converging in probability to 0 as $n \to \infty$. The true meaning of identity (1.2) is the absorption of the error $\mathcal{D}_n$ and the equivalence (up to $o_{\mathbb{P}}(1)$ term) between the random variable $\mathbf{V}_n(\boldsymbol{\theta})$ and its estimated version $\mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right)$. This plays an instrumental role in the asymptotic theories, such as in the proof of the optimality in statistical test problems. Observe that in the equality (1.1), $\mathcal{D}_n$ is assumed known. But in practice, $\mathcal{D}_n$ is, in general, unspecified. This corresponds to the first difficulty. Notice that Taylor expansion is a useful tool for the specifying of the quantity $\mathcal{D}_n$. In general, this last quantity depends on an asymptotically no degenerate error $e_n = \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$. Based on confidence intervals, it has been proved in Lounis (2017) that e_n is equivalent to $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_N)$, where $N = \lfloor n^{1+\mu} \rfloor + 1$, μ denotes some positive constant, and the script "$\lfloor \cdot \rfloor$" corresponds to the integer part. The main drawback of this methodology is its dependence on a level confidence. This also presents an additional problem. Recall that study of consistency of an estimator is the first step and plays a major role. In Lounis (2011), consistency has not been investigated in a precise and detailed way. A sufficient condition was stated, and its checking remains difficult in practice. In this sense, we propose a new method to obtain a modified consistent estimator. Lastly, the random variable $\mathbf{V}_n$ is real. This reduces the possibilities of applications and more specifically in the multidimensional case (i.e., $\mathbf{V}_n$ takes value in $\mathbb{R}^d$). This corresponds to the forth difficulty. Notice that the constructing of ME mainly follows two steps. The first one consists to specify $\mathcal{D}_n$. The second one is the avoiding of the effect of this error. The main purpose of this paper is to provide solutions to the previously stated problems. Consequently, applications are provided in estimation theory and testing problems. First, without using confidence intervals, we develop a new and simple method for the evaluation of the error

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}).$$

This enables us to specify $\mathcal{D}_n$. Subsequently, we extend our results to the case where $\mathbf{V}_n$ takes value in $\mathbb{R}^d$ and we show that our method is also valid when the sample size is equal to N, the whole sample size. Finally, an explicit form of ME based on all observations is given. In addition, a method to show the consistency of our estimator is developed. Note that the construction of ME in the multidimensional case is a direct consequence of Gauss elimination method. As application of the preceding results, we discuss the estimation of unknown parameters in statistical tests. We therefore prove one of the most important property, which is the optimality.

We now explain the link between the ME and statistical testing problems. We first note that a large variety of classical tests and their power functions are described by a random variable called *central sequence*. This central sequence corresponds to the principal part in logarithm likelihood ratio. As explained previously, we show the existence of an estimator $\overline{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ such that

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right) + o_{\mathbb{P}}(1), \tag{1.3}$$

where $\mathbf{V}_n$ takes value in $\mathbb{R}^d$. It is easy to see that when $\mathbf{V}_n$ is the central sequence, the above equality is just a particular case. This means that the equivalence between central sequences $\mathbf{V}_n(\boldsymbol{\theta})$ and $\mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right)$ is established. Because the power functions are described from central sequences, equality (1.3) suffices to deduce the optimality. In addition, computation of the limiting distribution under alternative hypothesis is one consequence of the contiguity property.

The paper is organized as follows. Section 2 is devoted to describing our main results. Section 3 provides some applications in testing problems and empirical processes. In order to illustrate the performances of the proposed methodology, some finite sample results are reposted in Section 4. Some concluding remarks are given in Section 5. All mathematical developments are relegated to the Section 6.

2. Estimation in multidimensional case

In this section, we present the multidimensional version of the modified estimator. In this context both the parameter $\boldsymbol{\theta}$ and the sequence $\mathbf{V}_n$ take values in $\mathbb{R}^d$. This new version extend the results obtained in Lounis (2011) from a real random sequence $\mathbf{V}_n$ to a multidimensional sequence $\mathbf{V}_n$. We first introduce some notations and state the conditions that we will use in our analysis.

2.1. *Assumptions and notation*

Throughout all the paper, $\mathbb{R}^d$ denotes the d-dimensional real vector space equipped with the Euclidian norm $\|\cdot\|$, and $d \geq 1$ denotes a positive integer. Let $Y_1, \ldots, Y_n$ be observations with a $\mathbb{P}_{\boldsymbol{\theta}}$ that depends on a parameter $\boldsymbol{\theta}$ ranging over some open set $\boldsymbol{\Theta}$ in $\mathbb{R}^d$. We denote by $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_{n,1}, \ldots, \widehat{\boldsymbol{\theta}}_{n,d})$ the $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d)$. Thought, $\mathbf{V}_n$ denotes a random function defined and differentiable around $\boldsymbol{\theta}$ and take values in $\mathbb{R}^d$, defined as follows

$$\begin{aligned} \mathbf{V}_n \ : \ & \mathbb{R}^d \to \mathbb{R}^d \\ & x \mapsto \Big(\mathbf{V}_{n,1}(x), \ldots, \mathbf{V}_{n,d}(x)\Big). \end{aligned}$$

In order to describe our methodology, the following identity will be imposed

(A.1)

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_n + o_{\mathbb{P}}(1), \tag{2.1}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

where $\mathcal{D}_n$ is a random vector such that, $\mathcal{D}_n = (\mathcal{D}_{n,1}, \ldots, \mathcal{D}_{n,d})$ and for each $i = 1, \ldots, d$,

$$\mathcal{D}_{n,i} = O_{\mathbb{P}}(1).$$

We mention that up to an $o_{\mathbb{P}}(1)$ term, the quantity $\mathcal{D}_n$ corresponds to the vector error in the estimation of the random sequence $\mathbf{V}_n(\boldsymbol{\theta})$. We introduce the $d \times d$ matrix $\mathbf{A}_n$ by

$$\mathbf{A}_n(\boldsymbol{\theta}) = \begin{bmatrix} \dfrac{\partial \mathbf{V}_{n,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \cdots & \dfrac{\partial \mathbf{V}_{n,1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_d} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \mathbf{V}_{n,d}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_1} & \cdots & \dfrac{\partial \mathbf{V}_{n,d}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_d} \end{bmatrix}. \tag{2.2}$$

Subsisting $\boldsymbol{\theta}$ by $\widehat{\boldsymbol{\theta}}_n$ into (2.2) gives

$$\mathbf{A}_n\left(\widehat{\boldsymbol{\theta}}_n\right) = \begin{bmatrix} \dfrac{\partial \mathbf{V}_{n,1}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_1} & \cdots & \dfrac{\partial \mathbf{V}_{n,1}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_d} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \mathbf{V}_{n,d}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_1} & \cdots & \dfrac{\partial \mathbf{V}_{n,d}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_d} \end{bmatrix}. \tag{2.3}$$

In a sequel, it will be assumed that:

(A.2) For all positive integer n, the matrix $\mathbf{A}_n\left(\widehat{\boldsymbol{\theta}}_n\right)$ is invertible.

Our goal is to construct another estimate $\overline{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ that make the error $\mathcal{D}_n$ in (2.1) negligible. More precisely, we shall construct an estimate satisfying to the following identity

$$\mathbf{V}_n(\boldsymbol{\theta}) \;=\; \mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right) + o_{\mathbb{P}}(1). \tag{2.4}$$

Clearly, equality (2.4) gives an equivalence between the random variable $\mathbf{V}_n(\boldsymbol{\theta})$ and its estimated version $\mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right)$ and this plays an instrumental role in estimation and testing problems. We present this in more details in the next sections.

In addition, we suppose that $\mathbf{V}_n(\boldsymbol{\theta})$ is expressed as follows

$$\mathbf{V}_n(\boldsymbol{\theta})' \;=\; \left(\frac{-1}{\sqrt{n}}\sum_{i=1}^{n}\Upsilon_1(\boldsymbol{\theta}, \mathbf{Y}_i), \ldots, \frac{-1}{\sqrt{n}}\sum_{i=1}^{n}\Upsilon_d(\boldsymbol{\theta}, \mathbf{Y}_i)\right), \tag{2.5}$$

here $\Upsilon_j, \;\; j = 1, \ldots, d$ is a function defined from $\mathbb{R}^d$ to $\mathbb{R}$ and differentiable around $\boldsymbol{\theta}$, and $\mathbf{Y}_i = (Y_{i-1}, \ldots, Y_{i-\ell})$.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

2.2. *Main results*

Let $X_1, \ldots, X_n, \ldots, X_N$ be observations from a random variable X with a distribution $\mathbb{P}_\lambda$, where λ is a parameter of Λ in $\mathbb{R}$. We denote by λ_n a $\sqrt{n}$-consistent estimate of λ. Then, we have our fundamental statement.

Proposition 2.1 ***(Evaluation of the error)***
Let n and N be natural positive numbers such that $n = o(N)$. Then, we have

$$\sqrt{n}(\lambda - \lambda_n) = \sqrt{n}(\lambda_N - \lambda_n) + o_{\mathbb{P}}(1).$$

It is worth noticing that this proposition is more general than (Lounis, 2017, Proposition 1) where the result depends on some confidence intervals depending on some level. This is an advantage of Proposition 2.1 that is free of such restriction. Consider again (2.1), and observe that, in a general the quantity $\mathcal{D}_n$ is unknown. In order to specify it, we apply Proposition 2.1 and make use of the following assumption.

(A.3) For each $(i, j) \in \{1, \ldots, d\} \times \{1, \ldots, d\}$, and for all consistent estimate $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$, the random variables

$$\left\{ \frac{1}{\sqrt{n}} \frac{\partial \mathbf{V}_{n,i}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i} \right\} \quad \text{and} \quad \left\{ \frac{1}{\sqrt{n}} \frac{\partial^2 \mathbf{V}_{n,i}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} \right\}$$

are bounded in probability.

We state the following proposition.

Proposition 2.2 ***(Equivalence between true and estimated version sequences)***
Let n and N be natural positive numbers such that $n = o(N)$. Suppose that $(\mathbf{A.1})$-$(\mathbf{A.3})$ hold. Then, for all $j \in \{1, \ldots, d\}$,

$$\mathbf{V}_{n,j}(\boldsymbol{\theta}) = \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_{n,j} + o_{\mathbb{P}}(1), \tag{2.6}$$

where

$$\mathcal{D}_{n,j} = \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i} \left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}\right). \tag{2.7}$$

This last statement gives an explicit form of the error vector $\mathcal{D}_n$ and by the way completes the results obtained in Lounis (2011, 2013). Making use of the last result, we construct our ME which satisfies the identity (2.4). This is summarized in the following theorem.

Theorem 2.3 ***(ME when $\mathbf{V}_n : \mathbb{R}^d \mapsto \mathbb{R}^d$)***
Let n and N be natural positive numbers such that $n = o(N)$. Assume that $(\mathbf{A.1})$-$(\mathbf{A.3})$ hold. Then there exists an estimator $\overline{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$, such that

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathbf{V}_n\left(\overline{\boldsymbol{\theta}}_n\right) + o_{\mathbb{P}}(1), \tag{2.8}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

and

$$\overline{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n + \mathbf{A}_n^{-1}\left(\widehat{\boldsymbol{\theta}}_n\right)\mathcal{D}_n. \tag{2.9}$$

We shall now investigate the consistency of the modified estimate. We start by processing the case when $\mathbf{V}_n : \mathbb{R}^d \to \mathbb{R}$. From the Proposition (2.2), it follows that

$$\mathbf{V}_n(\boldsymbol{\theta}) = \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_n + o_{\mathbb{P}}(1),$$

where

$$\mathcal{D}_n = \sum_{i=1}^{d} \frac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}).$$

As explained in Lounis (2011)[Section 1.1], the ME $\overline{\boldsymbol{\theta}}_n = (\overline{\boldsymbol{\theta}}_{n,1}, \ldots, \overline{\boldsymbol{\theta}}_{n,d})$ is deduced from a $\sqrt{n}$-consistency estimator $\widehat{\boldsymbol{\theta}}_n = (\widehat{\boldsymbol{\theta}}_{n,1}, \ldots, \widehat{\boldsymbol{\theta}}_{n,d})$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d)$. This is done by disrupting only one component of $\widehat{\boldsymbol{\theta}}_n$ and keeping the other components unchanged. For instance, if we only disrupt $\widehat{\boldsymbol{\theta}}_n$ in its j-th component, then we obtain an explicit form of $\overline{\boldsymbol{\theta}}_n$ described as

$$\begin{aligned} \overline{\boldsymbol{\theta}}_{n,i} &= \widehat{\boldsymbol{\theta}}_{n,i}, \;\; i \neq j, \;\; i = 1, \ldots, d, \\ \overline{\boldsymbol{\theta}}_{n,j} &= \widehat{\boldsymbol{\theta}}_{n,j} + \vartheta_n, \end{aligned} \tag{2.10}$$

where

$$\vartheta_n = \frac{\mathcal{D}_n}{\left\{\dfrac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_j}\right\}}. \tag{2.11}$$

By the fact that the estimator $\widehat{\boldsymbol{\theta}}_n$ is d-dimensional, it results from this procedure d different ME. The natural question is : Is all these estimators are consistent? The answer is given in the following theorem.

Theorem 2.4 ***(Consistency when*** $\mathbf{V}_n : \mathbb{R}^d \mapsto \mathbb{R}$***)***
Let n and N be natural positive numbers such that $n = o(N)$. Suppose that $(\mathbf{A.1})$ and $(\mathbf{A.3})$ hold. Then the ME $\overline{\boldsymbol{\theta}}_n = (\overline{\boldsymbol{\theta}}_{n,1}, \ldots, \overline{\boldsymbol{\theta}}_{n,d})$ of $\boldsymbol{\theta}$ defined as

$$\begin{aligned} \overline{\boldsymbol{\theta}}_{n,i} &= \widehat{\boldsymbol{\theta}}_{n,i}, \quad i \neq j, \quad\; i \in \{1, \ldots, d\}, \\ \overline{\boldsymbol{\theta}}_{n,j} &= \boldsymbol{\theta}_{n,j} + \vartheta_n \end{aligned}$$

where

$$\vartheta_n = \frac{\mathcal{D}_n}{\max\limits_{1\leq i\leq d}\left\{\dfrac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}\right\}}. \tag{2.12}$$

is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}$.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

This last result means that, to obtain a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}$, it suffices to disrupt $\widehat{\boldsymbol{\theta}}_n$ at the component j of $\widehat{\boldsymbol{\theta}}_n$ that satisfies the following identity

$$\frac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_j} = \max_{1\leq i\leq d}\left\{\frac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}\right\}.$$

Now we state a general result concerning the case when $\mathbf{V}_n : \mathbb{R}^d \mapsto \mathbb{R}^d$.

Theorem 2.5 ***(Consistency when*** $\mathbf{V}_n : \mathbb{R}^d \mapsto \mathbb{R}^d$***)***
Let n and N be natural positive numbers such that $n = o(N)$. Suppose that $(\mathbf{A.1})$-$(\mathbf{A.3})$ hold. Then,

If we assume that

$$\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)\right| \geq n^{\frac{d}{2}},$$

then ME $\overline{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent estimate of $\boldsymbol{\theta}$.

Theorem 2.5 gives a sufficient condition for the obtaining of $\sqrt{n}$-consistency in multidimensional case.

Our result concerns rather a sub samples of size n. It is judicious to study this problem by using all the N observations. As stated in Proposition (2.2), equivalence between $\mathbf{V}_n(\boldsymbol{\theta})$ and $\mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)$ is obtained under assumption that $n = o(N)$. The extension to large samples requires a similar condition. To this end, consider the following decomposition

$$\begin{aligned} N &= \lfloor N^{(1-\eta)}\rfloor + m \\ &=: s + m, \end{aligned} \tag{2.13}$$

where, η is some positive real in $]0;1[$. In a sequel, in order to have $s > m$, we choose η sufficiently small. Let us denote by $\widehat{\boldsymbol{\theta}}_s$, $\widehat{\boldsymbol{\theta}}_m$ and $\widehat{\boldsymbol{\theta}}_N$ the estimators of $\boldsymbol{\theta}$ based respectively on samples $Y_1,\dots,Y_s$, $Y_{1+s},\dots,Y_N$ and $Y_1,\dots,Y_N$. Therefore, we state one of our main results.

Theorem 2.6 ***(Extension to total observations)***
Suppose that $(\mathbf{A.1})$-$(\mathbf{A.3})$ hold. Then, we obtain the following assertions.

1. *We have*

$$\mathbf{V}_{N,j}\left(\boldsymbol{\theta}\right) = \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \mathcal{D}_{s,m,j} + o_{\mathbb{P}}(1), \;\; j = 1,\dots,d. \tag{2.14}$$

with

$$\begin{aligned} \mathcal{D}_{s,m,j} &= -\sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)\right] + \sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{s,i}\right) \\ &\quad + \sqrt{\frac{m}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{m,i}\right) \\ &=: \sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)\right] + \sqrt{\frac{s}{N}}\mathcal{D}_{s,j} + \sqrt{\frac{m}{N}}\mathcal{D}_{m,j}. \end{aligned}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

2. *There exists an estimator $\overline{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$ such that*

$$\mathbf{V}_N(\boldsymbol{\theta}) = \mathbf{V}_N\left(\overline{\boldsymbol{\theta}}_N\right) + o_{\mathbb{P}}(1), \tag{2.15}$$

with

$$\overline{\boldsymbol{\theta}}_N \;=\; \widehat{\boldsymbol{\theta}}_s + \mathbf{A}_N^{-1}\left(\widehat{\boldsymbol{\theta}}_s\right)\mathcal{D}_{s,m}, \tag{2.16}$$

and

$$\mathcal{D}_{s,m} \;=\; \begin{bmatrix} \mathcal{D}_{s,m,1} \\ \vdots \\ \mathcal{D}_{s,m,d} \end{bmatrix}. \tag{2.17}$$

3. *If we have*

$$\left|\det(\mathbf{A}_N\left(\widehat{\boldsymbol{\theta}}_s\right)\right| \geq N^{\frac{d}{2}},$$

then ME $\overline{\boldsymbol{\theta}}_n$ is $\sqrt{N}$-consistent estimate of $\boldsymbol{\theta}$.

Remark 2.7 *Notice that we can obtain $B = \binom{s}{N}$ samples of size s. Let us denote $\widehat{\boldsymbol{\theta}}_s^{(k)}$ the estimate based on the k-th sample, for $k = 1, \ldots, B$ and $\widehat{\boldsymbol{\theta}}_m^{(k)}$ based on the reminder sample. An application of Theorem 2.6, implies, in turn, that we have*

$$\mathbf{V}_{N,j}\left(\boldsymbol{\theta}\right) = \frac{1}{B}\sum_{k=1}^{B}\left\{\mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s^{(k)}\right) + \mathcal{D}_{s,m,j}^{(k)}\right\} + o_{\mathbb{P}}(1), \quad j = 1, \ldots, d,$$

where

$$\begin{aligned}
\mathcal{D}_{s,m,j} \;=\; & -\sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s^{(k)}\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)\right] + \sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s^{(k)}\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{s,i}^{(k)}\right) \\
& + \sqrt{\frac{m}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m^{(k)}\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{m,i}^{(k)}\right) \\
=: \; & \sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m^{(k)}\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s^{(k)}\right)\right] + \sqrt{\frac{s}{N}}\mathcal{D}_{s,j}^{(k)} + \sqrt{\frac{m}{N}}\mathcal{D}_{m,j}^{(k)}.
\end{aligned}$$

Likewise, we have

$$\mathbf{V}_N(\boldsymbol{\theta}) = \mathbf{V}_N\left(\overline{\boldsymbol{\theta}}_N^{(k)}\right) + o_{\mathbb{P}}(1),$$

with

$$\overline{\boldsymbol{\theta}}_N^{(k)} \;=\; \frac{1}{B}\sum_{k=1}^{B}\left\{\widehat{\boldsymbol{\theta}}_s^{(k)} + \mathbf{A}_N^{-1}\left(\widehat{\boldsymbol{\theta}}_s^{(k)}\right)\mathcal{D}_{s,m}^{(k)}\right\},$$

and

$$\mathcal{D}_{s,m}^{(k)} = \begin{bmatrix} \mathcal{D}_{s,m,1}^{(k)} \\ \vdots \\ \mathcal{D}_{s,m,d}^{(k)} \end{bmatrix}.$$

Clearly, this last statement provides an equivalence between the random variable $\mathbf{V}_N(\boldsymbol{\theta})$ and its estimated version $\mathbf{V}_N\left(\overline{\boldsymbol{\theta}}_N\right)$ with using all the N observations. This extends the results obtained in Lounis (2017) where only a subset samples of size n was considered.

Recall that is this constructing, ME is mainly based on decomposition (2.13) and equality (2.14) that will be rewritten as

$$\mathbf{V}_{N,j}(\boldsymbol{\theta}) - \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_N\right) = \mathcal{D}_{s,m,j} + o_{\mathbb{P}}(1), \quad j = 1, \ldots, d.$$

Then, to aim to construct a ME very close to the unknown parameter, it suffices to minimize the error $\mathcal{D}_{s,m}$ in its components $\mathcal{D}_{s,m,j}$ for $j = 1, \ldots, d$. Therefore, we determinate the value of η_0 satisfying the following criterion

$$\eta_0 = \underset{0<\eta<1}{\arg\min} \sum_{j=1}^{d} \left|\mathcal{D}_{s,m,j}\right|, \text{ such that, } s = \lfloor N^{1-\eta} \rfloor \text{ and } s > m. \tag{2.18}$$

Notice that when the two reals η_1 and η_2 may be different, the equality $\lfloor N^{1-\eta_1} \rfloor = \lfloor N^{1-\eta_2} \rfloor$ will be satisfied. This explains that in a general, the value of η_0 is not unique. Moreover, this gives an answer on an issue that was left open in Lounis (2017) and concerning the proper choice of the value of η.

Observe that, in (2.18), η_0 depends on the first partial derivative of the sequence $\mathbf{V}_N$. We will give now an extension of our study to the case when this error is expressed in terms of a high order partial derivatives. In our previous presentation, we have seen that Taylor expansion is a fundamental tool to specify the estimation's error. To be precise, return back to the Theorem 2.6, and observe that for each component $j = 1, \ldots, d$, errors $\mathcal{D}_{s,m,j}$ are expressed in terms of the partial derivatives with order 1 of the random variable $\mathbf{V}_{N,j}$. We now propose to study this problem when the error $\mathcal{D}_{s,m,j}$ is described with higher partial derivatives. Consequently the error of estimation will be reduced and a best rate of convergence of ME will be obtained. We start by processing the case when $\mathbf{V}_n : \mathbb{R}^d \mapsto \mathbb{R}$. After that we extend our results to the case when $\mathbf{V}_n$ takes values in $\mathbb{R}^d$. To this end, it will assumed that $\mathbf{V}_n$ is $p-$differentiable around $\boldsymbol{\theta}$, where p is an integer such that $p \geq 2$. In this setting, assumption **(A.3)** is replaced by the following.

(A.4) For all consistent estimate $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$, for all $\alpha = 1, \ldots, p$ and $i = 1, \ldots, d$, the random variables

$$\left\{ \frac{1}{\sqrt{n}} \frac{\partial^{\alpha} \mathbf{V}_{n,i}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_{i_1} \ldots \partial \boldsymbol{\theta}_{i_k}} \right\},$$

are bounded in probability, where $\sum_{j=1}^{k} i_j = \alpha,$ and $k \geq 1$.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

We need to prove the following technical result.

Proposition 2.8 ***(Evaluation of the error in general case)***
Let n, N and q be natural numbers such that $n = o(N)$ and $q \geq 1$. Then

$$n^{\frac{q}{2}}(\lambda - \lambda_n)^q = n^{\frac{q}{2}}(\lambda_N - \lambda_n)^q + o_{\mathbb{P}}(1).$$

It is easy to see that when $q = 1$, we obtain Proposition 2.1. We are ready to state the following.

Theorem 2.9 **(ME with order p when $\mathbf{V}_n : \mathbb{R}^d \mapsto \mathbb{R}^d$)**
Suppose that $(\mathbf{A.1})$-$(\mathbf{A.2})$ and $(\mathbf{A.1})$ hold. Then, we obtain

1. $$\mathbf{V}_{N,j}(\boldsymbol{\theta}) = \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \widehat{\mathcal{D}}^{(p)}_{s,m,j} + o_{\mathbb{P}}(1), \tag{2.19}$$

   where

   $$\begin{aligned}\widehat{\mathcal{D}}^{(p)}_{s,m,j} &= -\sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)\right] \\ &+\sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{s,i}\right) \\ &+\sqrt{\frac{m}{N}}\widehat{\mathcal{D}}_{s,j},\end{aligned}$$

   and

   $$\widehat{\mathcal{D}}_{s,j} = \boldsymbol{\nabla}\mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)\cdot(\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_s) + \cdots + \frac{1}{(p-1)!}\left(D^{p-1}\mathbf{V}_{s,j}\right)\left(\widehat{\boldsymbol{\theta}}_s\right)\cdot(\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_s)^{p-1}.$$

2. *There exists an estimator $\overline{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$, such that*

   $$\mathbf{V}_N(\boldsymbol{\theta}) = \mathbf{V}_N\left(\overline{\boldsymbol{\theta}}_N\right) + o_{\mathbb{P}}(1),$$

   and

   $$\overline{\boldsymbol{\theta}}_N = \widehat{\boldsymbol{\theta}}_s + \mathbf{A}_N^{-1}\left(\widehat{\boldsymbol{\theta}}_s\right)\widehat{\mathcal{D}}^{(p)}_{s,m,j}.$$

3. *If we assume that*

   $$\left|\det(\mathbf{A}_N\left(\widehat{\boldsymbol{\theta}}_s\right)\right| \geq N^{\frac{d}{2}},$$

   then ME $\overline{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent estimate of $\boldsymbol{\theta}$.

It is possible to propose an estimate based on several samples in a similar way as in Remark 2.7. This will given as follows. From Theorem 2.9, we have

$$\mathbf{V}_{N,j}(\boldsymbol{\theta}) = \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s^{(k)}\right) + \widehat{\mathcal{D}}^{(p)^{(k)}}_{s,m,j} + o_{\mathbb{P}}(1), \tag{2.20}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

where

$$\begin{aligned}
\widehat{\mathcal{D}}^{(p)^{(k)}}_{s,m,j} &= -\sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}^{(k)}_s\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}^{(k)}_m\right)\right] \\
&+\sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}^{(k)}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}^{(k)}_{s,i}\right) \\
&+\sqrt{\frac{m}{N}}\widehat{\mathcal{D}}^{(k)}_{s,j},
\end{aligned}$$

and

$$\widehat{\mathcal{D}}_{s,j} = \nabla \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}^{(k)}_s\right)\cdot(\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}^{(k)}_s) + \cdots + \frac{1}{(p-1)!}\left(D^{p-1}\mathbf{V}_{s,j}\right)\left(\widehat{\boldsymbol{\theta}}^{(k)}_s\right)\cdot(\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}^{(k)}_s)^{p-1}.$$

As an estimator, one can use $\overline{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$, such that

$$\mathbf{V}_N(\boldsymbol{\theta}) = \frac{1}{B}\sum_{i=1}^{B}\mathbf{V}_N\left(\overline{\boldsymbol{\theta}}^{(k)}_N\right) + o_{\mathbb{P}}(1),$$

where

$$\overline{\boldsymbol{\theta}}^{(k)}_N = \widehat{\boldsymbol{\theta}}^{(k)}_s + \mathbf{A}_N^{-1}\left(\widehat{\boldsymbol{\theta}}^{(k)}_s\right)\widehat{\mathcal{D}}^{(p)^{(k)}}_{s,m,j}.$$

3. Some applications

In the first part, we recall that one of the most important property in testing problems is the asymptotic optimality. For discriminating between two models, we use Likelihood Ratio statistic test. Is is worth noticing that from the LAN property the contiguity follows. We then deduce the limiting distribution under alternative hypothesis. Estimating of the unknown parameter in the times series setting often induces an error which may deteriorate the asymptotic power. We then show that the applying of our procedure permits to maintain the test optimality. In the second part, we discuss how our results may be extended the framework of the empirical processes. More precisely, we study the empirical marked process considered in Stute (1997) for testing the hypothesis that a regression function to belong to a specified parametric class of functions. We define a new transformation that gives the equivalence between the process with the true parameter and its estimated version. We introduce a multidimensional version of marked processes and we show then that our method enables to keep the same asymptotic properties of the initial process.

3.1. *LAN property and testing problems*

Based on the approach of quadratic mean differentiability, Hwang and Basawa (1993) established a LAN property for a class of a non linear time series models. The corresponding context is that the error process is not necessarily gaussian. Our main goal in this section, is to apply our results in testing problems derived from this LAN property. In the sequel of this section, we use a notation similar to that used in Hwang and Basawa (1993) including some changes absolutely necessary for our setting. Let $Y_1, Y_2, \ldots$ be a sequence of a random variables on the probability space $(\Omega, \mathbb{F}, \mathbb{P}_\theta)$, generated from the model

$$Y_t \;=\; H_\theta(Y_{t-1}, z_t) + \epsilon_t, \;\text{ for }\; t = 1, 2, \ldots, \tag{3.1}$$

where ϵ_t and z_t are two zero mean processes independent of Y_s where $s < t$. Let ϵ_t be a sequence of i.i.d. random variables and not necessarily gaussian. $H_\theta(\cdot, \cdot)$ is a known function. Some specification of $H_\theta(\cdot, \cdot)$ show that the model (3.1) embodies a large class of non linear times series models. Among others, we cite $RCAR(1)$, $ERA(1)$, $TAR(1)$ and $RCTAR(1)$. For more details, refer to Hwang and Basawa (1993). $\mathbb{P}_\theta$ is a probability measure indexed by an unknown parameter θ of $\boldsymbol{\Theta} \subset \mathbb{R}^d$. $\mathbb{F}_n$ denotes the σ-field generated by $(Y_1, \ldots, Y_n)$ and $\mathbb{P}_{n,\theta}$ corresponds to the restriction of P_θ on $\mathbb{F}_n$.

The log-likelihood ratio is defined by the quantity :

$$\tilde{\lambda}_n = \log(p_{n,\theta_n}/p_{n,\theta}), \quad \theta_n = \theta + h/\sqrt{n},$$

where $p_{n,\theta}$ is the density function corresponding to $\mathbb{P}_{n,\theta}$ and h is a fixed $d \times 1$ vector of real numbers.

For the special case when z_t degenerate at 0, the model (3.1) will be rewritten as follows

$$Y_t \;=\; H_\theta(Y_{t-1}) + \epsilon_t, \;\text{ for }\; t = 1, 2, \ldots \tag{3.2}$$

By imposing the following assumption

(C1) There exists a square-integrable random variable (under H_0) $K_\theta(Y_{t-1})$ and a positive constant c (which may depend on θ) such that for all $\tilde{\theta}$ with $|\tilde{\theta} - \theta| < c$,

$$\left|\boldsymbol{\nabla} H_\theta(t-1)(\tilde{\theta})\right| \leq K_\theta(Y_{t-1}),$$

where $\boldsymbol{\nabla} H_\theta(t-1)(\tilde{\theta})$ represents $\partial H_\theta(t-1)\Big/\partial\theta$ at $\theta = \tilde{\theta}$,

Hwang and Basawa (1993) deduced the LAN property of the model (3.2), under $\mathbb{P}_\theta$- probability. This is precisely described as follows.

$$\lambda_n \;=\; h^\top S_n(\theta_0) - \frac{1}{2} h^\top I(\theta_0) h + o_{\mathbb{P}}(1), \tag{3.3}$$

where

$$\begin{aligned} \lambda_n &= \log\left(L_n(\theta_n)\Big/L_n(\theta_0)\right), \\ S_N(\theta_0) &= -n^{-\frac{1}{2}} \sum_{i=1}^{N} \boldsymbol{\nabla} H_{t-1}(\theta_0) \frac{\dot{f}(\epsilon_i)}{f(\epsilon_i)} : \quad d \times 1 \;\text{ vector}, \\ I(\theta) &= \mathbb{E}_\theta\left[\left(\boldsymbol{\nabla} H_{t-1}(\theta_0)\right)\left(\boldsymbol{\nabla} H_{t-1}(\theta_0)\right)^\top\right] \cdot i_E : \quad d \times d \;\text{ matrix}, \end{aligned} \tag{3.4}$$

where i_E is defined to be $\mathbb{E}\left[\frac{\dot{f}(\epsilon_i)}{f(\epsilon_i)}\right]^2$, $L_N(\theta)$ is the likelihood function of $(Y_1, \ldots, Y_N)$ under $\mathbb{P}_\theta$, conditional on Y_0, and h a fixed $d \times 1$ vector. To test

$$H_0 : \theta = \theta_0$$

versus

$$H_1^N : \theta = \theta_0 + \ h/\sqrt{N},$$

a likelihood statistic test T_N is derived from the LAN property (3.3) of the model (3.2), with

$$T_N = -\log\left(L_N(\theta_0)/L_N(\widehat{\theta}_{ML})\right). \tag{3.5}$$

Here, $\widehat{\theta}_{ML}$ denotes a one step estimator maximum likelihood of θ described as

$$\widehat{\theta}_{ML} \ = \ \theta_N + N^{\frac{-1}{2}} I^{-1}(\theta_N) S_N(\theta_N), \tag{3.6}$$

where θ_N is a preliminary $\sqrt{N}$-consistent estimator of θ. In addition, it has been shown that this test is optimal, in the sense of Hwang and Basawa (1993), pp. 94-95. From Hall and Mathiason (1990), T_N converges in distribution under H_0 and under H_1^N, respectively, to $\chi^2(d)$ and $\chi^2(\mu, d)$. The non-centrality parameter μ is equal to $h' I(\theta_0) h$. By considering the expression (3.7), one remark that the statistic test and its limiting distributions are depending on an parameter θ_0. But, in practical use, this value remains unspecified. Then, we show that it is possible to replace this unknown parameter by an estimator which insures the preservation of the asymptotic optimality. This is treated later in the simulation study.

3.2. *Equivalence of processes in multidimensional case*

In a parametric context, some general methods for testing goodness-of-fit are studied in several papers, among other, we refer to Khmaladze (1981, 1993), Koul and Stute (1999), Khmaladze and Koul (2004), Delgado *et al.* (2005), Koul *et al.* (2005), Stute *et al.* (2006), Escanciano (2007), Alvarez-Andrade and Bouzebda (2014) and references therein. Most of these papers describes a transform whose the corresponding tests are ADF. These methods are complicate in practical situations. This section provides an extension of our methods to the multidimensional processes. Our construction mainly uses some results obtained in Stute (1997) corresponding to unidimensional marked empirical process. Our technical argument consists to construct an estimate which insures the equivalence between the multidimensional marked process and its estimated version. We also show that our transformation enables to obtain limiting laws with are distribution free. These results are particularly important because our transformation shares the same properties with the initial processes and can be easily used for performing statistical tests. The application of the Proposition 2.1 permits to achieve successfully our construction. In the sequel of this section, we use a notation similar to that used in Stute (1997) including some changes. Through (X, Y) denotes a random vector in $\mathbb{R}^{d+d}$ equipped by its appropriate Euclidian norm. It will assumed that $Y = (Y_1, \ldots, Y_d)$ is integrable with its associate regression function, for $x \in \mathbb{R}^d$,

$$\begin{aligned} m(x) \ &= \ \Big(m_1(x), \ldots, m_d(x)\Big), \\ &=: \ \Big(\mathbb{E}\left[Y_1 | X = x\right], \ldots, \mathbb{E}\left[Y_d | X = x\right]\Big). \end{aligned}$$

Stute (1997) introduced a marked empirical process based on residuals in order to test the hypothesis that the regression function $m(\cdot)$ is assumed to belong to a given parametric family

$$\mathcal{M} = \Big\{ m(\cdot, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta} \Big\},$$

Here $\boldsymbol{\Theta}$ is a subset of $\mathbb{R}^d$. This is also equivalent to test

$$H_0 : m \in \mathcal{M} \ \text{ versus } \ H_1 : m \notin \mathcal{M}. \tag{3.7}$$

Let us introduce the multidimensional standardized process $\mathcal{R}_n(x)$ by the following, for $x \in \mathbb{R}^d$,

$$\mathcal{R}_n(x) \;=\; \Big(\mathcal{R}_{n,1}(x), \ldots, \mathcal{R}_{n,d}(x) \Big), \tag{3.8}$$

where, for $j = 1, \ldots, d$,

$$\begin{aligned} \mathcal{R}_{n,j}(x) \;&=\; \mathcal{R}_{n,j}(x; \boldsymbol{\theta}) \\ &=\; n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} \Big[Y_{j,i} - m_j(X_i, \boldsymbol{\theta}) \Big]. \end{aligned} \tag{3.9}$$

Note that $Y_{j,1,}, \ldots, Y_{j,n}$ is a sample of size n from the random variable Y_j, $j = 1, \ldots, d$. As explained in Stute (1997) and by setting

$$\mathcal{R}_{n,j}(-\infty) \;=\; 0,$$

and

$$\mathcal{R}_{n,j}(x) \;=\; n^{-1/2} \sum_{i=1}^{n} \Big[Y_{j,i} - m_j(X_i) \Big], \tag{3.10}$$

each component of the multidimensional process $\mathcal{R}_n(x)$ is continuously extended to $-\infty$ and $+\infty$. Notice that $\mathcal{R}_{n,j}(x)$ is a sum of an i.i.d. random variables conditionally centered at X_i with variances

$$T_j(x) = \int_{-\infty}^{x} \operatorname{Var}(Y_j \mid u) F(du). \tag{3.11}$$

An application of the CLT, gives

$$\mathcal{R}_{n,j}(x) \stackrel{\mathcal{D}}{\to} \mathcal{N}\left(0, T_j(x)\right). \tag{3.12}$$

By replacing $\boldsymbol{\theta}$ by its $\sqrt{n}$ consistent estimates $\widehat{\boldsymbol{\theta}}_{j,n} = \left(\widehat{\boldsymbol{\theta}}_{j,n,1}, \ldots, \widehat{\boldsymbol{\theta}}_{j,n,d} \right)$ in the equality (3.9), we find the estimated version of these empirical processes described by

$$\mathcal{R}^{\mathbb{1}}_{n,j}(x) \;=\; \mathcal{R}_{n,j}(x)(\widehat{\boldsymbol{\theta}}_{j,n})$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

$$= \quad n^{-1/2}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\leq x\}}\Big[Y_{j,i}-m_j(X_i,\widehat{\boldsymbol{\theta}}_{j,n})\Big],\ \ j=1,\ldots,d. \tag{3.13}$$

Testing problem (3.7), is equivalent to

$$H_0:\ \ m_j=m_j(\cdot,\boldsymbol{\theta}_0)\ \text{ versus }\ H_1:\ \ m_j\neq m_j(\cdot,\boldsymbol{\theta}_0),\ \ j=1,\ldots,d, \tag{3.14}$$

where $\boldsymbol{\theta}_0$ an interior point in $\boldsymbol{\Theta}$. We start by establishing a link between $\mathcal{R}_n(x)$ and its estimated version $\mathcal{R}_n^{\mathbb{1}}(x)$. By using a same assumptions as those used in Stute (1997), we treat separately each component $\mathcal{R}_{n,j}$. After that, we present our main results.

3.3. *Assumptions and main results*

The following assumptions are required. Let $F(\cdot)$ denote the unknown distribution function of Y.

(D.1) We assume that

$$\mathbb{E}Y_j^2<\infty,\ \text{for } j\in\{1,\ldots,d\}.$$

(D.2) (i) For all $j\in\{1,\ldots,d\}$, $m_j(x,\boldsymbol{\theta})$ is continuously differentiable at each $\boldsymbol{\theta}$ an interior point in $\boldsymbol{\Theta}$.

(ii) Set

$$g_j(x,\boldsymbol{\theta})=\frac{\partial m_j(x,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\Big(g_{j,1}(x,\boldsymbol{\theta}),\ldots,g_{j,d}(x,\boldsymbol{\theta})\Big)^{\top}.$$

There exists an F- integrable functions $M_j(x)$ such that

$$|g_{j,i}x,\boldsymbol{\theta})|\leq M_j(x),\ \ (i,j)\in\{1,\ldots,d\}\times\{1,\ldots,d\}.$$

We denote by

$$G_j(x,\boldsymbol{\theta})=\Big(G_{j,1}(x,\boldsymbol{\theta}),\ldots,G_{j,d}(x,\boldsymbol{\theta})\Big)^{\top},$$

where,

$$G_{j,i}(x,\boldsymbol{\theta})=\int_{-\infty}^{x}g_{j,i}(u,\boldsymbol{\theta})F(du),\ \ (i,j)\in\{1,\ldots,d\}\times\{1,\ldots,d\}. \tag{3.15}$$

Now we are ready to state our first result in the multidimensional case.

Theorem 3.1 ***(Explicit link between multidimensional empirical marked process ant its estimated version)***
Suppose that ***(D.1)-(D.2)*** *are satisfied. Then under $H_0: m_j=m_j(\cdot,\boldsymbol{\theta}_0)$, we have, uniformly in x,*

$$\mathcal{R}_{N,j}(x)\quad=\quad\mathcal{R}_{N,j}^{\mathbb{1}}(x)+\mathbf{G}_j\left(x,\widehat{\boldsymbol{\theta}}_{j,s},\widehat{\boldsymbol{\theta}}_{j,m}\right)+o_{\mathbb{P}}(1),\quad j=1,\ldots,d,, \tag{3.16}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

where

$$\mathbf{G}_j\left(x,\widehat{\boldsymbol{\theta}}_{j,s},\widehat{\boldsymbol{\theta}}_{j,m}\right) = \sqrt{\frac{s}{N}}\left[s^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N}-\widehat{\boldsymbol{\theta}}_{j,s})^\top\cdot G_j(x,\widehat{\boldsymbol{\theta}}_{j,s})\right] + \sqrt{\frac{m}{N}}\left[m^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N}-\widehat{\boldsymbol{\theta}}_{j,m})^\top\cdot G_j(x,\widehat{\boldsymbol{\theta}}_{j,m})\right].$$

Theorem 3.1 is similar as Theorem 1.2 of Stute (1997). In our work, equality (3.16) is not depending on an unknown parameter. Under similar assumptions as in the Theorem 3.1, we are ready to state our more main result.

Theorem 3.2 ***(Equivalence between processes)***
For each $j\in\{1,\ldots,d\}$, there exists a transformation $S=S_{d_j}$ from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that

$$\begin{aligned}\mathcal{R}_{N,j}(x)(\boldsymbol{\theta}) &= \mathcal{R}_{N,j}(x)(\bar{\boldsymbol{\theta}}_{j,N}) + o_{\mathbb{P}}(1)\\ &:= \mathcal{R}_{N,j}(x)\left(S_{d_j}(\widehat{\boldsymbol{\theta}}_N)\right) + o_{\mathbb{P}}(1),\end{aligned} \tag{3.17}$$

where

$$\begin{aligned}S_{d_j}(\widehat{\boldsymbol{\theta}}_N) &= \widehat{\boldsymbol{\theta}}_N + P,\\ P^\top &= (p_1,\ldots,p_d),\end{aligned}$$

and

$$p_{d_j} = \frac{\mathbf{G}_j\left(x,\widehat{\boldsymbol{\theta}}_{j,s},\widehat{\boldsymbol{\theta}}_{j,m}\right)}{-N^{\frac{-1}{2}}\sum_{i=1}^N \dfrac{\partial m_j(X_i,\widehat{\boldsymbol{\theta}}_{j,m})}{\partial\theta_{d_j}}\mathbb{1}_{\{X_i\le x\}}}, \tag{3.18}$$

$$p_i = 0, \text{ for } i\in\{1,\ldots,d\}\backslash\{d_j\}.$$

The notation S_{d_j} above means that with a quantity p_{d_j}, the first estimator $\widehat{\boldsymbol{\theta}}_{j,N}$ was only disrupted at the component d_j. This transformation is no more than a translation in a component (the d_j-th one).

Clearly (3.17) provides an equivalence between the considered process $\mathcal{R}_{N,j}(x)(\boldsymbol{\theta})$ and its estimated version $\mathcal{R}_{N,j}(x)(\bar{\boldsymbol{\theta}}_{j,N})$. This is very useful in testing problems as we can see in the following. Observe that in this case, the disruptions described in (3.18) are also processes. Notice that the results of this section may be completed by a deep investigation in order to give a full characterization of the modified processes. This needs some results about the tightness of the modified processes that can studied in a similar as in Stute (1997).

3.4. *Testing problem*

Return back to testing problem (3.7) considered in Stute (1997), that we will describe in simple situation, for fixed x. As mentioned already in (3.12), for each $j=1,\ldots,d$, we have, as N tends to ∞,

$$\mathcal{R}_{N,j}(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0,T_j(x)). \tag{3.19}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Corresponding to the case when $d = 1$, we deduce that $\mathcal{R}_N(x) = \mathcal{R}_{N,1}(x)$ depends on the unknown parameter $\boldsymbol{\theta}$. Notice that the last convergence in (3.19), is given in Stute (1997), in terms of convergence to some gaussian processes. This will be of interest to investigate this result in our setting, this requires non trivial mathematics that we leave for future work. From (3.17) and by applying Slutsky's Lemma, uniformly in x, $\mathcal{R}_{N,1}(x)(\bar{\boldsymbol{\theta}}_{1,N})$ has a same limiting distribution as $\mathcal{R}_{N,1}(x)(\boldsymbol{\theta})$. Then H_0 is rejected when $\mathcal{R}_{N,1}(x)(\bar{\boldsymbol{\theta}}_{1,N})$ exceeds a critical value. Now we focus our interest on testing the hypothesis given in (3.14). Notice that the limiting distribution of $\mathcal{R}_N(x)$ is unknown. In general, the fact that each component $\mathcal{R}_N(x)$ follows a gaussian distribution is insufficient to characterize the limiting law of $\mathcal{R}_N(x)$. Then, in order to study this problem, it is required that

(D.3) we have

$$\mathrm{Cov}(Y_i, Y_j) = 0, \text{ for } 1 \leq i, j \leq d, \text{ and } i \neq j.$$

Under **(D.1)** $-$ **(D.3)**, we have, for fixed x,

$$\mathbb{T}_N = \sum_{i=1}^{d} \frac{\mathcal{R}_{N,i}^2(x)}{T_i(x)} \xrightarrow{\mathcal{D}} \chi^2(d-1). \tag{3.20}$$

An application of this result, leads to reject the null hypothesis H_0, whenever the value of the statistic $\mathbb{T}_k$ exceeds $q_{1-\alpha}$, namely, the $(1-\alpha)$-quantile of the χ^2 law with $d-1$ degrees of freedom. The corresponding test is then, asymptotically of level α, when $n \to \infty$. The critical region is, accordingly, given by

$$CR := \{\mathbb{T}_N > q_{1-\alpha}\}.$$

The convergence in (3.20), can be naturally extended to Cramér-von Mises statistics based on the quadratic form of the modified processes.

4. Simulations studies

In this section, series of experiments were conducted in order to examine the performance of the proposed methodology. We provide numerical illustrations regarding the performance of the new estimators and the impact of the modified estimate on the asymptotic power function. The computing program codes were implemented in R.

4.1. *Estimation and Testing problem in Exponential AR(1) (EAR(1))*

We generate $r = 500$ samples of sizes $N = 60, 65, 70, 75, 80, 85, 90, 95, 100, 150$ and 200 through the model described as follows :

$$Y_t = \left[(\theta_0 + a_N) + (\theta_1 + b_N)\exp(-Y_{t-1}^2)\right] Y_{t-1} + \epsilon_t, \tag{4.1}$$

where

$$\theta_0 = 0.5,\ \theta_1 = 5,\ a_N = b_N = 1/N.$$

$\{\epsilon_t\}$ is an i.i.d. sequence of innovations with standard gaussian distribution and a positive density function f. For a given observations $Y_1,\ldots,Y_N$, we propose to test the hypothesis that $Y_i, i=1,\ldots,N$ follow the model

$$\begin{aligned} Y_t &= \left[\theta_0+\theta_1\exp(-Y_{t-1}^2)\right]Y_{t-1}+\epsilon_t \\ &=: H_\theta(Y_{t-1})+\epsilon_t. \end{aligned} \tag{4.2}$$

against the hypothesis that $Y_t, t=1,\ldots,N$ follow the model

$$Y_t=\left[(\theta_0+N^{-\frac{1}{2}})+(\theta_1+N^{-\frac{1}{2}})\exp(-Y_{t-1}^2)\right]Y_{t-1}+\epsilon_t. \tag{4.3}$$

This testing problem is equivalent to test $H_0:\theta=(\theta_0,\theta_1)$ versus

$$H_1^N:\theta=\left(\theta_0+N^{-\frac{1}{2}},\theta_1+N^{-\frac{1}{2}}\right).$$

Note that the model (4.2) is a particular case of $EAR(1)$ model. As mentioned in Section 3.3, its LAN property was established under assumption **(C1)**. Let us at first checking this last assumption. In fact, by a simple computation, we find

$$\left|\boldsymbol{\nabla}\partial H_\theta(\tilde{\theta})\right|\leq\max\left(|\tilde{\theta_0}|+c,(|\tilde{\theta_1}|+c)\exp(-Y_{t-1}^2)\right)|Y_{t-1}|\leq K_\theta(Y_{t-1}),$$

where c is a positive constant such that $|\tilde{\theta}-\theta|<c$,

$$C=\max(|\tilde{\theta_0}|+c,(|\tilde{\theta_1}|+c)\exp(-Y_{t-1}^2)),$$

and

$$K_\theta(Y_{t-1})=C|Y_{t-1}|.$$

The existence of the third moment of ϵ_t, implies that $\mathbb{E}_\theta Y_{t-1}^2<\infty$. Therefore, **(C1)** is satisfied. Making use Hwang and Basawa (1993)[Theorem 2.3], we deduce the LAN property of the model (4.2).

4.2. *Estimation*

At a first step, the parameter $\theta=(\theta_0,\theta_1)$ is estimated by the least squared estimator (LSE). From this preliminary $\sqrt{n}$-consistent estimator, we construct a modified estimate (ME) and we compare the mean-squared errors MSE1 and MSE2 of LSE and ME respectively. In the next step, we explain our method of the construction of ME. Decompose each sample of size N into two samples of sizes, the large one with length $s=\lfloor N^{(1-l)}\rfloor$ and the second one with length m, respectively. Here $l=0.009,\ \ 0.0095$ and 0.01. With the observations $Y_1,\ldots,Y_s$, we compute the LSE of the parameter (θ_0,θ_1) of model (4.2). With the m remainder observations $Y_{1+s},\ldots,Y_N$ we compute the LSE of model (4.2). From the equality (2.16), we obtain the modified estimate (ME). In the sequel, we denote by $\widehat{\boldsymbol{\theta}}_s^{(j)}$, and $\widehat{\boldsymbol{\theta}}_m^{(j)}$, for $j=1,\ldots,r$ the LSE estimates obtained from the r samples based on observations $Y_1,\ldots,Y_s$ and $Y_{1+s},\ldots,Y_N$, respectively. Therefore the central sequence S_N and the Fisher information I were computed in Hwang and Basawa (1993)[Example 4.2] and described as

$$
\begin{aligned}
S_N(\theta_0) &= -N^{-\frac{1}{2}} \begin{pmatrix} \sum_{i=1}^{N} Y_{i-1} \frac{\dot{f}(\epsilon_i)}{f(\epsilon_i)} \\ \sum_{i=1}^{N} Y_{i-1} \exp(-Y_{i-1}^2) \frac{\dot{f}(\epsilon_i)}{f(\epsilon_i)} \end{pmatrix} \\
&=: \mathbf{V}_N(\theta_0), \\
I(\theta_0) &= \mathbb{E}_\theta \begin{bmatrix} Y_{i-1}^2 & Y_{i-1}^2 \exp(-Y_{i-1}^2) \\ Y_{i-1}^2 \exp(-Y_{i-1}^2) & Y_{i-1}^2 \exp(-2Y_{i-1}^2) \end{bmatrix} \cdot i_E
\end{aligned}
$$

where i_E denotes

$$
\mathbb{E}_\theta \left[\frac{\dot{f}(\epsilon_i)}{f(\epsilon_i)} \right]^2 .
$$

The matrix $\mathbf{A}_n$ defined in (2.2) will be expressed in this setting by

$$
\mathbf{A}_N(\widehat{\boldsymbol{\theta}}_s^{(j)}) = \begin{bmatrix} \frac{-1}{\sqrt{N}} \sum_{i=1}^{N} Y_{i-1,j}^2 & \frac{-1}{\sqrt{N}} \sum_{i=1}^{N} Y_{i-1,j}^3 \exp(-Y_{i-1}^2) \\ \frac{-1}{\sqrt{N}} \sum_{i=1}^{N} Y_{i-1}^2 \exp(-Y_{i-1}^2) & \frac{-1}{\sqrt{N}} \sum_{i=1}^{N} Y_{i-1}^3 \exp(-2Y_{i-1}^2) \end{bmatrix},
$$

Here $Y_{1,j}, \ldots, Y_{N,j}$ is the j-th generated sample of size N. Simple computations shows that the matrix determinants of $\mathbf{A}_N(\widehat{\boldsymbol{\theta}}_s^{(j)})$ are not equal to zero, this readily implies that $\mathbf{A}_{N,j}$ is invertible. An application of Theorem 2.6 implies the existence of modified estimates $\bar{\boldsymbol{\theta}}_N^{(j)}$ of the parameter $\boldsymbol{\theta}_0$, such that

$$
\bar{\boldsymbol{\theta}}_N^{(j)} = \widehat{\boldsymbol{\theta}}_s^{(j)} + \mathbf{A}_N{}^{-1}(\widehat{\boldsymbol{\theta}}_s^{(j)}) \mathcal{D}_{s,m}^{(j)},
$$

where $\mathcal{D}_{s,m}^{(j)}$ is defined in (2.17). For each value of N, we do a comparison between the mean-squared error of LSE (MSE1) and the mean-squared error of ME (MSE2). Notice that MSE1 and MSE2 are approximated respectively by their empirical means

$$
\frac{1}{r} \sum_{i=1}^{r} \left\| \widehat{\boldsymbol{\theta}}_{N,0}^{(j)} - \boldsymbol{\theta}_0 \right\|^2 \quad \text{and} \quad \frac{1}{r} \sum_{i=1}^{r} \left\| \bar{\boldsymbol{\theta}}_N^{(j)} - \boldsymbol{\theta}_0 \right\|^2 .
$$

These results are summarized in the following tables.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Table 1. MSE comparison, $l = 0.009$

N	$s = \lfloor N^{1-l} \rfloor$	m	MSE1	MSE2
60	57	3	0.0001328802	0.619587
65	62	3	0.01303067	119.7262
70	67	3	0.002378401	1.031119
75	72	3	0.0002070228	0.3235394
80	76	4	0.001508324	0.05253198
85	81	4	0.002550674	0.1111781
90	86	4	0.0001203558	0.07579942
95	91	4	2.970162×10^{-5}	0.04505101
100	95	5	0.0005801212	0.08251387
150	143	7	7.87997×10^{-5}	0.06945373
200	190	10	2.153016×10^{-5}	0.06072364

Table 2. MSE comparison, $l = 0.0095$

N	$s = \lfloor N^{1-l} \rfloor$	m	MSE1	MSE2
60	57	3	0.003865599	25.81255
65	62	3	0.0001293232	1.762621
70	67	3	0.001843774	0.03215461
75	72	3	4.467617×10^{-5}	0.0344221
80	77	3	0.003200346	0.01045798
85	81	4	0.0004392172	0.0899404
90	86	4	0.0001552694	0.09981933
95	91	4	6.01109×10^{-5}	0.05338201
100	96	4	0.0015192	0.101397
150	143	7	0.0002538218	0.05161148
200	191	9	0.00102925	0.05753669

Table 3. MSE comparison, $l = 0.01$

N	$s = \lfloor N^{1-l} \rfloor$	m	MSE1	MSE2
60	57	3	0.000569337	12.70678
65	62	3	0.0005972831	1.990746
70	67	3	0.001097378	7.73527
75	71	4	0.000130873	0.1309442
80	76	4	0.0001155424	0.09538544
85	81	4	0.0001469284	0.08391853
90	86	4	0.001685851	0.09553644
95	90	5	7.395901×10^{-5}	0.1018716
100	95	5	0.00153452	0.1309031
150	95	5	8.108649×10^{-5}	0.07078288
200	189	11	0.0005927485	0.05685946

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

We remark that LSE behaves better than the ME when the sample size is moderate. Notice that when the sample size is large, the two estimates are very similar.

4.3. *Testing problem*

To test $H_0 : \theta = \theta_0$, we use the statistic test T_N defined in (4.4) by

$$T_N(\theta_0) \;\; = \;\; -\log\left(L_N(\theta_0)/L_N(\widehat{\theta}_{ML})\right).$$

Remark that in practice θ_0 is unknown. From LAN property (3.3), $T_N(\theta_0)$ is expressed in term of the central sequence $S_N(\theta_0)$ and the Fisher information $I(\theta_0)$. By applying the Theorem 2.6, there exists an estimate $\bar{\boldsymbol{\theta}}_N$ such that

$$S_N(\bar{\boldsymbol{\theta}}_N) = S_N(\theta_0) + o_{\mathbb{P}}(1).$$

The continuity of $\cdot \to I(\cdot)$ implies that $I(\theta_0)$ and $I(\bar{\boldsymbol{\theta}}_N)$ are very close. Consequently, under H_0, the two sequences of tests $T_N(\theta_0)$ and $T_N(\bar{\boldsymbol{\theta}}_N)$ are equivalent. By the contiguity, we also deduce this equivalence under H_1^N. This enables to replace θ_0 by its estimate ME $\bar{\boldsymbol{\theta}}_N$. To evaluate the performance of our methodology, we use the same $r = 500$ samples and we do a comparison between the empirical power functions obtained from the statistic tests (4.4), with the true value θ_0, the LSE estimate and the ME estimate. By fixing the level u of tests at 5%, we compute for each value of N the empirical power functions of to the statistical tests obtained with the true parameter (Power0), with the LSE (Power1) and with the ME (Power2). Let us recall that the limit of the empirical power is given by

$$\lim_{N\to\infty} \mathbb{E}(\mathbb{1}_{\{T_N \geq q_{1-\alpha}\}}) = \lim_{N\to\infty} \mathbb{P}(T_N \geq q_{1-\alpha}).$$

Numerical approximation routines can be used to compute the power. This can approximated from r samples by computing the r statistics based on each sample, that is

$$\frac{1}{r}\sum_{j=1}^{r} \mathbb{1}\{T_N^j \geq q_{1-\alpha}\},$$

where $(1-\alpha)$-quantile of the χ^2 law with 2 degrees of freedom. This will be done for the statistics under comparison, $T_N(\theta_0)$, $T_n(\widehat{\boldsymbol{\theta}}_N)$ and $T_N(\bar{\boldsymbol{\theta}}_N)$ corresponding to true parameter, LSE and ME respectively. For sample sizes $N = 60, 65, 70, 75, 80, 85, 90, 100, 150$ and 200 we then compare in each case the absolute deviation dev1 between Power0 and Power1 with the absolute deviation dev2 between Power0 and Power1. All these results are summarized in the following tables.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Table 4. Power comparison with Likelihood Ratio Test Statistic, l=0.009

N	Power0	Power1	Power2	Dev1	Dev2
60	0.718	0.678	0.66	0.04	0.058
65	0.684	0.632	0.632	0.052	0.052
70	0.726	0.708	0.706	0.018	0.02
75	0.706	0.686	0.67	0.02	0.036
80	0.74	0.73	0.732	0.01	0.008
85	0.716	0.706	0.718	0.01	0.002
90	0.738	0.706	0.724	0.032	0.014
95	0.772	0.726	0.76	0.046	0.012
100	0.754	0.728	0.732	0.026	0.022
150	0.856	0.818	0.834	0.038	0.022
200	0.846	0.814	0.826	0.032	0.02

Table 5. Power comparison with Likelihood Ratio Test Statistic, l=0.0095

N	Power0	Power1	Power2	Dev1	Dev2
60	0.678	0.664	0.66	0.014	0.018
65	0.694	0.654	0.654	0.04	0.04
70	0.706	0.662	0.678	0.044	0.028
75	0.728	0.696	0.702	0.032	0.026
80	0.72	0.714	0.692	0.006	0.028
85	0.785	0.702	0.702	0.01	0.01
90	0.758	0.74	0.756	0.0188	0.002
95	0.766	0.732	0.74	0.034	0.026
100	0.788	0.772	0.786	0.016	0.002
150	0.836	0.8	0.82	0.036	0.016
200	0.848	0.824	0.84	0.024	0.008

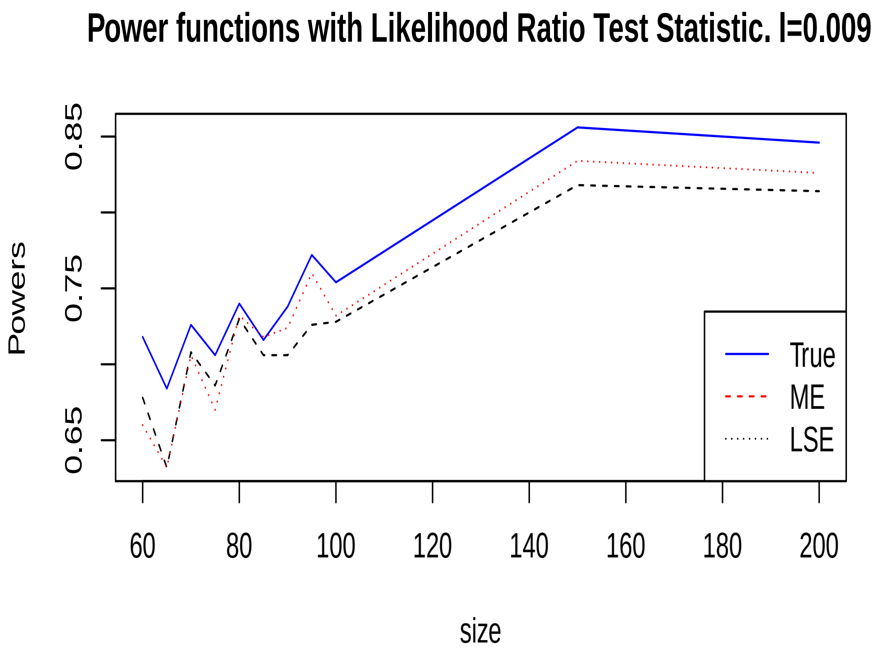
Table 6. Power comparison with Likelihood Ratio Test Statistic, l=0.01

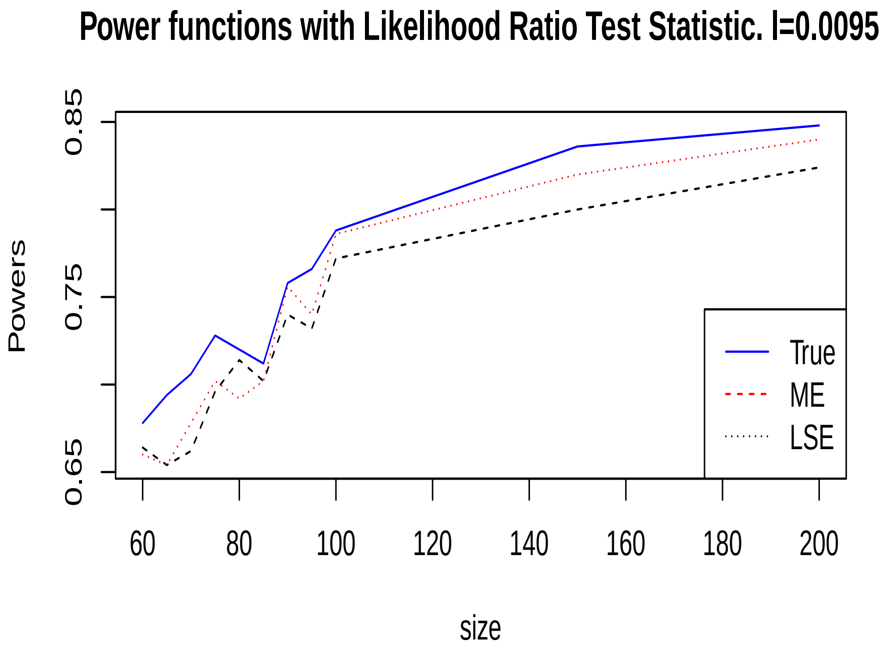
N	Power0	Power1	Power2	Dev1	Dev2
60	0.684	0.67	0.632	0.014	0.052
65	0.674	0.65	0.662	0.024	0.012
70	0.704	0.674	0.656	0.03	0.048
75	0.722	0.708	0.728	0.014	0.006
80	0.746	0.716	0.734	0.03	0.012
85	0.786	0.752	0.762	0.034	0.024
90	0.762	0.732	0.756	0.03	0.006
95	0.744	0.736	0.742	0.008	0.002
100	0.756	0.742	0.766	0.014	0.01
150	0.828	0.784	0.816	0.044	0.012
200	0.844	0.816	0.822	0.028	0.022

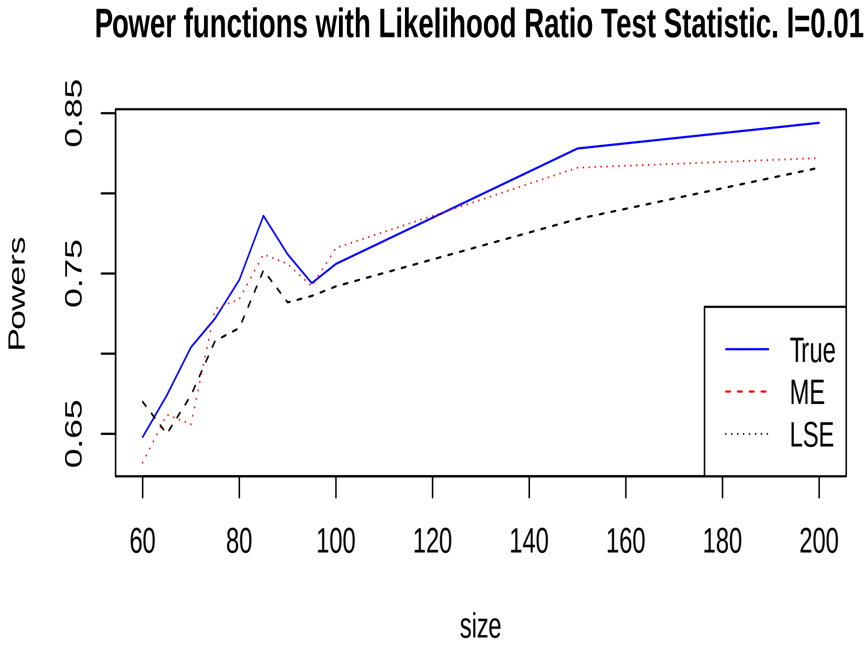
© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

It is shown by Tables $4-6$, when samples are large, the power of the statistics based on the modified is better the one based on LSE estimate. This corresponds to the asymptotic framework in which our main results are proven. From Tables $1-3$, one can that the quadratic errors are smaller for the classical estimator. In summary, ME estimate is usually further from the true value than the classical estimator. We can see also for large samples the proposed statistics is very close to the statistics when the parameter is known which is corroborated in the following Figures comparing graphically the power of the 5%-level the statistical tests. As the picture highlights, the power of the proposed test based on the modified estimate is generally the optimal one. In order to extract methodological recommendations for the use of the proposed statistics in this work, it will be interesting to conduct extensive Monte Carlo experiments to compare our procedures with other alternatives presented in the literature, but this would go well beyond the scope of the present paper.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr


Power functions with Likelihood Ratio Test Statistic. l=0.009
Powers
0.65
0.75
0.85
60
80
100
120
140
160
180
200
size
True
ME
LSE


Power functions with Likelihood Ratio Test Statistic. l=0.0095
Powers
0.65
0.75
0.85
60
80
100
120
140
160
180
200
size
True
ME
LSE



5. Concluding Remarks

In this present work the modified estimator defined from $\mathbb{R}^d$ to $\mathbb{R}$ in Lounis (2011) was extended from $\mathbb{R}^d$ to from $\mathbb{R}^d$. In addition the consistency is deeply studied and developed. The method of evaluating error through confidence intervals introduced in Lounis (2017) induces a random confidence level. Proposition 2.1 enables to evaluate the error without any restriction. In Lounis (2017) equivalence between a random sequence and its estimated version concerns rather a sub-sample. Extension of these results to the total number of observations is obtained in the present work. Criterion (2.18) gives a precise answer of a question which was be left open in Lounis (2017). We finally extend our finding to the marked empirical processes. It would be of interest to provide a complete investigation of this problem, that goes well beyond the scope of the present paper, we leave this problem open for future. Construction of ME in multidimensional case is essentially based on assumption that Jacobian of $\mathbf{V}_n$ is asymptotically not equal to zero. In future it would be of interest to relax this condition.

6. Proof of the results

This section is devoted to the proofs of our results. The previously presented notation continues to be used in the following.

Proof of Proposition 2.1

Consider the following decomposition

$$\sqrt{n}(\lambda - \lambda_n) = \sqrt{n}(\lambda - \lambda_N) + \sqrt{n}(\lambda_N - \lambda_n).$$

Remark that

$$\sqrt{n}(\lambda - \lambda_N) = \sqrt{\frac{n}{N}} \times \sqrt{N}(\lambda - \lambda_N). \tag{6.1}$$

Since $n = o(N)$, it follows that

$$\begin{aligned}\sqrt{n}(\lambda - \lambda_N) &= o_{\mathbb{P}}(1) \times O_{\mathbb{P}}(1), \\ &= o_{\mathbb{P}}(1). \end{aligned} \tag{6.2}$$

Hence, the proof is complete. □

Proof of Proposition 2.2

It is sufficient to show that for each component $\mathbf{V}_{n,j}$. By Taylor expansion of $\mathbf{V}_{n,j}$ of order 1 in the neighborhood of $\widehat{\boldsymbol{\theta}}_n$, it follows that

$$\begin{aligned}\mathbf{V}_{n,j}(\boldsymbol{\theta}) &= \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) \\ &+ \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n), \end{aligned} \tag{6.3}$$

where $\tilde{\boldsymbol{\theta}}_n^{(j)}$ between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n$, and $H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})$ is the Hessian matrix evaluated at $\tilde{\boldsymbol{\theta}}_n^{(j)}$. Similarly, (6.3) will be rewritten as follows

$$\begin{aligned}\mathbf{V}_{n,j}(\boldsymbol{\theta}) &= \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}) \\ &+ \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n). \end{aligned} \tag{6.4}$$

Observe that

$$\frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}) = \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\sqrt{n}\ \partial \boldsymbol{\theta}_i} \times \sqrt{n}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}).$$

By making use of Proposition 2.1, it follows that

$$\begin{aligned}\frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}) &= \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\sqrt{n}\ \partial \boldsymbol{\theta}_i} \times \sqrt{n}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}) \\ &+ \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\sqrt{n}\partial \boldsymbol{\theta}_i} \times o_{\mathbb{P}}(1). \end{aligned}$$

From the assumption (**A.3**), it results that

$$\frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}) = \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}) + o_{\mathbb{P}}(1),$$

which, in turn, implies that

$$\sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}) \;=\; \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}) + o_{\mathbb{P}}(1).$$

Combining this latter equality with (6.4) gives

$$\begin{aligned} \mathbf{V}_{n,j}(\boldsymbol{\theta}) \;&=\; \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}) \\ &\quad + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) + o_{\mathbb{P}}(1). \end{aligned} \tag{6.5}$$

It remains to evaluate the quantity

$$\frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n).$$

Notice that we have

$$H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)}) \;=\; \begin{bmatrix} \dfrac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_1^2} & \cdots & \dfrac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_d} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_d \boldsymbol{\theta}_1} & \cdots & \dfrac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_d^2} \end{bmatrix},$$

and

$$(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top \;=\; \left(\sum_{i=1}^{d}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i})\frac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_1}, \ldots, \sum_{i=1}^{d}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i})\frac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_d}\right).$$

Let us introduce, for each $j = 1, \ldots, d$,

$$h_{n,i,j} = \frac{1}{\sqrt{n}}\; \sqrt{n}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i})\frac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\sqrt{n}\;\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_k}\sqrt{n}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{n,k}).$$

By simple computation, we conclude easily that

$$\frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) \;=\; \sum_{1 \le i,k \le d}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i})\frac{\partial^2 \mathbf{V}_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_k}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{n,k}).$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

$$= \sum_{1\leq i,k\leq d} h_{n,i,j}.$$

Making use of the condition (**A.3**) and by the fact that

$$\sqrt{n}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}) = O_{\mathbb{P}}(1), \quad \sqrt{n}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{n,k}) = O_{\mathbb{P}}(1),$$

it follows, in turn, that

$$h_{n,i,j} \quad = \quad o_{\mathbb{P}}(1),$$

which implies that

$$\frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top H_{n,j}(\tilde{\boldsymbol{\theta}}_n^{(j)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) = o_{\mathbb{P}}(1). \tag{6.6}$$

This when combined with (6.5), readily implies that

$$\mathbf{V}_{n,j}(\boldsymbol{\theta}) = \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_{n,j} + o_{\mathbb{P}}(1),$$

where

$$\mathcal{D}_{n,j} \quad = \quad \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}).$$

Hence the proof is complete. □

Proof of Theorem 2.3

From the Proposition 2.2, we have for all $j = 1\ldots, d$.

$$\mathbf{V}_{n,j}(\boldsymbol{\theta}) = \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_{n,j} + o_{\mathbb{P}}(1), \tag{6.7}$$

where

$$\mathcal{D}_{n,j} \quad = \quad \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}). \tag{6.8}$$

For all $j = 1\ldots, d$, consider the tangent space Γ_j of the map of $\mathbf{V}_{n,j}$ at $\widehat{\boldsymbol{\theta}}_n$, defined by

$$\Gamma_j := \left\{(x, \mathbf{V}_{n,j}(x)), \quad \mathbf{V}_{n,j}(x) - \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) = \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)^\top \cdot (x - \widehat{\boldsymbol{\theta}}_n)\right\},$$

here, the script " $\cdot$ " denotes the inner product and recall

$$\boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)^\top = \left(\frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_1}, \ldots, \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_d}\right).$$

The modified estimator $\overline{\boldsymbol{\theta}}_n$ is obtained by disrupting each component of the first estimator $\widehat{\boldsymbol{\theta}}_n$. To be more precise, write

$$\overline{\boldsymbol{\theta}}_n^{\top} = (\widehat{\boldsymbol{\theta}}_{n,1} + p_1, \ldots, \widehat{\boldsymbol{\theta}}_{n,d} + p_d), \tag{6.9}$$

where $p_1, \ldots, p_d$ are unknown real values to determinate and $\overline{\boldsymbol{\theta}}_n$ is constructed on the tangent space Γ_j, this implies that

$$\mathbf{V}_{n,j}\left(\overline{\boldsymbol{\theta}}_n\right) = \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)^\top \cdot (\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n) + \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right). \tag{6.10}$$

Let us define $\mathcal{D}_{n,j}$ as follows

$$\mathcal{D}_{n,j} = \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)^\top \cdot (\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n). \tag{6.11}$$

Notice that, for each $j = 1, \ldots, d$, we have

$$\mathbf{V}_{n,j}(\boldsymbol{\theta}) \;=\; \mathbf{V}_{n,j}\left(\overline{\boldsymbol{\theta}}_n\right) + o_{\mathbb{P}}(1),$$

and hence

$$\mathbf{V}_{n}(\boldsymbol{\theta}) \;=\; \mathbf{V}_{n}\left(\overline{\boldsymbol{\theta}}_n\right) + o_{\mathbb{P}}(1). \tag{6.12}$$

By combining the preceding equations, we readily infer that

$$\sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i} \times p_i = \mathcal{D}_{n,j}, \tag{6.13}$$

where

$$\mathcal{D}_{n,j} = \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i} (\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}). \tag{6.14}$$

We have, for $k = 1, \ldots, d$

$$\sum_{i=1}^{d} \frac{\partial \mathbf{V}_{n,k}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i} \times p_i \;=\; \mathcal{D}_{n,k}. \tag{6.15}$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Remark that (6.15) is equivalent to the matrix system $\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)X = \mathcal{D}_n$, where

$$\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n) = \begin{bmatrix} \frac{\partial \mathbf{V}_{n,1}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_1} & \cdots & \frac{\partial \mathbf{V}_{n,1}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{V}_{n,d}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_1} & \cdots & \frac{\partial \mathbf{V}_{n,d}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_d} \end{bmatrix}, \quad X = \begin{bmatrix} p_1 \\ \vdots \\ p_d \end{bmatrix}, \quad \text{and} \quad \mathcal{D}_n = \begin{bmatrix} \mathcal{D}_{n,1} \\ \vdots \\ \mathcal{D}_{n,d} \end{bmatrix}.$$

By condition (**A.2**), this is a Cramér system that admits a unique solution

$$X = A_n^{-1}(\widehat{\boldsymbol{\theta}}_n) + \mathcal{D}_n.$$

This in connection with equality (6.9), implies that

$$\overline{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n + A_n^{-1}(\widehat{\boldsymbol{\theta}}_n)\mathcal{D}_n. \tag{6.16}$$

Therefore the proof is complete. □

Proof of the Theorem 2.4

Consider the following decomposition

$$\sqrt{n}(\overline{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \;=\; -\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n) + \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}). \tag{6.17}$$

As defined in (2.12), we observe that all components of the random vector $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n)$ are null, except the j-th one which is given by

$$\sqrt{n}(\boldsymbol{\theta}_{n,j} - \overline{\boldsymbol{\theta}}_{n,j}) \;=\; \sqrt{n}\vartheta_n = \sqrt{n}\frac{\mathcal{D}_n}{\max\limits_{1\leq i\leq d}\left\{\frac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}\right\}}.$$

Applying the triangle inequality, we obtain

$$\begin{aligned}\sqrt{n}\left|\boldsymbol{\theta}_{n,j} - \overline{\boldsymbol{\theta}}_{n,j})\right| \;&=\; \sqrt{n}\frac{|\mathcal{D}_n|}{\left|\max_{1\leq i\leq d}\left\{\frac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}\right\}\right|} \\ &\leq\; \sum_{i=1}^{d}\sqrt{n}\left|\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}\right|.\end{aligned}$$

Making use Proposition 2.1, we find

$$\sqrt{n}\left|\boldsymbol{\theta}_{n,j} - \overline{\boldsymbol{\theta}}_{n,j}\right| \;\leq\; \sum_{i=1}^{d}\sqrt{n}\left|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}\right| + o_{\mathbb{P}}(1). \tag{6.18}$$

Finally, inequality (6.18) implies that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n) = O_{\mathbb{P}}(1).$$

Combining this last equality with (6.17) gives

$$\sqrt{n}(\boldsymbol{\theta}_{n,j} - \overline{\boldsymbol{\theta}}_{n,j}) \;=\; O_{\mathbb{P}}(1). \tag{6.19}$$

Consider again the equality (2.12) and by using again Proposition 2.1, we find

$$\begin{aligned}\left|\boldsymbol{\theta}_{n,j} - \overline{\boldsymbol{\theta}}_{n,j}\right| &= \frac{|\mathcal{D}_n|}{\left|\max_{1\leq i\leq d}\left\{\frac{\partial \mathbf{V}_n\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_i}\right\}\right|},\\ &\leq \sum_{i=1}^{d}\left|\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{n,i}\right| = \sum_{i=1}^{d}\left|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_{n,i}\right| + o_{\mathbb{P}}(1).\end{aligned}$$

From the last equality, we deduce that

$$(\widehat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n) = o_{\mathbb{P}}(1).$$

Thus the proof is complete. □

Proof of the Theorem 2.5

From (2.9), we readily obtain

$$\begin{aligned}\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n &= \mathbf{A}_n^{-1}(\widehat{\boldsymbol{\theta}}_n)\mathcal{D}_n\\ &= \frac{1}{\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))}\mathrm{com}(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))\mathcal{D}_n,\end{aligned} \tag{6.20}$$

with

$$\mathrm{com}(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)) \;=\; \begin{bmatrix} A_{1,1}(n) & \cdots & A_{1,d}(n)\\ \cdots & \ddots & \vdots\\ A_{d,1}(n) & \cdots & A_{d,d}(n)\end{bmatrix},$$

and, for $(i,j) \in \{1,\ldots,d\} \times \{1,\ldots,d\}$,

$$\begin{aligned}A_{i,j}(n) &= (-1)^{i+j}\det(M_{i,j}(n)),\\ &= n^{\frac{d-1}{2}}(-1)^{i+j}\sum_{\sigma\in S_{n-1}}\left[\epsilon(\sigma)\prod_{k\in\{1,\ldots,d\},\;\; k\neq i,\;\; \sigma(k)\neq j}\frac{1}{\sqrt{n}}\frac{\partial \mathbf{V}_{n,k}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_{\sigma(k)}}\right],\end{aligned}$$

here $M_{i,j}(n)$ is the submatrix extracted from $\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)$ by throwing away row i and column j.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

From (6.20) and by taking the norm, we write

$$\sqrt{n}\left\|\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right\| \;=\; \frac{\sqrt{n}}{\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))\right|}\left\|\mathrm{com}(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))\mathcal{D}_n\right\|, \tag{6.21}$$

Similarly, the matrix $\mathrm{com}(\mathbf{A}_n)$ will be rewritten as

$$\mathrm{com}(\mathbf{A}_n) \;=\; n^{\frac{d-1}{2}} L_n, \tag{6.22}$$

where, the $d \times d$-matrix L_n is defined as follows, for $(i,j) \in \{1,\ldots,d\} \times \{1,\ldots,d\}$,

$$L_n(i,j) \;=\; (-1)^{i+j} \sum_{\sigma \in S_{n-1}} \left[\epsilon(\sigma) \prod_{k\in\{1,\ldots,d\}, k\neq i, \sigma(k)\neq j} \frac{1}{\sqrt{n}} \frac{\partial \mathbf{V}_{n,k}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}_{\sigma(k)}}\right]$$

Combining (6.21) with (6.23) gives

$$\begin{aligned}\sqrt{n}\left\|\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right\| \;&=\; \frac{n^{\frac{d}{2}}}{\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))\right|}\left\|L_n\mathcal{D}_n\right\|, \\ &=: \; \frac{n^{\frac{d}{2}}}{\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))\right|}\left\|Q_n\right\|. \end{aligned} \tag{6.23}$$

From (**A.3**), each element $L_n(i,j)$ of the matrix L_n is bounded in probability. Because $\mathcal{D}_n$ is also bounded in probability and by doing the product between the matrix L_n and the $(n \times 1)$-vector $\mathcal{D}_n$, we deduce that the $(n \times 1)$-vector Q_n is bounded in probability in $\mathbb{R}^d$. By the fact that

$$\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)\right| \geq n^{\frac{d}{2}},$$

we conclude that

$$\sqrt{n}\left(\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right) = O_{\mathbb{P}}(1). \tag{6.24}$$

We have from (6.23)

$$\left\|\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n\right\| \;=\; \frac{1}{\sqrt{n}} \times \frac{n^{\frac{d}{2}}}{\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n))\right|}\left\|Q_n\right\|.$$

As already mentioned Q_n is bounded in probability and

$$\left|\det(\mathbf{A}_n(\widehat{\boldsymbol{\theta}}_n)\right| \geq n^{\frac{d}{2}}.$$

Consequently $\overline{\boldsymbol{\theta}}_n - \widehat{\boldsymbol{\theta}}_n$ converges in probability to 0 in $\mathbb{R}^d$ as n tends to $+\infty$. By considering the following decomposition and using triangle inequality, we find

$$\left\|\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}_n\right\| = \left\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n\right\| + \left\|\widehat{\boldsymbol{\theta}}_n - \overline{\boldsymbol{\theta}}_n\right\|,$$

and

$$\sqrt{n}\left\|\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}_n\right\| = \sqrt{n}\left\|\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n\right\| + \sqrt{n}\left\|\widehat{\boldsymbol{\theta}}_n-\overline{\boldsymbol{\theta}}_n\right\|.$$

By the fact that $\widehat{\boldsymbol{\theta}}_n$ is a $\sqrt{n}$-consistent combined with the last previous results, thus, the proof of Theorem is complete.

□

6.1. *Proof of the Theorem 2.6*

1. Return back to Proposition (2.2) and recall that equivalence between sequences was obtained in the case when $n = o(N)$. By considering the decomposition (2.13) and for η sufficiently small, we obtain

$$m < \lfloor N^{1-\eta} \rfloor = s. \tag{6.25}$$

We have

$$\frac{\lfloor N^{1-\eta} \rfloor}{N} \leq \frac{N^{1-\eta}}{N} \longrightarrow 0 \quad \text{as} \quad N \longrightarrow \infty.$$

This when combined with the inequality (6.25) gives,

$$s = o\left(N\right) \quad \text{and} \quad m = o\left(N\right). \tag{6.26}$$

Obvsiouly, (6.26) enables to apply Proposition (2.2). By considering the decomposition (2.5) one can write

$$\begin{aligned}
\mathbf{V}_{N,j}\left(\boldsymbol{\theta}\right) &= \frac{-1}{\sqrt{N}}\sum_{i=1}^{N}\Upsilon_j\left(\boldsymbol{\theta},\mathbf{Y}_i\right) \\
&= \sqrt{\frac{s}{N}}\frac{-1}{\sqrt{s}}\sum_{i=1}^{s}\Upsilon_j\left(\boldsymbol{\theta},\mathbf{Y}_i\right) + \sqrt{\frac{m}{N}}\frac{-1}{\sqrt{m}}\sum_{i=1+s}^{N}\Upsilon_j\left(\boldsymbol{\theta},\mathbf{Y}_i\right) \\
&=: \sqrt{\frac{s}{N}}\mathbf{V}_{s,j}\left(\boldsymbol{\theta}\right) + \sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\boldsymbol{\theta}\right).
\end{aligned}$$

Making use of the Proposition (2.2), we obtain

$$\begin{aligned}
\mathbf{V}_{N,j}\left(\boldsymbol{\theta}\right) &= \sqrt{\frac{s}{N}}\mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right) \\
&\quad + \sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{s,i}\right) \\
&\quad + \sqrt{\frac{m}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{m,i}\right) + \left(\sqrt{\frac{s}{N}}+\sqrt{\frac{m}{N}}\right)o_{\mathbb{P}}(1) \\
&= \sqrt{\frac{s}{N}}\mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)
\end{aligned}$$

$$
\begin{aligned}
&+\sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{s,i}\right)\\
&+\sqrt{\frac{m}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{m,i}\right)+o_{\mathbb{P}}(1)\\
= \ & \sqrt{\frac{s}{N}}\mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right) \ +\sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)\\
&-\sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)+\sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)\\
&+\sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{s,i}\right)\\
&+\sqrt{\frac{m}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{m,i}\right)+o_{\mathbb{P}}(1)\\
= \ & \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s\right)+\mathcal{D}_{s,m,j}+o_{\mathbb{P}}(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{D}_{s,m,j} \ = \ & -\sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)+\sqrt{\frac{m}{N}}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)\\
&+\sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{s,i}\right)\\
&+\sqrt{\frac{m}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{m,i}\right)\\
=: \ & \sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)-\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)\right]+\ \sqrt{\frac{s}{N}}\mathcal{D}_{s,j}+\sqrt{\frac{m}{N}}\mathcal{D}_{m,j}.
\end{aligned}
$$

It remains to check that for $j\in\{1,\ldots,d\}$, $\mathcal{D}_{s,m,j}$ is bounded in probability in $\mathbb{R}$. We can write

$$
\sqrt{\frac{m}{N}}\mathcal{D}_{m,j}=\sqrt{\frac{m}{N}}\times\sum_{i=1}^{d}\frac{1}{\sqrt{m}}\frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\sqrt{m}\left(\widehat{\boldsymbol{\theta}}_{N,i}-\widehat{\boldsymbol{\theta}}_{m,i}\right).
$$

Since $m<N$ and by condition $(\mathbf{A.3})$, we deduce that

$$
\sqrt{\frac{m}{N}}\mathcal{D}_{m,j}=O_{\mathbb{P}}(1). \tag{6.27}
$$

With a similar reasoning we show that

$$
\sqrt{\frac{s}{N}}\mathcal{D}_{s,j}=O_{\mathbb{P}}(1). \tag{6.28}
$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Since $s = o(N)$ and we make use of Proposition (2.2), we infer that

$$\mathbf{V}_{m,j}(\boldsymbol{\theta}) = \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right) + \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{m,i}\right) + o_{\mathbb{P}}(1), \tag{6.29}$$

$$\mathbf{V}_{m,j}(\boldsymbol{\theta}) = \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \sum_{i=1}^{d} \frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{s,i}\right) + o_{\mathbb{P}}(1). \tag{6.30}$$

Doing the difference between (6.29) and (6.30) gives

$$\begin{aligned}\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right) &= \sum_{i=1}^{d} \frac{1}{\sqrt{m}} \frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i} \sqrt{m}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{s,i}\right) \\ &\quad - \sum_{i=1}^{d} \frac{1}{\sqrt{m}} \frac{\partial \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)}{\partial \boldsymbol{\theta}_i} \sqrt{m}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{m,i}\right) + o_{\mathbb{P}}(1)\end{aligned} \tag{6.31}$$

Using again the assumption (**A3**) permits to deduce that

$$\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)$$

is bounded in probability and then by multiplying by $\sqrt{\frac{m}{N}}$ we find

$$\sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right)\right] = O_{\mathbb{P}}(1). \tag{6.32}$$

From (6.27), (6.28) and (6.32), we obtain

$$\mathcal{D}_{s,m,j} = O_{\mathbb{P}}(1).$$

Finally, we deduce that we have

$$\mathbf{V}_{N,j}(\boldsymbol{\theta}) = \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \mathcal{D}_{s,m,j} + o_{\mathbb{P}}(1). \tag{6.33}$$

2. From (6.33), we have

$$\begin{aligned}\mathbf{V}_{N,1}(\boldsymbol{\theta}) - \mathbf{V}_{N,1}\left(\widehat{\boldsymbol{\theta}}_s\right) &= \mathcal{D}_{s,m,1} + o_{\mathbb{P}}(1) \\ &\vdots \\ \mathbf{V}_{N,d}(\boldsymbol{\theta}) - \mathbf{V}_{N,d}\left(\widehat{\boldsymbol{\theta}}_s\right) &= \mathcal{D}_{s,m,d} + o_{\mathbb{P}}(1).\end{aligned}$$

Following a same reasoning and steps as in the proof of Theorem 2.3, we show the existence of a modified estimate such that

$$\overline{\boldsymbol{\theta}}_N = \widehat{\boldsymbol{\theta}}_s + \mathbf{A}_N^{-1}\left(\widehat{\boldsymbol{\theta}}_s\right)\mathcal{D}_{s,m},$$

where

$$\mathcal{D}_{s,m} = \begin{bmatrix} \mathcal{D}_{s,m,1} \\ \vdots \\ \mathcal{D}_{s,m,d} \end{bmatrix}.$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

3. By replacing $\mathbf{A}_n\left(\widehat{\boldsymbol{\theta}}_n\right)$ and $\mathcal{D}_n$, respectively by $\mathbf{A}_N\left(\widehat{\boldsymbol{\theta}}_s\right)$ and $\mathcal{D}_{s,m}$ and doing a similar reasoning as in the proof of Theorem 2.5, we show the desired result of consistency.

Hence the proof is complete. □

6.2. Proof of the Theorem 2.9

The proof is organized in three steps. The first one consists to establish an explicit link between $\mathbf{V}_{n,j}(\boldsymbol{\theta})$ and $\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)$. The second step gives an extension from sub-samples size n to the sample of size N. The last gives an equivalence between $\mathbf{V}_{N,j}(\boldsymbol{\theta})$ with it's estimated version. In a sequel $n = o(N)$, p an integer such $p \geq 2$ and $j = 1, \ldots, d$.

1. By Taylor expansion of $\mathbf{V}_{n,j}$ of order $p-1$ in the neighborhood of $\widehat{\boldsymbol{\theta}}_n$, it follows that

$$\begin{aligned}
\mathbf{V}_{n,j}(\boldsymbol{\theta}) &= \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)\cdot(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n) \\
&\quad + \cdots + \frac{1}{(p-1)!}\left(D^{p-1}\mathbf{V}_{n,j}\right)\left(\widehat{\boldsymbol{\theta}}_n\right)\cdot(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n)^{p-1} \\
&\quad + \frac{1}{(p)!}\left(D^{p}\mathbf{V}_{n,j}\right)\left(\tilde{\boldsymbol{\theta}}_n^{(j)}\right)\cdot(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}_n^{(j)})^{p} \\
&=: \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_{n,j} + \frac{1}{(p)!}\left(D^{p}\mathbf{V}_{n,j}\right)\left(\tilde{\boldsymbol{\theta}}_n^{(j)}\right)\cdot(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}_n^{(j)})^{p},
\end{aligned} \tag{6.34}$$

where $\tilde{\boldsymbol{\theta}}_n^{(j)}$ is between $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}_n$, and for all positive integer β we have

$$\begin{aligned}
&\left(D^{\beta}\mathbf{V}_{n,j}\right)\left(\widehat{\boldsymbol{\theta}}_n\right)\cdot(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_n)^{\beta} \\
&= \sum_{\beta_1,\ldots,\beta_d=1}^{\beta} \frac{\partial^{\beta}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial^{\alpha_1}\theta_1\cdots\partial^{\alpha_d}\theta_d}(\boldsymbol{\theta}_{\alpha_1}-\widehat{\boldsymbol{\theta}}_{n,1})^{\beta_1}\cdots(\boldsymbol{\theta}_{\alpha_d}-\widehat{\boldsymbol{\theta}}_{n,d})^{\beta_d},
\end{aligned} \tag{6.35}$$

such that

$$\sum_{j=1}^{d}\beta_j = \beta.$$

Consider the equality (6.35) and remark that

$$\begin{aligned}
&\frac{\partial^{\boldsymbol{\beta}}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial^{\alpha_1}\theta_1\cdots\partial^{\alpha_d}\theta_d}\prod_{j=1}^{d}(\boldsymbol{\theta}_{\alpha_j}-\widehat{\boldsymbol{\theta}}_{n,j})^{\beta_j} \\
&= \frac{1}{\sqrt{n}}\frac{\partial^{\boldsymbol{\beta}}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial^{\alpha_1}\theta_1\cdots\partial^{\alpha_d}\theta_d}\left(\prod_{j=1}^{d}\left[\sqrt{n}(\boldsymbol{\theta}_{\alpha_j}-\widehat{\boldsymbol{\theta}}_{n,j})\right]^{\beta_j}\right)\times\frac{1}{n^{\frac{\beta-1}{2}}}.
\end{aligned}$$

Since $\sqrt{n}(\boldsymbol{\theta}_{\alpha_j} - \widehat{\boldsymbol{\theta}}_{n,j})$, for $j = 1, \ldots, d$, are bounded in probability and making use of the condition **(A.4)**, we deduce that

$$\frac{1}{\sqrt{n}} \frac{\partial^{\beta} \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial^{\alpha_1}\theta_1 \cdots \partial^{\alpha_d}\theta_d} \left(\prod_{j=1}^{d} \left[\sqrt{n}(\boldsymbol{\theta}_{\alpha_j} - \widehat{\boldsymbol{\theta}}_{n,j}) \right]^{\beta_j} \right) = O_{\mathbb{P}}(1).$$

For $\beta \geq 2$, the quantity $\frac{1}{n^{\frac{\beta-1}{2}}}$ tends to 0 as $n \longrightarrow \infty$, then we deduce that

$$\frac{1}{(p)!} \left(D^{p}\mathbf{V}_{n,j}\right) \left(\tilde{\boldsymbol{\theta}}_n^{(j)}\right) \cdot (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_n^{(j)})^p = o_{\mathbb{P}}(1).$$

For $\beta = 1$,

$$\frac{1}{n^{\frac{\beta-1}{2}}} = 1,$$

then we deduce that the random variable $\mathcal{D}_{n,j}$ is just bounded in probability. Combining these last equalities with (6.34), we find

$$\mathbf{V}_{n,j}(\boldsymbol{\theta}) \;=\; \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \mathcal{D}_{n,j} + o_{\mathbb{P}}(1). \tag{6.36}$$

Observe that the quantity

$$\mathcal{D}_{n,j} = \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) + \cdots + \frac{1}{(p-1)!}\left(D^{\boldsymbol{p-1}}\mathbf{V}_{n,j}\right)\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^{p-1},$$

depends on the unknown parameter $\boldsymbol{\theta}$. Consider the following decomposition

$$\boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) \;=\; \sum_{j=1}^{d} \frac{1}{\sqrt{n}} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \theta_j} \sqrt{n}(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_{n,j}).$$

By the condition **(A.3)** and by applying Proposition 2.1, we obtain

$$\begin{aligned}\boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) \;&=\; \left(\sum_{j=1}^{d} \frac{1}{\sqrt{n}} \frac{\partial \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)}{\partial \theta_j} \sqrt{n}(\boldsymbol{\theta}_{N,j} - \widehat{\boldsymbol{\theta}}_{n,j}) \right) + o_{\mathbb{P}}(1) \\ &=: \; \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_n) \;+\; o_{\mathbb{P}}(1).\end{aligned} \tag{6.37}$$

By a similar reasoning and making use of Proposition 2.8 combined with the condition **(A.4)**, it follows that for all integer $\alpha \geq 2$

$$\left(D^{\alpha}\mathbf{V}_{n,j}\right)\left(\widehat{\theta}_n\right) \cdot (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^{\alpha} \;=\; \left(D^{\alpha}\mathbf{V}_{n,j}\right)\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_n)^{\alpha} + o_{\mathbb{P}}(1). \tag{6.38}$$

By combining (6.37) with (6.38), implies that equation (6.36) may be rewritten as follows

$$\mathbf{V}_{n,j}(\boldsymbol{\theta}) \;=\; \mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) + \widehat{\mathcal{D}}_{n,j} + o_{\mathbb{P}}(1), \tag{6.39}$$

where

$$\widehat{\mathcal{D}}_{n,j} = \boldsymbol{\nabla}\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_n) + \cdots + \frac{1}{(p-1)!}\left(D^{p-1}\mathbf{V}_{n,j}\right)\left(\widehat{\boldsymbol{\theta}}_n\right) \cdot (\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_n)^{p-1}. \tag{6.40}$$

Notice that equality (6.39) gives an explicit link between the random variable $\mathbf{V}_{n,j}(\boldsymbol{\theta})$ with it's estimated version $\mathbf{V}_{n,j}\left(\widehat{\boldsymbol{\theta}}_n\right)$.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

2. Applying Theorem 2.6 to see that (6.39) can be expressed as follows

$$\mathbf{V}_{N,j}(\boldsymbol{\theta}) = \mathbf{V}_{N,j}\left(\widehat{\boldsymbol{\theta}}_s\right) + \widehat{\mathcal{D}}^{(p)}_{s,m,j} + o_{\mathbb{P}}(1), \tag{6.41}$$

with

$$\begin{aligned} \widehat{\mathcal{D}}^{(p)}_{s,m,j} &= -\sqrt{\frac{m}{N}}\left[\mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_s\right) - \mathbf{V}_{m,j}\left(\widehat{\boldsymbol{\theta}}_m\right)\right] \\ &+ \sqrt{\frac{s}{N}}\sum_{i=1}^{d}\frac{\partial \mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_s\right)}{\partial \boldsymbol{\theta}_i}\left(\widehat{\boldsymbol{\theta}}_{N,i} - \widehat{\boldsymbol{\theta}}_{s,i}\right) \\ &+ \sqrt{\frac{m}{N}}\widehat{\mathcal{D}}_{s,j} \end{aligned}$$

and

$$\begin{aligned} \widehat{\mathcal{D}}_{s,j} &= \boldsymbol{\nabla}\mathbf{V}_{s,j}\left(\widehat{\boldsymbol{\theta}}_n\right)\cdot\left(\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_s\right) \\ &+ \cdots + \frac{1}{(p-1)!}\left(D^{p-1}\mathbf{V}_{s,j}\right)\left(\widehat{\boldsymbol{\theta}}_s\right)\cdot\left(\boldsymbol{\theta}_N - \widehat{\boldsymbol{\theta}}_s\right)^{p-1}. \end{aligned}$$

3. By using similar reasoning as in the proof of Theorem 2.3, we shall prove the existence of an estimate $\overline{\boldsymbol{\theta}}_N$ of $\boldsymbol{\theta}$, such that

$$\mathbf{V}_N(\boldsymbol{\theta}) = \mathbf{V}_N\left(\overline{\boldsymbol{\theta}}_N\right) + o_{\mathbb{P}}(1),$$

and

$$\overline{\boldsymbol{\theta}}_N = \widehat{\boldsymbol{\theta}}_s + A_N^{-1}\left(\widehat{\boldsymbol{\theta}}_s\right)\widehat{\mathcal{D}}^{(p)}_{s,m}.$$

4. The proof is similar as in Theorems (2.5) and (2.6).

Hence the proof of the theorem is complete. □

Proof of Theorem 3.1

Let n be a positive integer such that $n = o(N)$. In the sequel, for all $j = 1, \ldots, d$,

$$\widehat{\boldsymbol{\theta}}_{j,n} = \left(\widehat{\boldsymbol{\theta}}_{j,n,1}, \ldots, \widehat{\boldsymbol{\theta}}_{j,n,d}\right)$$

denotes a $\sqrt{n}-$consistent estimate of $\boldsymbol{\theta}_0 = \left(\boldsymbol{\theta}_{0,1}, \ldots, \boldsymbol{\theta}_{0,d}\right)$. Following similar reasonig as in the proof of Stute (1997)[Theorem 1.2], we have, for $j = 1, \ldots, d$,

$$\begin{aligned} \mathcal{R}^{\mathbb{1}}_{n,j}(x) &= \mathcal{R}_{n,j}(x) - n^{\frac{1}{2}}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{j,n})^{\top}\cdot G_j(x, \boldsymbol{\theta}_0) + o_{\mathbb{P}}(1), \quad (6.42) \\ n^{\frac{1}{2}}(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}}_{j,n})^{\top}\cdot G_j(x, \boldsymbol{\theta}_0) &= \sum_{i=1}^{d} G_{j,i}(x, \boldsymbol{\theta}_0)\left(n^{\frac{1}{2}}(\boldsymbol{\theta}_{0,i} - \widehat{\boldsymbol{\theta}}_{j,n,i})\right). \end{aligned}$$

From **(D.2)**(ii) and (3.15), we deduce that

$$G_{j,i}(x,\boldsymbol{\theta}_0) \leq |G_{j,i}(x,\boldsymbol{\theta}_0)| < \infty.$$

An application of Proposition 2.1 gives

$$G_{j,i}(x,\boldsymbol{\theta}_0) \times \sqrt{n}\left(\boldsymbol{\theta}_{0,i} - \widehat{\boldsymbol{\theta}}_{j,n,i}\right) = G_{j,i}(x,\boldsymbol{\theta}_0) \times \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n,i}\right) + o_{\mathbb{P}}(1).$$

This last equality in connection with (6.42) gives, for $j = 1, \ldots, d$,

$$\mathcal{R}^{\mathbb{1}}_{n,j}(x) \;=\; \mathcal{R}_{n,j}(x) - n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot G_j(x,\boldsymbol{\theta}_0) + o_{\mathbb{P}}(1). \tag{6.43}$$

Thus, it will be rewritten as, for $j = 1, \ldots, d$,

$$\mathcal{R}_{n,j}(x) \;=\; \mathcal{R}^{\mathbb{1}}_{n,j}(x) + n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot G_j(x,\boldsymbol{\theta}_0) + o_{\mathbb{P}}(1). \tag{6.44}$$

Consider the following decomposition

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot G_j(x,\boldsymbol{\theta}_0) \;&=\; n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot \Big(G_j(x,\boldsymbol{\theta}_0) - G_j(x,\widehat{\boldsymbol{\theta}}_{j,n})\Big) \\ &\quad + n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot G_j(x,\widehat{\boldsymbol{\theta}}_{j,n}). \end{aligned}$$

Equation (3.15) ensures the continuity of each component $G_{j,i}(x,\boldsymbol{\theta})$ at each $\boldsymbol{\theta}$ in interior of the set $\boldsymbol{\Theta}$. Since $\widehat{\boldsymbol{\theta}}_{j,n}$ is a consistent estimate of $\boldsymbol{\theta}_0$, and according to the continuous mapping theorem, we deduce that

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot \Big(G_j(x,\boldsymbol{\theta}_0) - G_j(x,\widehat{\boldsymbol{\theta}}_{j,n})\Big) = o_{\mathbb{P}}(1).$$

Finally equality (6.44) will be rewritten, for $j = 1, \ldots, d$, as follows

$$\mathcal{R}_{n,j}(x) \;=\; \mathcal{R}^{\mathbb{1}}_{n,j}(x) + n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,n})^{\top} \cdot G_j(x,\widehat{\boldsymbol{\theta}}_{j,n}) + o_{\mathbb{P}}(1). \tag{6.45}$$

Consider now the empirical process with all observations N and doing a similar decomposition as in the proof of Theorem 2.6, we write

$$\begin{aligned} \mathcal{R}_{N,j}(x) \;&=\; N^{-1/2}\sum_{i=1}^{N} \mathbb{1}_{\{X_i \leq x\}}\Big[Y_{j,i} - m_j(X_i,\boldsymbol{\theta})\Big] \\ &=\; \sqrt{\frac{s}{N}}\,\mathcal{R}_{s,j}(x) + \sqrt{\frac{m}{N}}\,\mathcal{R}_{m,j}(x), \end{aligned} \tag{6.46}$$

Because $s = o(N)$, $m = o(N)$ and $m < s < N$, we may apply (6.45) on (6.46) and obtain

$$\begin{aligned} \mathcal{R}_{N,j}(x) \;&=\; \sqrt{\frac{s}{N}}\Big[\mathcal{R}^{\mathbb{1}}_{s,j}(x) + s^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,s})^{\top} \cdot G_j(x,\widehat{\boldsymbol{\theta}}_{j,s})\Big] \\ &\quad + \sqrt{\frac{m}{N}}\Big[\mathcal{R}^{\mathbb{1}}_{m,j}(x) + m^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,m})^{\top} \cdot G_j(x,\widehat{\boldsymbol{\theta}}_{j,m})\Big] + o_{\mathbb{P}}(1) \end{aligned}$$

$$
\begin{aligned}
&= \mathcal{R}^{\mathbb{1}}_{N,j}(x) + \sqrt{\frac{s}{N}}\left[s^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,s})^\top \cdot G_j(x, \widehat{\boldsymbol{\theta}}_{j,s})\right] \\
&\quad + \sqrt{\frac{m}{N}}\left[m^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}}_{j,N} - \widehat{\boldsymbol{\theta}}_{j,m})^\top \cdot G_j(x, \widehat{\boldsymbol{\theta}}_{j,m})\right] + o_{\mathbb{P}}(1) \\
&=: \mathcal{R}^{\mathbb{1}}_{N,j}(x) + \mathbf{G}_j\left(x, \widehat{\boldsymbol{\theta}}_{j,s}, \widehat{\boldsymbol{\theta}}_{j,m}\right) + o_{\mathbb{P}}(1).
\end{aligned}
$$

It is easy to check that for $j \in \{1, \ldots, d\}$, the quantity $\mathbf{G}_j\left(x, \widehat{\boldsymbol{\theta}}_{j,s}, \widehat{\boldsymbol{\theta}}_{j,m}\right)$ is bounded. Hence the Theorem is established. □

Proof of Theorem 3.2

For each $j = 1, \ldots, d$, let Γ_j be a tangent space of the regression function m_j at $\widehat{\boldsymbol{\theta}}_{j,N}$ defined by

$$
\Gamma_j := \left\{(z, m_j(\cdot, z)) \in \mathbb{R}^d \times \mathbb{R}, \text{ such that, } m_j(\cdot, z) - m_j(\cdot, \widehat{\boldsymbol{\theta}}_{j,N}) = \boldsymbol{\nabla} m_j(\cdot, \widehat{\boldsymbol{\theta}}_{j,N}) \cdot \left(z - \widehat{\boldsymbol{\theta}}_{j,N}\right)\right\} \quad (6.47)
$$

For each $j = 1, \ldots, d$, and with disrupting a component of the first estimate $\widehat{\boldsymbol{\theta}}_{j,N}$ and keeping the others unchanged, we construct another estimate $\bar{\boldsymbol{\theta}}_{j,N}$. More precisely, if we disrupt $\widehat{\boldsymbol{\theta}}_{j,N}$ at the d_j-th component with a quantity p_j, that we describe as follows

$$
\bar{\boldsymbol{\theta}}_{j,N,r} = \widehat{\boldsymbol{\theta}}_{j,N,r}, \quad \text{for } r = 1, \ldots, d\,, \quad \text{and} \quad r \neq d_j, \quad (6.48)
$$

$$
\bar{\boldsymbol{\theta}}_{j,N,d_j} = \widehat{\boldsymbol{\theta}}_{j,N,d_j} + p_j, \quad \text{for } p_j \neq 0. \quad (6.49)
$$

From (6.47) and according to the definition of the modified estimator $\bar{\boldsymbol{\theta}}_{j,N}$, we obtain, for $j = 1, \ldots, d$,

$$
\begin{aligned}
m_j(X_1, \bar{\boldsymbol{\theta}}_{j,N}) - m_j(X_1, \widehat{\boldsymbol{\theta}}_{j,N}) &= \frac{\partial m_j(X_1, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} \times p_j, \\
&\ \vdots \\
m_j(X_N, \bar{\boldsymbol{\theta}}_{j,N}) - m_j(X_N, \widehat{\boldsymbol{\theta}}_{j,N}) &= \frac{\partial m_j(X_N, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} \times p_j.
\end{aligned}
$$

By multiplying each equality by its corresponding indicator function, for $j = 1, \ldots, d$, we find

$$
\begin{aligned}
\left(m_j(X_1, \bar{\boldsymbol{\theta}}_{j,N}) - m_j(X_1, \widehat{\boldsymbol{\theta}}_{j,N})\right)\mathbb{1}_{\{X_1 \leq x\}} &= \frac{\partial m_j(X_1, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} p_j \mathbb{1}_{\{X_1 \leq x\}}, \\
&\ \vdots \\
\left(m_j(X_N, \bar{\boldsymbol{\theta}}_{j,N}) - m_j(X_N, \widehat{\boldsymbol{\theta}}_{j,N})\right)\mathbb{1}_{\{X_N \leq x\}} &= \frac{\partial m_j(X_N, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} p_j \mathbb{1}_{\{X_N \leq x\}}.
\end{aligned}
$$

By summing and multiplying the result by $-N^{\frac{-1}{2}}$, we find

$$
-N^{\frac{-1}{2}} \sum_{i=1}^{N} \left(m_j(X_i, \bar{\boldsymbol{\theta}}_{j,N}) - m_j(X_i, \widehat{\boldsymbol{\theta}}_{j,N})\right)\mathbb{1}_{\{X_i \leq x\}} = -N^{\frac{-1}{2}} p_j \sum_{i=1}^{N} \frac{\partial m_j(X_i, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} \mathbb{1}_{\{X_i \leq x\}}. \quad (6.50)
$$

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Let $\mathcal{R}^2_{N,j}(x)$ be the process obtained from $\mathcal{R}_{N,j}(x)$ by replacing $\boldsymbol{\theta}_0$ by its estimate $\bar{\boldsymbol{\theta}}_{j,N}$. By straightforward calculus and making use of (6.50), we readily obtain

$$\begin{aligned}\mathcal{R}^2_{N,j}(x) - \mathcal{R}^{\mathbb{1}}_{N,j}(x) &= -N^{\frac{-1}{2}} \sum_{i=1}^{N} \Big(m_j(X_i, \bar{\boldsymbol{\theta}}_{j,N}) - m_j(X_i, \widehat{\boldsymbol{\theta}}_{j,N}) \Big) \mathbb{1}_{\{X_i \le x\}} \\ &= -N^{\frac{-1}{2}} p_j \sum_{i=1}^{N} \frac{\partial m_j(X_i, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} \mathbb{1}_{\{X_i \le x\}}. \end{aligned} \tag{6.51}$$

From Theorem 3.1, we find

$$\mathcal{R}_{N,j}(x) - \mathcal{R}^{\mathbb{1}}_{N,j}(x) = \mathbf{G}_j \left(x, \widehat{\boldsymbol{\theta}}_{j,s}, \widehat{\boldsymbol{\theta}}_{j,m} \right) + o_{\mathbb{P}}(1). \tag{6.52}$$

By imposing the following identity

$$-N^{\frac{-1}{2}} \sum_{i=1}^{N} \frac{\partial m_j(X_i, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} \mathbb{1}_{\{X_i \le x\}} = \mathbf{G}_j \left(x, \widehat{\boldsymbol{\theta}}_{j,s}, \widehat{\boldsymbol{\theta}}_{j,m} \right),$$

We readily obtain

$$p_j = \frac{\mathbf{G}_j \left(x, \widehat{\boldsymbol{\theta}}_{j,s}, \widehat{\boldsymbol{\theta}}_{j,m} \right)}{-N^{\frac{-1}{2}} \sum_{i=1}^{N} \frac{\partial m_j(X_i, \widehat{\boldsymbol{\theta}}_{j,N})}{\partial \theta_{d_j}} \mathbb{1}_{\{X_i \le x\}}},$$

and we infer that, for $j = 1, \ldots, d$,

$$\mathcal{R}_{N,j}(x) - \mathcal{R}^{\mathbb{1}}_{N,j}(x) = \mathcal{R}^2_{N,j}(x) - \mathcal{R}^{\mathbb{1}}_{N,j}(x) + o_{\mathbb{P}}(1). \tag{6.53}$$

Obviously, equality (6.52) implies that, for $j = 1, \ldots, d$,

$$\mathcal{R}_{N,j}(x) = \mathcal{R}^2_{N,j}(x) + o_{\mathbb{P}}(1),$$

where

$$\mathcal{R}^2_{N,j}(x) = \mathcal{R}_{N,j}(x)(\bar{\boldsymbol{\theta}}_{j,N}).$$

This implies that

$$\mathcal{R}_{N,j}(x) = \mathcal{R}_{N,j}(x)(\bar{\boldsymbol{\theta}}_{j,N}) + o_{\mathbb{P}}(1).$$

Thus the proof is achieved. □

Acknowledgements

The authors would like to thank the anonymous referee and the Editor in chief for their constructive comments on an earlier version of this manuscript.

Bibliography

Alvarez-Andrade, S. and Bouzebda, S. (2014). Asymptotic results for hybrids of empirical and partial sums processes. *Statist. Papers*, **55**(4), 1121–1143.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag, New York. Reprint of the 1993 original.

Cheng, R. (2017). *Non-standard parametric statistical inference*. Oxford University Press, Oxford.

Delgado, M. A., Hidalgo, J., and Velasco, C. (2005). Distribution free goodness-of-fit tests for linear processes. *Ann. Statist.*, **33**(6), 2568–2609.

Drost, F. C., Klaassen, C. A. J., and Werker, B. J. M. (1997). Adaptive estimation in time-series models. *Ann. Statist.*, **25**(2), 786–817.

Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statist. Sinica*, **17**(1), 115–138.

Hall, W. J. and Mathiason, D. J. (1990). On large-sample estimation and testing in parametric models. *International Statistical Review / Revue Internationale de Statistique*, **58**(1), 77–97.

Hwang, S. Y. and Basawa, I. V. (1993). Asymptotic optimal inference for a class of nonlinear time series models. *Stochastic Process. Appl.*, **46**(1), 91–113.

Hwang, S. Y. and Basawa, I. V. (2001). Nonlinear time series contiguous to $\mathrm{AR}(1)$ processes and a related efficient test for linearity. *Statist. Probab. Lett.*, **52**(4), 381–390.

Hwang, S. Y. and Basawa, I. V. (2003). Estimation for nonlinear autoregressive models generated by beta-ARCH processes. *Sankhyā*, **65**(4), 744–762.

Khmaladze, E. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Teor. Veroyatnost. i Primenen.*, **26**(2), 246–265.

Khmaladze, E. V. (1993). Goodness of fit problem and scanning innovation martingales. *Ann. Statist.*, **21**(2), 798–829.

Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. Statist.*, **32**(3), 995–1034.

Koul, H. L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Ann. Statist.*, **34**(2), 994–1012.

Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.*, **27**(1), 204–236.

Koul, H. L., Stute, W., and Li, F. (2005). Model diagnosis for SETAR time series. *Statist. Sinica*, **15**(3), 795–817.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. california Publ. Statist.*, **3**, 37–98.

Le Cam, L. (1974). *Notes on asymptotic methods in statistical decision theory*. Centre de Recherches Mathématiques, Université de Montréal, Montreal, Que.

Le Cam, L. (1986). *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr

Lindsey, J. K. (1996). *Parametric statistical inference*. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York.

Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory*, **9**(4), 539–569.

Lounis, T. (2011). Asymptotically optimal tests when parameters are estimated. *Albanian J. Math.*, **5**(4), 193–214.

Lounis, T. (2013). Local asymptotically optimal test in Arch model. *Rev. Roumaine Math. Pures Appl.*, **58**(4), 333–392.

Lounis, T. (2017). Optimal tests in $\mathrm{AR}(m)$ time series model. *Comm. Statist. Simulation Comput.*, **46**(2), 1583–1610.

Pfanzagl, J. (1994). *Parametric statistical theory*. De Gruyter Textbook. Walter de Gruyter & Co., Berlin. With the assistance of R. Hamböker.

Robinson, P. M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *Ann. Statist.*, **33**(4), 1800–1842.

Roussas, G. G. (1972). *Contiguity of probability measures: some applications in statistics*. Cambridge University Press, London-New York. Cambridge Tracts in Mathematics and Mathematical Physics, No. 63.

Shiryaev, A. N. and Spokoiny, V. G. (2000). *Statistical experiments and decisions*, volume 8 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific Publishing Co., Inc., River Edge, NJ. Asymptotic theory.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, **25**(2), 613–641.

Stute, W., Presedo Quindimil, M., González Manteiga, W., and Koul, H. L. (2006). Model checks of higher order time series. *Statist. Probab. Lett.*, **76**(13), 1385–1396.

Swensen, A. R. (1985). The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *J. Multivariate Anal.*, **16**(1), 54–70.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

© 2019 ISTE OpenScience – Published by ISTE Ltd. London, UK – openscience.fr